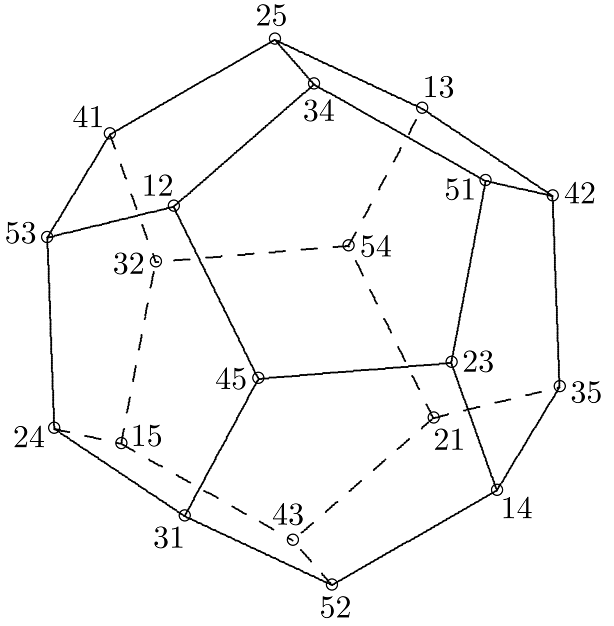
Abstract Algebra I

(Math 5550)



G. Eric Moorhouse

Department of Mathematics
UNIVERSITY OF WYOMING

Preface

These lecture notes were used in teaching Abstract Algebra I (Math 709, later Math 5550) at the University of Wyoming in the Spring 1990, 1993 and 2004 semesters. Many of the exercises included were also used in homework assignments and tests during the semester. I expect there remain several gaps, misprints and errors in these notes. These will be posted on the website

`http://www.uwyo.edu/moorhouse/courses/5550/errata.html`

I would appreciate receiving any suggestions and corrections which I can include at this site, and which will be useful in future revisions of these notes.

The main topics presented are groups, rings, fields and Galois theory. Several standard topics are omitted (such as group presentations, factorization theorems for rings, linear representations and characters of finite groups) or stated without proof (e.g. the Jordan-Hölder Theorem) in order to reach the goal of rudimentary Galois Theory by the end of the course, a goal we proved to be practical and worthwhile. In such a course as this, which is required for all our graduate students, not just those specializing in pure mathematics, I believe that abstract group theory is less significant than understanding the concept of group actions. Accordingly I have stressed permutation actions of groups, which shows how groups arise 'in nature' and stays fairly close to the historical origins of finite group theory (as groups of permutations, including Galois actions on zeroes of polynomials).

In these notes, I assume no prior expertise in groups or rings. I do expect the reader has reasonable experience with sets and functions, elementary number theory, and an appreciation for the notion of a proof. Some knowledge of linear algebra is also assumed, in particular matrix arithmetic, and the concepts of basis and dimension. I do not expect that these notes would be suitable, however, for any of our undergraduate courses, except perhaps for individual study in special cases.

With so many good abstract algebra textbooks on the market, why these lecture notes? Mathematical textbooks tend to include far more material in every chapter than a typical first algebra course can reasonably cover. This is the natural outcome of the motivation of authors and publishers to offer a small number of current titles of a very comprehensive nature, in order to accommodate all the whims of every instructor, whose job it is then to pick and choose what portion of the textbook appeals to her/him. Unfortunately the student must then work harder to extract from the textbook what the instructor requires. By contrast, these contain no more and no less than students in my class are expected to know.

In order to get through as much of the most basic material as possible, many of the standard results are found only in the homework exercises. Accordingly, *students are expected to solve all of the exercises;* any exercises which cannot complete themselves as assigned homework, they should later seek help in learning to solve. In order to help keep these expectations realistic, ample detailed hints are provided for many of the exercises. *Exception:* Those problems designated by asterisks (*) are considered more challenging and may be considered supplementary.

My desire in these notes has been to be as informal and visual as possible. For example, many results are outlined in discussion format without being dignified as Theorem so-and-so.

Eric Moorhouse
1993, 2004, 2005

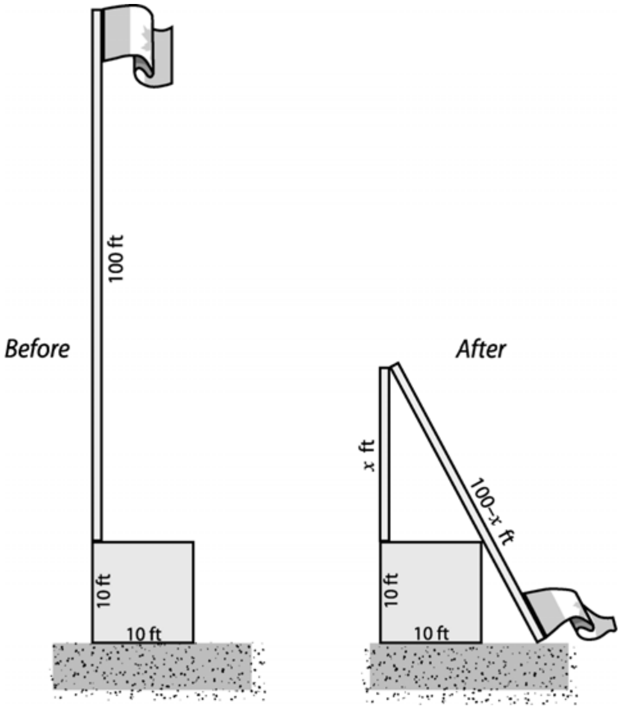
Prologue

The goal of this course is to cover certain topics in groups (especially finite groups), rings and fields. Many standard topics in group theory and ring theory will be bypassed so that we will have time to reach our goal by the end of the course, namely some basic Galois theory.

By the time we have done a little Galois Theory, you will have attained an appreciation for the interrelationships between different areas of mathematics. It is these interrelationships that so often make mathematics, and especially Galois theory, so beautiful.

When I was in secondary school in Saint John, the Head of our school's Math/Science Department showed us a problem that he said was too difficult for him. Of course this made the problem irresistible to me! The problem was as follows:

> A 10 ft. × 10 ft. building has a 100 ft. flagpole erected on one of its upper corners. During a storm, the flagpole is cracked at a point of height x above the roof, such that the tip of the flagpole just touches the ground, while the flagpole also just touches another corner of the building (see diagram).



> Find x.

This led me to a cubic polynomial equation for x, namely

$$11x^3 - 490x^2 + 100x + 500 = 0.$$

Numerical approximation shows that this has three real roots: 44.317, 1.133, and the extraneous root -0.905. However, I was not interested in approximations. I wanted the real thing! After all, the quadratic equation $ax^2+bx+c=0$ could be solved *exactly* using the *Almighty Formula*

$$x = \frac{b \pm \sqrt{b^2-4ac}}{2a}\,.$$

Every day I would rush home from school to my pile of rough work, in a futile attempt to factor my cubic equation, or transform it into a quadratic equation, or even find another equation for x which would simplify the problem. It was weeks later that I discovered that every cubic equation could be transformed (by an affine change of variable) to one of the form $x^3+px=q$, for which Fontana's solutions are

$$x_1 = \sqrt[3]{\frac{q}{2}+\sqrt{\frac{p^3}{27}+\frac{q^2}{4}}} + \sqrt[3]{\frac{q}{2}-\sqrt{\frac{p^3}{27}+\frac{q^2}{4}}}\,,$$

$$x_2 = \ldots, \qquad x_3 = \ldots \quad .$$

I also discovered that similar solutions existed for the general quartic (degree 4) polynomial equations, and that these general solutions required the extraction of *fourth* roots, as expected. It was only much later that I learned why the general quintic (degree 5) equation has no such solution in terms of fifth roots (or in fact using *any* radicals. Indeed Galois theory shows that the question of whether a given polynomial equation is solvable by radicals, reduces to the question of whether a certain group (now known as the Galois group of the polynomial) is 'solvable'.

We mention another couple of classical problems whose answers depend (somewhat surprisingly) on field theory. One is the impossibility of trisecting an arbitrary angle using straightedge and compass. You will see once and for all why this is so. Another is the impossibility of finding an antiderivative

$$\int e^{x^2}\,dx$$

in elementary terms. This topic we may possibly present in an extra seminar, once we have covered enough field theory.

Let's forge ahead, then, with an introduction to group theory.

0. Integers

We review some preliminary facts concerning the ring $\mathbb{Z}$ of integers. (Although the definition of a ring does not appear until Section 14, nevertheless all these facts about $\mathbb{Z}$ should be familiar to you.)

Let $a, b \in \mathbb{Z}$. We say that a **divides** b (or a is a **divisor** of b, or b is a **multiple** of a) if $b = da$ for some $d \in \mathbb{Z}$. We write the statement 'a divides b' symbolically as $a \,\big|\, b$; if a does *not* divide b, we write $a \not\big| b$. The following results are well known.

0.1 Proposition. Let $a, b, c \in \mathbb{Z}$.

(i) If $a \,\big|\, b$ and $b \,\big|\, c$, then $a \,\big|\, c$.

(ii) If $a \,\big|\, b$ and $a \,\big|\, c$, then $a \,\big|\, (b + c)$ and $a \,\big|\, (b - c)$.

0.2 Theorem (Division Algorithm). Let $a, d \in \mathbb{Z}$ and suppose that $d > 0$. Then there exist unique integers q, r such that $a = qd + r$ and $0 \le r < d$.

The algorithm (i.e. procedure) for determining the quotient q and remainder r is familiar; for example when we divide 103 by 7, we obtain 14 as the quotient and 5 as the remainder:

$$\begin{array}{r|r} & 14 \\ \cline{2-2} 7 & 103 \\ & \underline{7} \\ & 33 \\ & \underline{28} \\ & 5 \end{array}$$

Thus $103 = 14 \cdot 7 + 5$. Calling Theorem 0.2 an 'algorithm' is rather a misnomer, but a popular one.

The divisors of 12 are $\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12$; the divisors of 18 are $\pm 1, \pm 2, \pm 3, \pm 6, \pm 9, \pm 18$. Thus the *common* divisors of 12 and 18 are $\pm 1, \pm 2, \pm 3, \pm 6$. The greatest (i.e. largest) of these common divisors is therefore 6. We write $\gcd(12, 18) = 6$ (the *greatest common divisor* of 12 and 18 is 6). Given any two integers $a, b \in \mathbb{Z}$, not both zero, we similarly denote their greatest common divisor by $\gcd(a, b)$. (Note that $\gcd(0, 0)$ is undefined since the common divisors of 0 and 0 are all the integers.) We say m and n are **relatively prime** if $\gcd(m, n) = 1$; for example, 12 and 35 are relatively prime.

Similarly, the multiples of 12 are $0, \pm 12, \pm 24, \pm 36, \pm 48, \ldots$; the multiples of 18 are $0, \pm 18, \pm 36, \pm 54, \ldots$; and the *common* multiples are $0, \pm 36, \pm 72, \ldots$. Of all the positive common multiples of 12 and 18, the least is 36, and we write $\operatorname{lcm}(12, 18) = 36$ (the *least common multiple* is 36).

By the Fundamental Theorem of Arithmetic, every positive integer has a unique prime factorization. We see from the prime factorizations $12 = 2^2 3^1$ and $18 = 2^1 3^2$ that $\gcd(12, 18) = 2^1 3^1 = 6$. Similarly, $\operatorname{lcm}(12, 18) = 2^2 3^2 = 36$. More generally, if $m, n \in \mathbb{Z}$ are nonzero, let $p_1, p_2, \ldots, p_k$ be the primes dividing at least one of m and n. Write the prime factorizations of m and n as

$$m = \prod_{i \in I} p_i^{r_i}, \qquad n = \prod_{i \in I} p_i^{s_i}$$

where $r_i, s_i \geq 0$ for all $i \in I = \{1, 2, \ldots, k\}$. Then

0.3 Proposition.

(i) $\gcd(m, n) = \prod_{i \in I} p_i^{\min\{r_i, s_i\}}$; and (ii) $\operatorname{lcm}(m, n) = \prod_{i \in I} p_i^{\max\{r_i, s_i\}}$.

From these formulae it follows easily that $mn = \gcd(m, n) \operatorname{lcm}(m, n)$, so determining the greatest common divisor of two given integers is just as hard (or just as easy) as finding their least common multiple. For large numbers, prime factorizations are not easily computed, so Proposition 0.3 will be of little value in computing $\gcd(m, n)$. However, Euclid's Algorithm gives us the answer quickly, by repeated application of the Division Algorithm. For example, we determine $\gcd(108, 74)$:

$$\begin{aligned} 108 &= 1 \cdot 74 + 34 \\ 74 &= 2 \cdot 34 + 6 \\ 34 &= 5 \cdot 6 + 4 \\ 6 &= 1 \cdot 4 + 2 \\ 4 &= 2 \cdot 2 + 0 \end{aligned}$$

and the last nonzero remainder gives $\gcd(108, 74) = 2$. Moreover reversing these steps, we are able to express 2 as an integer-linear combination of 108 and 74, thus:

$$\begin{aligned} 2 &= 6 - 4 \\ &= 6 - (34 - 5 \cdot 6) \\ &= 6 \cdot 6 - 34 \\ &= 6(74 - 2 \cdot 34) - 34 \\ &= 6 \cdot 74 - 13 \cdot 34 \\ &= 6 \cdot 74 - 13(108 - 74) \\ &= 19 \cdot 74 - 13 \cdot 108. \end{aligned}$$

More generally

0.4 Theorem (Euclid's Algorithm). Let m and n be integers, not both zero, and let $d = \gcd(m, n)$. Then there exist integers x, y such that $d = mx + ny$.

The algorithm described above (by example) succeeds very quickly even for very large integers m and n because, with repeated application of the Division Algorithm, the remainders decrease in size rather quickly. Again, it is a popular misnomer (to which we accede) to refer to Theorem 0.4 as Euclid's *Algorithm,* which is more precisely the algorithm described above for computing d, x and y.

It is perhaps already apparent that the properties listed above hold for rings other than $\mathbb{Z}$; in particular the ring $\mathbb{R}[X]$ of polynomials in X with real coefficients, admits a Division Algorithm, and so Euclid's Algorithm applies also in $\mathbb{R}[X]$. The rings $\mathbb{Z}$ and $\mathbb{R}[X]$ are examples of the class of rings known as Euclidean domains, considered in Section 15, for which these results apply. But because the elementary properties of groups make use of these elementary properties of the integers, we have reviewed these properties at the outset.

Exercises 0.

1. Let m and n be positive integers. Prove that there exist relatively prime integers m' and n' such that $m' \,|\, m$, $n' \,|\, n$ and $\operatorname{lcm}(m, n) = m'n'$.

   Hint: Write $m = \prod_{i \in I} p_i^{r_i}$ and $n = \prod_{i \in I} p_i^{s_i}$ in the notation of Proposition 0.3. Let $I_1 = \{i \in I : r_i \geq s_i\}$ and $I_2 = \{i \in I : r_i < s_i\}$. Consider $m' = \prod_{i \in I_1} p_i^{r_i}$ and $n' = \prod_{i \in I_2} p_i^{s_i}$.

Chapter I

Groups

1. Definitions

A **binary operation** on a set G is a function $G \times G \to G$. Examples of binary operations include:

addition of real numbers, i.e. $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad (x, y) \mapsto x + y$;
multiplication of real numbers, i.e. $\times : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad (x, y) \mapsto x \times y$;
punctiliation of real numbers, i.e. $* : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \quad (x, y) \mapsto (x^4 - y^3)/(x^2 + y^2 + 1)$.

(I'm sorry, I don't know what punctiliation is good for. I made it up just now.) Some books use a symbol like $*$ or $\circ$ to denote an arbitrary binary operation. However in the general case we shall use *juxtaposition* of elements to denote the binary operation, so that the image of the pair $(x, y) \in G \times G$ is simply denoted $xy \in G$. It is important to remember, though, that this operation need not correspond to any usual notion of 'multiplication'; indeed, the operation $(x, y) \mapsto xy$ may represent actual addition, or something very unfamiliar, like punctiliation.

Certain choices of a binary operation on a set G will make G into a group, while others will not. In the above three examples, $\mathbb{R}$ is a group under addition, but not under multiplication or punctiliation. For this reason, we might say that the pair $(\mathbb{R}, +)$ is a group, while $(\mathbb{R}, \times)$ and $(\mathbb{R}, *)$ are not. However, if G is a set on which we have one clearly defined binary operation, it is unambiguous and acceptable to say simply that G itself is a group.

A **group** is a set G, together with a binary operation on G (here denoted simply by juxtaposition of elements of G), such that

(i) there exists $e \in G$ such that $xe = ex = x$ for all $x \in G$;
(ii) for all $x \in G$, there exists $y \in G$ such that $xy = yx = e$; and
(iii) the binary operation is *associative,* i.e. $(xy)z = x(yz)$ for all $x, y, z \in G$.

If G is a group, then by (i), G contains an identity element for the binary operation. In particular, every group is nonempty. It is easy to see that the identity of G is unique, for if e_1 and e_2 are identities for G, then $e_1 = e_1 e_2 = e_2$. We will denote the (unique) identity of G by e (or sometimes 1 or 0).

Furthermore, 'inverses' (as in (ii)) are unique in G. For if $x \in G$ has two inverses $y_1, y_2 \in G$ such that $xy_1 = y_1 x = e = xy_2 = y_2 x$ then

$$y_1 = y_1 e = y_1(xy_2) = (y_1 x)y_2 = ey_2 = y_2,$$

as claimed. We will usually denote the (unique) inverse of x in G by x^{-1}. Furthermore we denote

$$x^m = ((\cdots((xx)x)x)\cdots x)x = \overbrace{xx\cdots x}^{m \text{ times}}$$

for any positive integer m. (The parentheses above are irrelevant, of course, by associativity.) It is then straightforward to check that

$$x^m x^n = x^{m+n}, \qquad (x^m)^n = x^{mn}$$

for *all* integers m, n, with the additional conventions that $x^0 = e$ and $x^{-m} = (x^m)^{-1} = (x^{-1})^m$ for $m < 0$.

Note that the binary operation for a group is *not* required to be commutative. If $xy = yx$ for all $x, y \in G$, then G is said to be **abelian**; otherwise G is **nonabelian**.

The **order** of G is by definition $|G|$, i.e. the cardinality of the set of group elements. There is no reason why $|G|$ must be finite, but if it is, we say that G is a **finite group**. The **order** of an element $x \in G$ is the smallest positive integer m such that $x^m = e$. (If no such integer exists, we say that x has **infinite order**.) For example, e is the unique element of order 1.

Exercises 1.

1. For elements x and y in a group G, prove that $(xy)^{-1} = y^{-1}x^{-1}$. (This is the so-called Shoe-Sock Theorem: The opposite of putting on socks and shoes, is to first remove the shoes, then remove the socks.)

2. Prove that if $x^2 = e$ for every element x in a group G, then G is abelian.

3. Let G be a group, and suppose $x \in G$ has order m. Show that $x^i = x^j$ iff $i \equiv j \bmod m$. In particular, $x^k = e$ iff k is divisible by m.

   Hint: Suppose $x^k = e$. By the Division Algorithm 0.2, write $k = qm + r$ where $0 \le r < m$. Deduce that $x^r = e$ and so $r = 0$. More generally if $x^i = x^j$ then $x^{i-j} = e$ and the previous reasoning applies.

4. Let x, y be elements of a group G such that $xy = yx$. Prove that if $m = |x|$ and $n = |y|$ are relatively prime integers, then $|xy| = mn$.

   Hint: Suppose $(xy)^k = e$. Since x and y commute, this means that $x^k y^k = e$. Consider the element $z = x^k = y^{-k}$. Then $z^m = (x^k)^m = (x^m)^k = e$ and $z^n = (y^{-k})^n = (y^n)^{-k} = e$. By Exercise 1.3, $|z|$ divides both m and n...

5. (a) Show that every element of a finite group has finite order.

   (b) Find an example of an infinite group in which every element has finite order.

6. Let x and y be elements of a group G such that $xy = yx$. Suppose that x has finite order m, and y has finite order n. Show that there exist $r, s \in \mathbb{Z}$ such that $|x^r y^s| = \operatorname{lcm}(m, n)$.

   Hint: There exist integers m', n' as in Exercise 0.1. Let $r = m/m'$ and $s = n/n'$ and apply Exercise 1.4.

7. Let G be a finite group of even order. Prove that G has an element of order 2.

   Hint: If G has no element of order 2, then the nonidentity elements of G are partitioned into pairs $\{g, g^{-1}\}$.

2. Examples

Example: Multiplicative Rationals. The set of nonzero rational numbers under multiplication, i.e. $\mathbb{Q}^\times = \mathbb{Q} \smallsetminus \{0\}$, forms an infinite abelian group. Of course e is really 1 in this case. Likewise both $\mathbb{R}^\times$ and $\mathbb{C}^\times$ are infinite abelian groups under multiplication. (However, the nonzero integers *do not* form a multiplicative group, owing to the lack of multiplicative inverses.)

Example: Additive Integers. The set of integers under addition forms an infinite abelian group. (In this case note that xy is really $x + y$, e is really 0, x^m means $x + x + \cdots + x = mx$, and x^{-1} is really $-x$.) Similarly each of the sets $\mathbb{R}$, $\mathbb{Q}$ and $\mathbb{C}$ forms an infinite abelian group under addition.

Example: General Linear Groups of Degree n. The set of all invertible $n \times n$ matrices with real entries, denoted $GL_n(\mathbb{R})$, forms an infinite group, under the usual matrix multiplication. This group is nonabelian for $n > 1$. Similarly we have the groups $GL_n(\mathbb{C})$ and $GL_n(\mathbb{Q})$.

Example: Symmetric Groups. A **permutation of degree n** is a bijection from the set $\{1, 2, 3, \ldots, n\}$ to itself. For a given n, the set of all such bijections forms a group under composition, called the **symmetric group of degree n**, and denoted S_n. This group has order $n!$, and is nonabelian for $n > 2$.

We introduce notation for permutations as follows. Every permutation is the product (i.e. composite) of disjoint 'cycles'. For example in S_6, the cycle of length 4 (or 4-cycle) denoted (2634) is the permutation defined by

$$\begin{array}{r|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \scriptstyle(2634) & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 1 & 6 & 4 & 2 & 5 & 3 \end{array} .$$

In this example, (2634) takes 1 to 1, 2 to 6, etc. We could also have denoted this permutation by (1)(2634)(5), but it is customary here to suppress writing the cycles (1) and (5) of length 1. Note also that (2634) = (3426) = (4236) = (6342). We will also express the action of (2634) by writing $1^{(2634)} = 1$, $2^{(2634)} = 6$, etc. (This is preferable to saying $(2634)(1) = 1$, $(2634)(2) = 6$, etc., since we don't want our use of parentheses to start getting ambiguous.) An example of a permutation which is the product of two disjoint 3-cycles is (124)(365), which has the effect

$$\begin{array}{rcccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \scriptstyle(124)(365) & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ & 2 & 4 & 6 & 1 & 3 & 5 \end{array} .$$

Composing the above permutations, we have

$$\begin{array}{rcccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
{\scriptstyle (2634)} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 1 & 6 & 4 & 2 & 5 & 3 \\
{\scriptstyle (124)(365)} & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & 2 & 5 & 1 & 4 & 3 & 6
\end{array} \quad ,$$

which shows that $(2634)(124)(365) = (1253)$. We multiply (or compose) permutations from left to right like this, since we want to be able to say

$$\left(2^{(2634)}\right)^{(124)(365)} = 6^{(124)(365)} = 5 = 2^{(1253)} = 2^{(2634)(124)(365)}.$$

(*Warning:* Some books define multiplication of permutations using right-to-left composition!) It is customary to write the identity of S_n as (1), which is an abbreviation for $(1)(2)(3)\cdots(n)$, the permutation fixing every point.

Example: Cyclic Groups. The **cyclic group** of order n may be written as $C_n = \{e, x, x^2, \ldots, x^{n-1}\}$ where $x^n = e$, and $x^m = e$ iff $n \,\big|\, m$; see Exercise 1.3. In this example, the order of x is n, which is the same as the order of C_n .

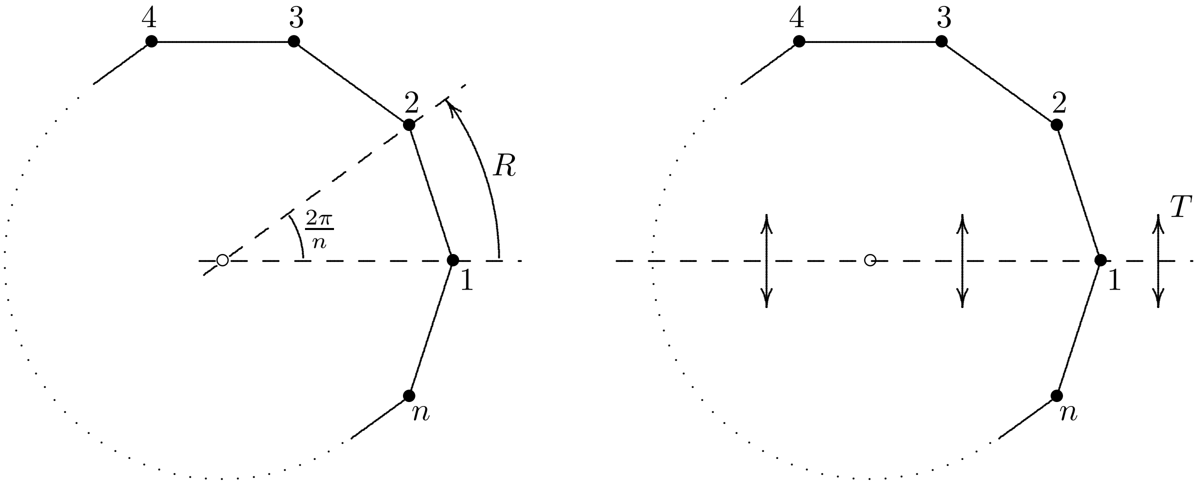
Example: Isometry Groups. An **isometry** of the Euclidean plane is a transformation $\mathbb{R}^2 \to \mathbb{R}^2$ which preserves distance. Isometries of the plane include rotations, reflections, translations and glide reflections. Examples of these are the maps

$$\begin{gathered}
(x, y) \mapsto (y, -x); \qquad (x, y) \mapsto (x, -y); \\
(x, y) \mapsto (x + 2, y - 3); \qquad (x, y) \mapsto (x + 1, -y)
\end{gathered}$$

respectively. The set of all isometries of the plane forms an infinite nonabelian group under composition. We may also consider the isometry group of Euclidean 3-space, an even larger nonabelian group.

Example: Dihedral Groups. The **dihedral group** of degree n and order $2n$ is the group D_n consisting of all symmetries of a regular n-gon in the plane. (*Warning:* Some books use the notation D_{2n} in place of D_n .) By a **symmetry** of an object in 2-space or in 3-space, we mean an isometry of the corresponding 2-space or 3-space which preserves the object.

Consider a regular n-gon in the plane with vertices labeled $1, 2, 3, \ldots, n$ in a counter-clockwise fashion as shown. This n-gon has n rotational symmetries $I, R, R^2, \ldots, R^{n-1}$ where R represents a counter-clockwise rotation through an angle of $\frac{2\pi}{n}$ about the center of the n-gon. It also has n reflective symmetries. The axes of these reflections consist of lines through the center of the n-gon, passing through the vertices of the n-gon and the midpoints of its sides. Let T be the reflective symmetry whose axis is the line joining vertex 1 with the center of the n-gon.



Every element of D_n has the form R^iT^j where $i \in \{0, 1, 2, \ldots, n-1\}$, $j \in \{0, 1\}$, and it is not hard to check that elements of D_n multiply according to the law

$$(R^iT^j)(R^kT^\ell) = \begin{cases} R^{i+k}T^\ell, & \text{if } j \text{ is even;} \\ R^{i-k}T^{\ell+1}, & \text{if } j \text{ is odd.} \end{cases}$$

Exercises 2.

1. List all orders of elements of S_5 , and the number of elements of each order. Check that the total number of elements equals $5! = 120$.

2. How many elements of S_6 commute with (12)? (Do not list them all!)

3. A **transposition** is a 2-cycle in S_n. For example, S_3 has exactly three transpositions: (12), (13) and (23). Observe that every 3-cycle in S_3 is a product of two transpositions: (123) = (12)(13), (132) = (13)(12). Show that every permutation in S_n is a product of at most $n-1$ transpositions. (These transpositions, however, are not necessarily disjoint!) This says that S_n is *generated by* its transpositions.

4. Let $n \geq 2$ be an integer. Consider the polynomial

$$f(X_1, X_2, \ldots, X_n) = \prod_{1 \leq i < j \leq n} (X_j - X_i),$$

which is homogeneous of degree $n(n-1)/2$. For any $\sigma \in S_n$, define a new polynomial by

$$f_\sigma(X_1, X_2, \ldots, X_n) = f(X_{1^\sigma}, X_{2^\sigma}, \ldots, X_{n^\sigma}).$$

For example if $n = 3$ then $f(X_1, X_2, X_3) = (X_2 - X_1)(X_3 - X_1)(X_3 - X_2)$ and $f_{(12)}(X_1, X_2, X_3) = f(X_2, X_1, X_3) = (X_1 - X_2)(X_3 - X_2)(X_3 - X_1)$ so that $f_{(12)} = -f$.

(a) For any $\sigma \in S_n$, show that $f_\sigma = \pm f$. Thus we may define the **sign** of σ, denoted $\operatorname{sgn}\sigma$, by

$$\operatorname{sgn} : S_n \to \{1, -1\}, \quad f_\sigma = (\operatorname{sgn}\sigma)f.$$

(b) If $\sigma \in S_n$ and $\tau \in S_n$ is a transposition, show that $\operatorname{sgn}(\sigma\tau) = -\operatorname{sgn}\sigma$. In particular for $\sigma = (1)$, this says that the sign of every transposition is -1.

(c) Using (b) and Exercise 3, prove that $\operatorname{sgn}(\sigma\pi) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\pi)$ for all $\sigma, \pi \in S_n$.

(d) Prove that the set $A_n = \{\sigma \in S_n : \operatorname{sgn}(\sigma) = 1\}$ is a group of order $n!/2$. This is called the **alternating group of degree** n.

(e) Show that any permutation $\sigma \in S_n$ is expressible either as a product of an even number of transpositions (in which case $\sigma \in A_n$), or as a product of an odd number of transpositions (in which case $\sigma \in S_n \smallsetminus A_n$); but that no permutation is expressible *both* as a product of an even number of transpositions, *and* as a product of an odd number of transpositions. So we call a permutation **even** if its sign is +1, or **odd** if its sign is −1.

(f) List all the permutations in A_4.

5. Show that every isometry of the Euclidean plane is the product of at most three reflections.

6. Show that every isometry of the Euclidean plane is (a) continuous, and (b) bijective.

7. Let G be the symmetry group of the solid in $\mathbb{R}^3$ consisting of all points (x, y, z) such that $|x| \leq 2$, $|y| \leq 2$, and $|z| \leq 0.1$. Say as much as you can about the structure of G, including its order, the number of elements of each order, whether or not G is abelian, and if possible relating G to any groups we have studied so far.

3. Isomorphism

If G and H are groups, then an **isomorphism** from G to H is a bijection $\phi : G \to H$ such that $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in G$. If there exists an isomorphism from G to H, then G and H are said to be **isomorphic**, and we write $G \cong H$. Clearly, group isomorphism is an equivalence relation. It is also clear that isomorphic groups have the same order, as well as the same number of elements of each order. In fact isomorphic groups have *all* the same abstract group-theoretical properties, so that two groups which are isomorphic are usually considered to be the same group.

For a finite group G, a **Cayley table** (or **group table** or **multiplication table**) is a square table specifying the products of all pairs of elements of the group. For example, if $G_1 = \{1, 3, 5, 7\}$ is the group whose operation is multiplication modulo 8, then a Cayley table for G_1 is given by

	1	3	5	7
1	1	3	5	7
3	3	1	7	5
5	5	7	1	3
7	7	5	3	1

Similarly, if G_2 is the group consisting of all diagonal 2×2 matrices with entries ± 1, under the usual matrix multiplication, then a Cayley table for G_2 is given by

	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$
$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$
$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$
$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$
$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}-1&0\\0&1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&-1\end{pmatrix}$	$\begin{pmatrix}1&0\\0&1\end{pmatrix}$

Observe that the Cayley table for G_1 becomes the Cayley table for G_2 if we replace $1 \mapsto \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, $3 \mapsto \left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)$, $5 \mapsto \left(\begin{smallmatrix}-1&0\\0&1\end{smallmatrix}\right)$, and $7 \mapsto \left(\begin{smallmatrix}-1&0\\0&-1\end{smallmatrix}\right)$ throughout. This defines an isomorphism from G_1 to G_2. In general, two finite groups G and H are isomorphic iff a group table for H may be obtained from a group table for G by renaming the elements appropriately, and possibly permuting rows and columns.

It is clear that given n, there are only finitely many groups of order n up to isomorphism; for if $G = \{x_1{=}e, x_2, \ldots, x_n\}$, then there are only finitely many ways to fill the n^2 entries of the table for G, and only a few of these possible tables can be expected to form groups. Clearly any row (or column) of a Cayley table for G must consist of all the elements $x_1{=}e, x_2, \ldots, x_n$ in *some* order. (Why?) We may therefore ask for a list of all groups of order n up to isomorphism.

If G is a group of order 2, say $G = \{e, x\}$, then there is only one way to complete the table

	e	x
e	e	x
x	x	

to a Cayley table, namely

	e	x
e	e	x
x	x	e

.

This shows that any group of order 2 is cyclic, i.e. isomorphic to C_2.

If G is a group of order 3, say $G = \{e, x, y\}$, then we start with a Cayley table for G as follows:

	e	x	y
e	e	x	y
x	x		
y	y		

The next row of the table fills in as either

	e	x	y
e	e	x	y
x	x	e	y
y	y		

or

	e	x	y
e	e	x	y
x	x	y	e
y	y		

.

But we cannot have two y's in the same column, so the table must complete uniquely as

	e	x	y
e	e	x	y
x	x	y	e
y	y	e	x

Thus $G = \{e, x, x^2{=}y\} \cong C_3$, i.e. every group of order 3 is cyclic.

How many isomorphism classes of groups are there of groups of order 4? Let $G = \{e, x, y, z\}$. Starting with

$$\begin{array}{c|cccc} & e & x & y & z \\ \hline e & e & x & y & z \\ x & x & & & \\ y & y & & & \\ z & z & & & \end{array} ,$$

the next row completes as either

$$\text{(i)}\ \begin{array}{c|cccc} & e & x & y & z \\ \hline e & e & x & y & z \\ x & x & e & z & y \\ y & y & & & \\ z & z & & & \end{array} , \qquad \text{(ii)}\ \begin{array}{c|cccc} & e & x & y & z \\ \hline e & e & x & y & z \\ x & x & y & z & e \\ y & y & & & \\ z & z & & & \end{array} , \qquad \text{or (iii)}\ \begin{array}{c|cccc} & e & x & y & z \\ \hline e & e & x & y & z \\ x & x & z & e & y \\ y & y & & & \\ z & z & & & \end{array} .$$

However, cases (ii) and (iii) are equivalent, since interchanging the last two columns of (iii), then interchanging the last two rows, then renaming $y \leftrightarrow z$, yields case (ii). So disregard case (iii). It is easy to see that (ii) leads uniquely to

$$\begin{array}{c|cccc} & e & x & y & z \\ \hline e & e & x & y & z \\ x & x & y & z & e \\ y & y & z & e & x \\ z & z & e & x & y \end{array} .$$

In this case $G = \{e, x, x^2{=}y, x^3{=}z\} \cong C_4$, i.e. G is cyclic. Case (i) may be completed to a group table in just two ways:

$$\text{(i.a)}\ \begin{array}{c|cccc} & e & x & y & z \\ \hline e & e & x & y & z \\ x & x & e & z & y \\ y & y & z & e & x \\ z & z & y & x & e \end{array} , \qquad \text{or (i.b)}\ \begin{array}{c|cccc} & e & x & y & z \\ \hline e & e & x & y & z \\ x & x & e & z & y \\ y & y & z & x & e \\ z & z & y & e & x \end{array} .$$

However, case (i.b) is $\{e, y, y^2{=}x, y^3{=}z\}$, which is cyclic, isomorphic to case (ii). Case (i.a) is *not* cyclic, since it has $g^2 = e$ for all $g \in G$. Case (i.a) is called the **Klein 4-group**. (Note that examples G_1 and G_2 above belong to this isomorphism class.) Thus there are exactly *two* groups of order 4 up to isomorphism: the cyclic group C_4, and the Klein 4-group.

Of course for infinite groups, or even large finite groups, tables are not helpful for specifying the group or checking isomorphism. As an example of an isomorphism of infinite groups, consider $\mathbb{R}$ under addition, and the positive real numbers $(0,\infty)$ under multiplication. Clearly $\log : (0,\infty) \to \mathbb{R}$ is an isomorphism; here $\log(xy) = \log(x) + \log(y)$ for all $x, y \in (0,\infty)$. (Here we have been careful to write '+' for the operation in $\mathbb{R}$, but multiplication for the operation in $(0,\infty)$.) However, $\mathbb{R}^\times = (-\infty,0) \cup (0,\infty)$ is *not* isomorphic to the additive group of $\mathbb{R}$. One way to see this is to note that $\mathbb{R}^\times$ has an element -1 of order two, whereas the additive group $\mathbb{R}$ has no element of finite order other than the identity.

Exercises 3.

1. Let G be a group of order 5. By considering possibilities for the Cayley table of G, prove that G is cyclic.

2. Let H and K be groups, with respective group operations. On the usual set-theoretic Cartesian product
$$H \times K = \{(h,k) : h \in H,\ k \in K\},$$
we define componentwise multiplication:
$$(h,k)(h',k') = (hh',kk') \in H \times K \quad \text{for } (h,k),(h',k') \in H \times K.$$
Show that this makes (H,K) into a group. This is called the **direct product** of H and K. Observe that the Klein 4-group is isomorphic to $C_2 \times C_2$, so that the only two groups of order 4 (up to isomorphism) are C_4 and $C_2 \times C_2$.

3. If $n = n_1 n_2 \cdots n_r$ where the positive integers $n_1, n_2, \ldots, n_r$ are mutually relatively prime (i.e. $\gcd(n_i,n_j) = 1$ whenever $i \neq j$), show that
$$C_n \cong C_{n_1} \times C_{n_2} \times \cdots \times C_{n_r}.$$

4. Show that the multiplicative group $\mathbb{R}^\times$ of nonzero real numbers, is isomorphic to the direct product of $(0,\infty)$ with a group of order 2.

5. Show that two cycles in S_n commute iff they are disjoint, or one is a power of the other.

6. Let $\sigma \in S_n$. Show that $|\sigma|$ is the least common multiple of the lengths of the disjoint cycles in σ.

7. Let $\mathbb{F}_p = \{0,1,2,\ldots,p-1\}$ be the field of integers modulo p, with addition and multiplication modulo p. Define the **general linear group over** $\mathbb{F}_p$ to be the multiplicative group of all invertible $n \times n$ matrices over $\mathbb{F}_p$.
   (a) Determine the order of $GL_n(\mathbb{F}_p)$.
   Hint: For a typical element $A \in GL_n(\mathbb{F}_p)$, consider how many possible choices there are for the first column of A, then how many choices for the second column of A, then how many choices for the third column of A, etc.
   (b) Show that $GL_2(\mathbb{F}_2) \cong S_3$.

8. Let $G = C_3 \times C_3 \times C_3$, and let H be the group of all 3×3 matrices of the form
$$\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}$$
with $a,b,c \in \mathbb{F}_3$. Then G and H are groups of order 27.
   (a) How many elements of each order does G have? How many elements of each order does H have?
   (b) Is H isomorphic to G? Explain.
   Hint: Is G abelian? Is H abelian?

4. Subgroups and Cosets

A subset $H \subseteq G$ is called a **subgroup** of G (denoted $H \leq G$) if H forms a group with respect to the binary operation on G, restricted to H. It is easy to see that H is a subgroup of G, iff $H \subseteq G$ is a *nonempty* subset which is closed under multiplication and inversion; or equivalently, $\emptyset \neq H \subseteq G$ and $xy^{-1} \in H$ whenever $x, y \in H$.

Examples of subgroups include the **trivial subgroup** $\{e\}$, often denoted simply by 1, and G itself. Also each $x \in G$ generates a cyclic subgroup, the set of all powers of x, denoted by

$$\langle x \rangle = \{x^k : k \in \mathbb{Z}\}.$$

If x is of finite order m, we simply have $\langle x \rangle = \{e, x, x^2, \ldots, x^{m-1}\} \cong C_m$. In particular, note that $|x| = |\langle x \rangle|$.

More generally, if $x_1, x_2, \ldots, x_r \in G$, then $\langle x_1, x_2, \ldots, x_r \rangle$ denotes the **subgroup generated by** $x_1, x_2, \ldots, x_r$; this is the set of all finite products formed by the **generators** $x_1, x_2, \ldots, x_r$ and their inverses $x_1^{-1}, x_2^{-1}, \ldots, x_r^{-1}$. This may also be described as the unique smallest subgroup of G containing $x_1, x_2, \ldots, x_r$; see Exercise 4.2.

The set of all subgroups of a finite group is often represented in a diagram showing their inclusions. For example, the subgroups of the cyclic group $C_{12} = \{e, x, x^2, \ldots, x^{11}\}$ (where $x^{12} = e$) are all of the form $\langle x^{12/d} \rangle \cong C_d$ for $d \mid 12$, and their inclusions are all represented by the diagram:

$$\begin{tikzcd}
 & & \langle x \rangle = C_{12} \arrow[dl, dash] \arrow[dr, dash] & \\
 & \langle x^2 \rangle \cong C_6 \arrow[dl, dash] \arrow[dr, dash] & & \langle x^3 \rangle \cong C_4 \arrow[dl, dash] \\
\langle x^4 \rangle \cong C_3 \arrow[dr, dash] & & \langle x^6 \rangle \cong C_2 \arrow[dl, dash] & \\
 & \langle e \rangle & &
\end{tikzcd}$$

For an arbitrary group G, we define the **center** of G as

$$Z(G) = \{z \in G : zx = xz \text{ for all } x \in G\}.$$

We check that $Z(G)$ is a subgroup of G, as follows. First, $ex = x = xe$ for all $x \in G$, so $e \in Z(G)$. If $z_1, z_2 \in Z(G)$, then

$$(z_1 z_2)x = z_1(z_2 x) = z_1(x z_2) = (z_1 x) z_2 = (x z_1) z_2 = x(z_1 z_2)$$

for all $x \in G$, so that $z_1 z_2 \in Z(G)$. Also $z_1 x = x z_1$ implies

$$z_1^{-1}x = z_1^{-1}(xz_1)z_1^{-1} = z_1^{-1}(z_1x)z_1^{-1} = xz_1^{-1}$$

for all $x \in G$, so $z_1^{-1} \in Z(G)$. This shows that $Z(G) \leq G$ as claimed. By definition, we have $Z(G) = G$ iff G is abelian.

If A and B are *subsets* of a group G, we define

$$AB = \{ab : a \in A,\ b \in B\}.$$

In particular for $H \leq G$ and $x \in G$, we have $Hx = \{hx : h \in H\}$, called the **right coset** of H containing x. It is easy to see that, for $x, y \in G$, $H \leq G$,

$$y \in Hx \iff yx^{-1} \in H \iff H = Hyx^{-1} \iff Hx = Hy \iff Hx \cap Hy \neq \emptyset.$$

(For example if $Hx \cap Hy \neq \emptyset$, then $h_1 x = h_2 y$ for some $h_1, h_2 \in H$, so that $yx^{-1} = h_1 h_2^{-1} \in H$, and the other conclusions follow similarly.) This means that the right cosets of H partition the elements of G. Furthermore for $x \in G$, the map

$$H \to Hx, \qquad h \mapsto hx$$

is clearly bijective. Especially this shows that if H is finite, then all right cosets of H have the same size, namely $|H|$. We define the **index** of H in G, denoted $[G : H]$, as the number of distinct right cosets of H in G. Since the right cosets of H have equal size and they partition the elements of G, we have

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

This proves

4.1 Lagrange's Theorem. If G is a finite group with subgroup H, then $|H|$ divides $|G|$.

The truth of Lagrange's Theorem is clear in the example of C_{12} detailed above. *Warning:* the 'converse' of Lagrange's Theorem is false! If $d \mid |G|$, it does not necessarily follow that G has a subgroup of order d; see Exercise 4.7.

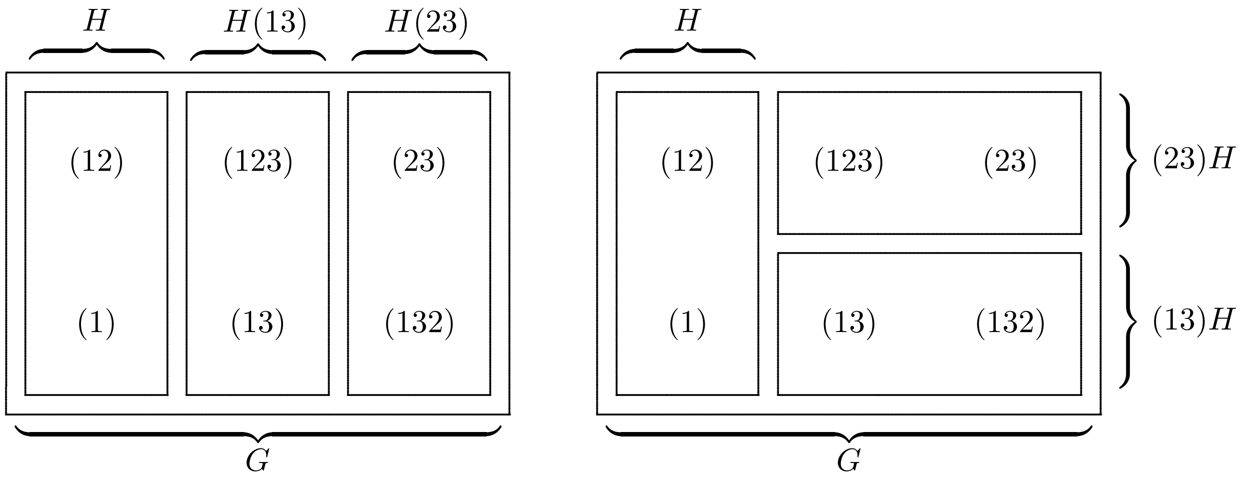
In the proof of Lagrange's Theorem, we partitioned G into right cosets Hx, $x \in G$. It is also true that G is partitioned into left cosets xH, $x \in G$. There are $[G : H]$ left cosets, each of size $|H|$, which gives another proof of Lagrange's Theorem. However, left cosets *do not* necessarily coincide with right cosets! For example if $G = S_3$ and $H = \langle (12) \rangle < G$, then the right cosets of H are

$$H = \{(1), (12)\}, \qquad H(13) = \{(13), (123)\}, \qquad H(23) = \{(23), (132)\}$$

whereas the left cosets are

$$H = \{(1),(12)\}, \qquad (13)H = \{(13),(132)\}, \qquad (23)H = \{(23),(123)\}.$$

The *number* of cosets is the same in each case (namely, $[G:H] = |G|/|H| = \frac{6}{2} = 3$) and the *size* of each coset is the same (namely $|H| = 2$), but the two partitions are *not* the same:



This motivates the definition of normal subgroup, in the next section.

Our formula $|G| = [G:H]|H|$ makes sense in the case of infinite groups, with the usual conventions (such as $m\infty = \infty$ whenever $1 \le m \le \infty$). An example of an infinite subgroup of finite index is given by the subgroup $(0,\infty)$ in the multiplicative group $\mathbb{R}^\times$. In this case the index $[\mathbb{R}^\times : (0,\infty)] = 2$ since there are just two cosets, namely $(0,\infty)$ and $(-\infty,0)$. Note that the subgroup $\langle -1 \rangle = \{1,-1\}$ of order 2 has infinite index in $\mathbb{R}^\times$, while the infinite subgroup $\mathbb{Q}^\times < \mathbb{R}^\times$ has infinite index (see Exercise 4.8).

4.2 Corollary. Let G be a finite group. Then every element of G has order dividing $|G|$.

Proof. Let $g \in G$. Then $|g| = |\langle g \rangle|$, which divides $|G|$ by Lagrange's Theorem. □

Exercises 4.

1. If $H, K \le G$, show that $H \cap K \le G$. More generally, if $\mathcal{S}$ is any nonempty collection of subgroups of G, show that
$$\bigcap \mathcal{S} = \bigcap_{H \in \mathcal{S}} H \le G.$$
2. Let S be any subset of a group G, and let $\langle S \rangle$ be the set of all products formed by elements of S and inverses of elements of S. (If $S = \varnothing$, then $\langle S \rangle = 1$.)
   (a) Show that $\langle S \rangle$ is a subgroup of G. (We call $\langle S \rangle$ the **subgroup generated by** S.)
   (b) Show that $\langle S \rangle$ is the intersection (see Exercise 4.1) of all subgroups of G containing S.

3. Let $H, K \leq G$. Recall that $HK = \{hk : h \in H,\ k \in K\}$. Is the product HK necessarily a subgroup? Prove this or give a counterexample.

4. If $H, K \leq G$ where G is a finite group, show that $|HK| = \frac{|H||K|}{|H \cap K|}$.

5. Consider a cyclic group $C_n = \langle g \rangle = \{e, g, g^2, \ldots, g^{n-1}\}$ here $g^n = e$. Prove that for every divisor $d|n$, C_n has exactly one subgroup of order d, namely the cyclic subgroup $\langle g^{n/d}$; and that these are the only subgroups of C_n.

6. Prove that every group of prime order is cyclic.

7. Prove that A_4 has no subgroup of order 6.

8. Let $G = \mathbb{R}^\times$, the multiplicative group of nonzero real numbers.
   (a) Show that the subgroup $(0, \infty)$ has index 2.
   (b) Show that the subgroup $\langle -1 \rangle$ has infinite index.
   (c) Show that the subgroup $\mathbb{Q}^\times$ has infinite index.

   Hint: Consider cardinalities.

9. Let G be a group. Define the **centralizer** of an arbitrary element $a \in G$ as $C_G(a) = \{g \in G : ga = ag\}$. Show that $C_G(a)$ is a subgroup of G.

10. Show that the center of $GL_2(\mathbb{C})$ consists of nonzero scalar multiples of the identity matrix.

11. Let G be a group of finite order n. By Corollary 4.2, we have $x^n = e$ for all $x \in G$. The *smallest* integer $k \geq 1$ such that $x^k = e$ for all $x \in G$, is called the **exponent** of G.
    (a) Prove that the exponent k divides n.

    Hint: Using the Division Algorithm, write $n = qk + r$ for some integers k, r where $0 \leq r < k$. Then use Corollary 4.2.

    (b) Suppose moreover that G is abelian. Prove that the exponent k is the maximum of $|x|$ for all $x \in G$. [This is not generally true for nonabelian groups, since S_3 has exponent 6, but its elements have orders 1, 2, 3 only.]

    Hint: Use Exercise 1.6.

12. Let G be a group. A *maximal subgroup* of G is a proper subgroup $M < G$ such that there is no subgroup strictly between M and G, i.e. if a subgroup H satisfies $M \leq H \leq G$ then either $H = M$ or $H = G$. The *Frattini subgroup* $\Phi(G) \leq G$ is defined as the intersection of all maximal subgroups of G (see Exercise 4.1).
    (a) Prove that $\Phi(G)$ is a normal subgroup of G.
    (b) Consider a subset $S \subseteq G$ such that $S \cup \Phi(G)$ generates G (see Exercise 4.2). Prove that S generates G. (Thus $Phi(G)$

5. Normal Subgroups and Quotient Groups

A subgroup $H \leq G$ is said to be a **normal subgroup** if $xH = Hx$ for all $x \in G$, i.e. if the partition of G into left cosets of H, coincides with the partition of G into right cosets of H. (*Warning:* This does *not* mean that $xh = hx$ for all $h \in H$.) We write $H \trianglelefteq G$ if H is a normal subgroup of G; or $H \triangleleft G$ if H is a normal *proper* subgroup of G. Note that $xH = Hx \iff x^{-1}Hx = x^{-1}xH = H$, by associativity. Furthermore if $x^{-1}Hx \subseteq H$ for all $x \in G$, then replacing x by x^{-1} gives $(x^{-1})^{-1}Hx^{-1} \subseteq H$, i.e. $H \subseteq x^{-1}Hx$. This proves

5.1 Proposition. A subgroup $H \le G$ is normal in G iff $x^{-1}Hx \subseteq H$ for all $x \in G$.

One nice thing about a normal subgroup $H \trianglelefteq G$ is that the product of two cosets of H is again a coset of H, since

$$(Hx)(Hy) = H(xH)y = H(Hx)y = (HH)(xy) = H(xy)$$

for all $x, y \in G$. (Note that $HH \subseteq H$ since H is closed under the group operation; conversely, $HH \supseteq H$ since $e \in H$.) Multiplication of cosets (or of any subset of G) is associative:

$$\big((Hx)(Hy)\big)(Hz) = (Hx)\big((Hy)(Hz)\big),$$

simply because multiplication of elements of G is associative. Also H acts as an 'identity' among cosets of H, since

$$H(Hx) = (HH)x = Hx, \qquad (Hx)H = H(xH) = H(Hx) = (HH)x = Hx$$

for all $x \in G$. Furthermore for any coset Hx we have

$$(Hx)(Hx^{-1}) = H(xx^{-1}) = He = H, \qquad (Hx^{-1})(Hx) = H(x^{-1}x) = He = H,$$

i.e. $H(x^{-1})$ acts as an inverse for Hx among the cosets of H. So whenever $H \trianglelefteq G$, we are led to define the **quotient group** (or **factor group**) of G modulo H, as the set of all cosets of H in G, under the operation of subset multiplication:

$$G/H = \{Hx : x \in G\}.$$

We have already shown that this is in fact a group of order $[G : H]$ with identity H and inverses $(Hx)^{-1} = H(x^{-1})$.

For example, consider $G = S_3$. We have already seen that the subgroup $\langle(12)\rangle$ is *not* normal in G. However, consider $H = \langle(123)\rangle$ of order 3. We have a partition of G into $[G : H] = \frac{6}{3} = 2$ left cosets of H, and likewise into 2 right cosets of H. The left cosets are in fact

$$H = \{(1), (123), (132)\}, \qquad (12)H = \{(12), (13), (23)\},$$

while the right cosets are

$$H = \{(1), (123), (132)\}, \qquad H(12) = \{(12), (13), (23)\}.$$

Since these two partitions are the same, we have $H \trianglelefteq G$. As an example of multiplication of two cosets of H, we have

$$\big(H(12)\big)\big(H(12)\big) = \{(12), (13), (23)\} \cdot \{(12), (13), (23)\} = \{(1), (123), (132)\} = H.$$

Thus $G/H = S_3/\langle(123)\rangle$ is a cyclic group of order 2, with Cayley table

$$\begin{array}{c|cc} & H & H(12) \\ \hline H & H & H(12) \\ H(12) & H(12) & H \end{array} \ .$$

The same argument shows that if G is *any* group having a subgroup H of index 2, then H is normal in G. This is because if $x \in G \smallsetminus H$ then the left cosets form a partition $\{H, xH\}$ of G, while the right cosets form a partition $\{H, Hx\}$ of G, so that the two partitions must coincide.

As further examples, note that $G \trianglelefteq G$ always, and G/G is the trivial group. Also the trivial subgroup $\langle e\rangle$ is normal in G, and the left (and right) cosets of $\langle e\rangle$ are the singleton subsets $\{x\}$ for $x \in G$ so that $G/\langle e\rangle \cong G$, where $\langle x\rangle \mapsto x$ is an obvious isomorphism. We say that a nontrivial group G is **simple** if its *only* normal subgroups are 1 and G itself. Cyclic groups of prime order are simple (see Exercises 4.6, 5.3), and there are also many nonabelian simple groups, of which the smallest is the alternating group A_5 of order 60 (see Exercise 9.7).

If G is an *abelian* group, then *every* subgroup of G is clearly normal. More generally, if every element of G commutes with every element of a subgroup $H \le G$ (equivalently, $H \le Z(G)$), then $H \trianglelefteq G$.

Exercises 5.

1. If $H, K \trianglelefteq G$, show that $H \cap K \trianglelefteq G$. More generally, if $\mathcal{S}$ is any nonempty collection of normal subgroups of G, show that
$$\bigcap \mathcal{S} = \bigcap_{H\in\mathcal{S}} H \trianglelefteq G.$$
Compare with Exercise 4.1.
2. If $H \le G$ and $K \trianglelefteq G$, show that $HK \le G$. If $H, K \trianglelefteq G$, show that $HK \trianglelefteq G$.
3. Prove that an abelian group is simple iff it is cyclic of prime order.
4. If $H \trianglelefteq K \trianglelefteq G$, does it necessarily follow that $H \trianglelefteq G$? Explain.
Hint: Consider the group A_4.

6. Homomorphisms

We have already defined group isomorphisms. If we drop the requirement of bijectivity, then we have simply a homomorphism of groups. If G and H are groups, then a map $\phi : G \to H$ is called a **homomorphism** if ϕ preserves the respective group operations, i.e. $\phi(xy) = \phi(x)\phi(y)$. We remind the reader that the two binary operations arising in the latter equation are interpreted quite differently: the operation between x and y takes place inside G, whereas the operation between $\phi(x)$ and $\phi(y)$ takes place inside H. To interpret the requirement of being a homomorphism in terms of group tables, is trickier than the interpretation for isomorphisms. [See however Gorenstein, *Scientific*

American (Dec. 1985), pp.104–115 for such an interpretation.] A homomorphism is called an **epimorphism** if it is surjective (that is, 'onto'); or a **monomorphism** if it is injective (that is, 'one-to-one'); and an **isomorphism** if it is bijective (that is, 'one-to-one' and 'onto').

As examples, note that every isomorphism is a homomorphism. Moreover if G is any group, then the unique map from G to 1 is a homomorphism, the so-called **trivial homomorphism**. A more interesting example is the map $\theta : G \to G$, $x \mapsto x^2$ which is a homomorphism iff G is abelian. To see this, note that the equation $xyxy = \theta(xy) = \theta(x)\theta(y) = x^2y^2$ simplifies to $xy = yx$ for all $x, y \in G$. Moreover θ need not be either bijective or trivial in this case; for example if $G = C_4 = \{e, x, x^2, x^3\}$ then $\theta(e) = \theta(x^2) = e$ and $\theta(x) = \theta(x^3) = x^2$.

Suppose that $\phi : G \to H$ is a homomorphism. Then $\phi(e) = \phi(ee) = \phi(e)\phi(e)$, so

$$\phi(e) = e.$$

Of course this means that $\phi(e_G) = e_H$ where e_G and e_H are the respective identities of G and H. Also

$$\phi(x^k) = \phi(x)^k$$

for all $x \in G$ and $k \in \mathbb{Z}$; for example $\phi(x^{-1})\phi(x) = \phi(x^{-1}x) = \phi(e) = e$ implies that $\phi(x^{-1}) = \phi(x)^{-1}$, and by induction on k we verify the above formula more generally.

Define the **kernel** and **image** of a homomorphism $\phi : G \to H$ by

$$\ker\phi = \{g \in G : \phi(g) = e\}, \qquad \phi(G) = \{\phi(g) : g \in G\}.$$

It is easy to show that $\ker\phi \leq G$ and $\phi(G) \leq H$. More significant is the fact that kernels of homomorphisms are always *normal* subgroups. (This is not true, however, for images.) To see that $\ker\phi \trianglelefteq G$, note that

$$\phi(x^{-1}gx) = \phi(x)^{-1}\phi(g)\phi(x) = \phi(x)^{-1}e\phi(x) = e$$

if $x \in G$ and $g \in \ker\phi$, so that $x^{-1}(\ker\phi)x \subseteq \ker\phi$ for all $x \in G$. Next we prove a converse of this fact: *Every normal subgroup is the kernel of some homomorphism.*

Given a group G and a normal subgroup $K \trianglelefteq G$, we have the **canonical homomorphism**

$$\pi : G \to G/K, \quad g \mapsto Kg.$$

This map π is onto by definition, and the fact that π is a homomorphism follows from $(Kx)(Ky) = Kxy$. Since K is the identity of G/K, we have

$$\ker\pi = \{g \in G : \pi(g) = K\} = \{g \in G : Kg = K\} = K.$$

Thus the given normal subgroup K is the kernel of a homomorphism (namely, the canonical homomorphism $\pi : G \to G/K$).

Now let $\phi : G \to H$ be any group homomorphism. Since $\ker\phi \trianglelefteq G$, we may form the quotient group $G/\ker\phi$. We show

6.1 First Isomorphism Theorem. $G/\ker\phi \cong \phi(G)$.

Proof. Let $K = \ker\phi \trianglelefteq G$, and define $\overline{\phi} : G/K \to \phi(G)$ by $\overline{\phi}(Kg) = \phi(g)$ for $g \in G$. We first observe that $\overline{\phi}$ is well-defined, for if $Kx = Ky$ for some $x, y \in G$, then $y = kx$ for some $k \in K$ and $\overline{\phi}(Ky) = \phi(y) = \phi(kx) = \phi(k)\phi(x) = e\phi(x) = \phi(x) = \overline{\phi}(Kx)$.

Now if $\overline{\phi}(Kx_1) = \overline{\phi}(Kx_2)$ for some $x_1, x_2 \in G$, then $\phi(x_1) = \phi(x_2)$, i.e. $\phi(x_1x_2^{-1}) = \phi(x_1)\phi(x_2)^{-1} = e$ so that $x_1x_2^{-1} \in \ker\phi = K$ and $Kx_1 = Kx_2$. This proves that $\overline{\phi}$ is one-to-one.

Clearly $\overline{\phi}$ is onto: every element of $\phi(G)$ is of the form $\phi(g) = Kg$ for some $g \in G$, and this is the same element as $\overline{\phi}(Kg)$. Thus $\overline{\phi}$ is bijective.

The fact that $\overline{\phi}$ is an isomorphism now follows from the fact that

$$\overline{\phi}\big((Kx)(Ky)\big) = \overline{\phi}(Kxy) = \phi(xy) = \phi(x)\phi(y) = \overline{\phi}(Kx)\overline{\phi}(Ky)$$

for all $x, y \in G$. □

Exercises 6.

1. Let $\phi : G \to H$ be a group homomorphism. Show that ϕ is one-to-one iff $\ker\phi = 1$.

2. Let $\phi : G \to H$ be a group homomorphism. Prove that for every $g \in G$ of finite order, the image $\phi(g)$ has finite order dividing $|g|$.

3. Let G be a group. Define $\phi : G \to G$, $x \mapsto x^2$ and define $\psi : G \to G$, $x \mapsto x^{-1}$. Prove that G is abelian iff ϕ is a homomorphism iff ψ is a homomorphism.

4. Let $G = GL_2(\mathbb{C})$, the multiplicative group of invertible 2×2 matrices with comples entries. Let $Z = Z(G)$. Recall (Exercise 4.10) that $Z = \{\alpha I : 0 \neq \alpha \in \mathbb{C}\}$ where $I \in G$ is the identity matrix. Since $Z \trianglelefteq G$, we may define the **projective general linear group**

$$PGL_2(\mathbb{C}) = G/Z.$$

Let $\mathbb{C}^* = \mathbb{C} \cup \{\infty\}$ be the **one-point compactification** of the complex plane. Let H be the set of all *fractional linear transformations* on $\mathbb{C}^*$, i.e. mappings of the form

$$\mathbb{C}^* \to \mathbb{C}^*, \quad z \mapsto \frac{az+b}{cz+d}$$

where $ad-bc \neq 0$. (Here $\frac{az+b}{cz+d}$ means the same as $\lim\limits_{w\to z} \frac{aw+b}{cw+d}$, in order to make these transformations defined, and in fact continuous, on $\mathbb{C}^*$.) Prove that H is a group under composition, and that

$$H \cong PGL_2(\mathbb{C}).$$

Hint: Use the First Isomorphism Theorem.

*5. Let G be a finite group, and suppose that $\phi : G \to G$ is a homomorphism such that

(i) $\phi(\phi(x)) = x$ for all $x \in G$; and

(ii) $\phi(x) = x$ only for $x = e$.

Show that G is abelian. Can you find a counterexample if G is infinite?

6. Let $K \trianglelefteq G$. If H is a subgroup of G containing K, then clearly $K \trianglelefteq H$ and we have $H/K \le G/K$.
   (a) Show that $H \mapsto H/K$ defines a bijection between $\mathcal{S}_1$ (the collection of all subgroups of G which contain K) and $\mathcal{S}_2$ (the collection of all subgroups of G/K).
   (b) For $H \in \mathcal{S}_1$, show that
   $$H \trianglelefteq G \quad \iff \quad H/K \trianglelefteq G/K.$$

7. Let H be a proper normal subgroup of G. We say that H is **maximal normal** in G if H is maximal among all proper normal subgroups of G, i.e. if $H \trianglelefteq K \triangleleft G$ implies $H = K$. Show that H is maximal normal in G iff G/H is simple.

   Hint: Use Exercise 6.6.

8. Let $H \le G$, $K \trianglelefteq G$. Recall that $HK \le G$; see Exercise 5.2.
   (a) Show that $H \cap K \trianglelefteq H$ and $K \trianglelefteq HK$. Thus we may construct the quotient groups $H/(H \cap K)$ and HK/K.
   (b) Prove the **Second Isomorphism Theorem** for groups:
   $$H/(H \cap K) \;\cong\; HK/K.$$
   Hint: Consider the map $\phi : H \to HK/K$, $h \mapsto hK$ and apply the First Isomorphism Theorem.

9. Suppose that $K \le H \trianglelefteq G$ and $K \trianglelefteq G$. By Exercise 6.6, we have $H/K \trianglelefteq G/K$.
   (a) Show that the map $\phi : G/K \to G/H$, $Kg \mapsto Hg$ is well-defined, and an epimorphism.
   (b) Prove the **Third Isomorphism Theorem** for groups:
   $$^{(G/K)}\!\big/_{(H/K)} \;\cong\; G/H$$
   by applying the First Isomorphism Theorem.

10. Let $K \le G$ be groups such that K is contained in the center of G. If G/K is cyclic, show that G is abelian.

11. (a) Give an example of an infinite group G and a homomorphism $\phi : G \to G$ which is one-to-one but not onto.
    (b) Give an example of an infinite group G and a homomorphism $\phi : G \to G$ which is onto but not one-to-one.

12. Let G be a group. Define the **commutator** of two elements $x, y \in G$ by
    $$[x, y] = x^{-1}y^{-1}xy.$$
    (a) Show that $[x, y] = e$ iff x and y commute.
    (b) Define the **derived subgroup** $G' \le G$ to be the subgroup generated by all commutators in G:
    $$G' = \langle [x, y] : x, y \in G \rangle.$$
    Show that $G' \trianglelefteq G$ and that G/G' is abelian.
    (c) Let $K \trianglelefteq G$. Show that $K \supseteq G'$ iff G/K is abelian. Thus G' is the *smallest* normal subgroup of G whose quotient group is abelian. (It makes sense to speak of the smallest such normal subgroup, because this is the same as the *intersection* of all normal subgroups $K \trianglelefteq G$ such that G/K is abelian, as in Exercise 4.1.)

13. Let G be a group. Define the **commutator** of two subgroups $H, K \le G$ to be the subgroup $[H, K] \le G$ generated by all commutators $[h, k]$ such that $h \in H$, $k \in K$. This generalizes the derived subgroup $G' = [G, G]$ considered in Exercise 6.12. Now suppose that $H, K \trianglelefteq G$ satisfy $H \cap K = 1$.
    (a) Show that every element of H commutes with every element of K.

    Hint: Show that $[H, K] \subseteq H \cap K$.

(b) Recall that $HK \trianglelefteq G$; see Exercise 5.2. Show that $HK \cong H \times K$.

Remarks: A subgroup of the form HK is known as an **internal direct product** of the subgroups H and K.

14. Let $n \geq 1$ and let p be prime. Consider the group $G = GL_n(\mathbb{F}_p)$ of invertible $n \times n$ matrices over the field $\mathbb{F}_p$ of order p; see Exercise 3.7. Recall that the determinant map $\det : G \to \mathbb{F}_p^\times$ is a homomorphism, where $\mathbb{F}_p^\times = \{1, 2, \dots, p{-}1\}$ is the multiplicative group of nonzero scalars.

(a) Show that $\det$ is surjective. Using the First Isomorphism Theorem, determine the order of the **special linear group** $SL_n(\mathbb{F}_p) = \{A \in G : \det(A) = 1\}$.

(b) Show that $GL_3(\mathbb{F}_5) \cong SL_3(\mathbb{F}_5) \times C_4$.

Hint: Use Exercise 6.13.

15. Let G be a finite abelian group and let p be a prime divisor of $|G|$. Show that G has an element of order p. (This is a special case of Cauchy's Theorem, Exercise 8.7.)

Hint: We may assume that G has a nontrivial proper subgroup H (why?) and by induction on $|G|$, we may assume that $p \not| |H|$. Show that G/H has an element of order p by induction, and use Exercise 6.2 to obtain an element of order p in G.

16. Let G be a finite abelian group, and let $g \in G$ of order equal to the exponent m of G. (See Exercise 4.11 where it was shown that such an element $g \in G$ necessarily exists.)

(a) If G is not cyclic, prove that there exists an element $k \in G$ of prime order such that $k \notin \langle g \rangle$.

Hint: The quotient group $G/\langle g \rangle$ has an element $k\langle g \rangle$ of prime order p; thus there exists $k \in G$ such that $k^p \in \langle g \rangle$ but $k \notin \langle g \rangle$. Since $(k^p)^{m/p} = h^m = e$ and the subgroup $\langle g \rangle$ has exactly m/p solutions of $x^p = e$, namely the elements of $\langle g^p \rangle$ (see Exercise 4.5) we have $k^p \in \langle g^p \rangle$, i.e. $k^p = g^{jp}$ for some integer j. Verify that $k' = kg^{-j}$ has the required properties.

(b) Prove that there exists a subgroup $H \leq G$ complementary to $\langle g \rangle$, i.e. $G = \langle g \rangle H$ with $\langle g \rangle \cap H = 1$, and thus $G \cong \langle g \rangle \times H$ by Exercise 6.13.

Hint: If G is not cyclic, there exists a nontrivial subgroup $K \leq G$ such that $K \cap \langle g \rangle = 1$ (why?). Denote the quotient group $\overline{G} = G/K$ and a typical element $\overline{x} = xK \in \overline{G}$. By induction $\overline{G} = \langle \overline{g} \rangle \overline{H}$ with $\langle \overline{g} \rangle \cap \overline{H} = 1$. Let $H \leq G$ be the preimage of $\overline{H}$ in G, as in Exercise 6.6.

17. Prove the *Fundamental Theorem of Finite Abelian Groups:* Every finite abelian group G is isomorphic to a direct product of cyclic groups.

Hint: Use Exercise 6.16.

7. Automorphisms

For any group G, we may consider the possibilities for an isomorphism from G to itself. Such a map is called an **automorphism** of G.

We have already seen that if G is abelian, then the map

$$\theta : G \to G, \quad x \mapsto x^2$$

is a homomorphism. In order for θ to be an automorphism of G, θ must be bijective. Assuming that G is finite, θ is bijective iff θ is one-to-one iff θ is onto. And this is equivalent to $\ker\theta = 1$ (see Exercise 6.1). All this says that for a finite group G, the map $x \mapsto x^2$ is an automorphism iff G is abelian with no elements of order two.

Given a group G, let $\operatorname{Aut} G$ denote the set of all automorphisms of G. Clearly $\operatorname{Aut} G$ contains the identity map $G \to G$, $x \mapsto x$. Now $\operatorname{Aut} G$ is a group under composition, called the **automorphism group** of G, or the **full automorphism group** of G (to distinguish it from other groups of automorphisms of G, i.e. subgroups of $\operatorname{Aut} G$).

A special set of automorphisms of G is the set of **inner automorphisms** $\operatorname{Inn} G = \{\psi_g : g \in G\}$. These automorphisms are defined by

$$\psi_g : G \to G, \quad x \mapsto x^{\psi_g} = g^{-1}xg.$$

(We will write x^{ψ_g} instead of $\psi_g(x)$, since ψ_g simply permutes the elements of G, and we wish to be consistent with our usual notation in placing permutations as superscripts.) If $g^{-1}xg = g^{-1}yg$, then cancelation yields $x = y$, so that ψ_g is one-to-one. Also, given $y \in G$, we have $(gyg^{-1})^{\psi_g} = y$, which shows that ψ_g is onto. For $x, y \in G$, we have

$$x^{\psi_g} y^{\psi_g} = (g^{-1}xg)(g^{-1}yg) = g^{-1}xyg = (xy)^{\psi_g},$$

so that ψ_g is a homomorphism, and so $\psi_g \in \operatorname{Aut} G$. Now

$$\left(x^{\psi_g}\right)^{\psi_h} = h^{-1}(g^{-1}xg)h = (gh)^{-1}x(gh) = x^{\psi_{gh}}$$

for all $x \in G$, and so $\psi_g\psi_h = \psi_{gh}$ for all $g, h \in G$. This means that $\operatorname{Inn} G$ is closed under composition. It is also easy to chow that $(\psi_g)^{-1} = \psi_{g^{-1}}$, so in particular $\operatorname{Inn} G$ is 'closed under inverses'. Together this shows that $\operatorname{Inn} G \leq \operatorname{Aut} G$. (Of course, if G is abelian, then ψ_g is the *identity* map $G \to G$, $x \mapsto x$. In this case $\operatorname{Inn} G$ is the trivial subgroup of $\operatorname{Aut} G$.)

More is true: we in fact have $\operatorname{Inn} G \trianglelefteq \operatorname{Aut} G$. To prove this, let $g \in G$, $\sigma \in \operatorname{Aut} G$. Then

$$x^{\sigma^{-1}\psi_g\sigma} = \left(\left(x^{\sigma^{-1}}\right)^{\psi_g}\right)^{\sigma} = \left(g^{-1}x^{\sigma^{-1}}g\right)^{\sigma} = \left(g^{-1}\right)^{\sigma}\left(x^{\sigma^{-1}}\right)^{\sigma}g^{\sigma} = \left(g^{\sigma}\right)^{-1}xg^{\sigma}$$

for all $x \in G$, i.e. $\sigma^{-1}\psi_g\sigma = \psi_{g^\sigma} \in \operatorname{Inn} G$. This shows that $\operatorname{Inn} G \trianglelefteq \operatorname{Aut} G$ as claimed.

Exercises 7.

1. Let G be a cyclic group of order n. Show that $\operatorname{Aut} G$ is abelian of order $\varphi(n)$, using the notation of Euler's function $\varphi(n) = |\{k \in \mathbb{Z} : 1 \leq k \leq n,\ \gcd(k, n) = 1\}|$.

   Hint: Let $G = \{e, x, x^2, \ldots, x^{n-1}\}$. Show that every automorphism of G has the form $\tau_k : G \to G$, $x^i \mapsto x^{ik}$ for some integer k relatively prime to n.

2. Determine the automorphism group of the Klein 4-group.

3. Let $G = C_p \times C_p \times \cdots \times C_p$ (n times). Determine $\operatorname{Aut} G$. (This is a group you have encountered previously in this course.)

   Remark: Compare with Exercises 3.7(b) and 7.2.

4. Let $G = SL_n(F)$ and consider the inverse-transpose map $\phi : g \mapsto g^{-T}$.

   (a) Prove that ϕ is an automorphism of G.

(b) Prove that the automorphism ϕ is inner, iff $n \le 2$.

5. A subgroup $H \le G$ is *characteristic* (denoted H *char* G) if $\phi(H) = H$ for every $\phi \in \mathrm{Aut}(G)$. Prove that

   (a) Every characteristic subgroup of G is normal.

   (b) The derived subgroup of G is characteristic (see Exercise 6.12).

6. A nontrivial group G is *characteristically simple* if its only characteristic subgroups are 1 and G itself. Prove that G is characteristically simple iff G is a direct product of isomorphic simple groups, i.e. $G \cong K^n = K \times K \times \cdots \times K$ (n times) for some simple group K and some positive integer n.

7. Let $H \trianglelefteq G$ be a *minimal normal subgroup,* i.e. H is a nontrivial normal subgroup for which the only normal subgroups of G contained in H, are 1 and H itself. Prove that every minimal normal subgroup of G is characteristic.

8. Permutation Groups and Actions

If $\mathcal{S}$ is any set, the **symmetric group on** $\mathcal{S}$ is by definition the set of all permutations of $\mathcal{S}$ (i.e. bijections $\mathcal{S} \to \mathcal{S}$), denoted $\mathrm{Sym}\,\mathcal{S}$. This forms a group under composition. We will generally assume that $|\mathcal{S}| = n < \infty$, in which case $\mathrm{Sym}\,\mathcal{S}$ is clearly isomorphic to S_n . A **permutation group on** $\mathcal{S}$ is by definition a subgroup $G \le \mathrm{Sym}\,\mathcal{S}$. In this case we say that G is a permutation group of **degree** n. The **stabilizer** of an element $a \in \mathcal{S}$ is by definition

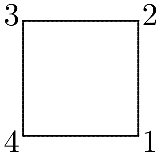
$$G_a = \{g \in G : a^g = a\}.$$

It is not hard to verify that G_a is in fact a *subgroup* of G. The **orbit** of $a \in \mathcal{S}$ is by definition the set of all possible images of a under G:

$$a^G = \{a^g : g \in G\}.$$

If $\mathcal{S}$ consists of a *single* orbit under G, we say that G is **transitive** on $\mathcal{S}$, or that G **permutes** $\mathcal{S}$ **transitively**; otherwise G is **intransitive**.

As a first example, consider the group $G = \langle(1234), (24)\rangle$, a dihedral group of order 8. This is a permutation group of degree 4 which may be considered as the symmetry group of the square



In this case the stabilizer of the vertex 1 is $G_1 = \langle(24)\rangle = \{(1), (24)\}$, and the orbit of the vertex 1 is $1^G = \{1, 2, 3, 4\}$. We observe that $|G_1||1^G| = 2\cdot 4 = 8 = |G|$, as evidence of a property that holds more generally; see Proposition 8.1 below. The subgroup $H = \langle(24)\rangle$ has three orbits on vertices:

$$1^H = \{1\}; \qquad 2^H = \{2, 4\} = 4^H; \qquad 3^H = \{3\}.$$

Observe that the orbits of H partition the vertex set $\{1,2,3,4\}$ into three orbits, each of size dividing $|H| = 2$. We have $|H_1||1^H| = |H||\{1\}| = 2\cdot 1 = 2 = |H|$; also $|H_2||2^H| = |\langle(1)\rangle||\{2,4\}| = 1\cdot 2 = 2 = |H|$. Note that G permutes $\{1,2,3,4\}$ transitively; the subgroup H is intransitive.

In our example above, the vertices played a special role. Our original abstract definition of a dihedral group, in Section 2, gave no special heed to vertices of the n-gon. The dihedral group just as well permutes edges of the n-gon, or diagonals, or points of the entire Euclidean plane. In our next example we consider some of these other actions of the dihedral group.

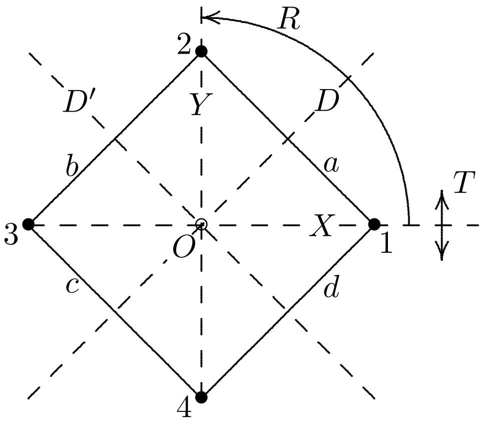
A **permutation action** (or **permutation representation**) of a group G on a set $\mathcal{S}$ is a homomorphism $\phi : G \to \mathrm{Sym}\,\mathcal{S}$. This associates, to each $g \in G$, a permutation $\phi(g) \in \mathrm{Sym}\,\mathcal{S}$. If ϕ is one-to-one, we say that the action is **faithful**; in this case $\phi(G) \cong G$, so we may identify G with the permutation group $\phi(G) \le \mathrm{Sym}\,\mathcal{S}$. However, if ϕ is not one-to-one, then distinct elements $g, h \in G$ may give rise to the same permutation $\phi(g) = \phi(h) \in \mathrm{Sym}\,\mathcal{S}$; in this case we say ϕ is **unfaithful**.

As an example, we again consider the dihedral group $G = D_4$ of order 8, but this time using our original notation of Section 2, thus: $G = \{I, R, R^2, R^3, T, RT, R^2T, R^3T\}$. We consider five actions of this group:

μ, the action of G on the vertex set $\mathcal{V} = \{1,2,3,4\}$;
σ, the action of G on the set of edges $\mathcal{E} = \{a,b,c,d\}$;
δ, the action of G on the set of diagonals $\mathcal{D} = \{D, D'\}$;
τ, the action of G on the set of coordinate axes $\mathcal{A} = \{X, Y\}$; and
γ, the action of G on the set consisting of the center $\mathcal{O} = \{O\}$.

One verifies that these five actions are as listed in the following table, in which () denotes the identity permutation:

g	$\mu(g)$	$\sigma(g)$	$\delta(g)$	$\tau(g)$	$\gamma(g)$
I	(1)	()	()	()	()
R	(1234)	$(abcd)$	$(D\,D')$	$(X\,Y)$	()
R^2	(13)(24)	$(ac)(bd)$	()	()	()
R^3	(1432)	$(adcb)$	$(D\,D')$	$(X\,Y)$	()
T	(24)	$(ad)(bc)$	$(D\,D')$	()	()
RT	(14)(23)	(ac)	()	$(X\,Y)$	()
R^2T	(13)	$(ab)(cd)$	$(D\,D')$	()	()
R^3T	(12)(34)	(bd)	()	$(X\,Y)$	()



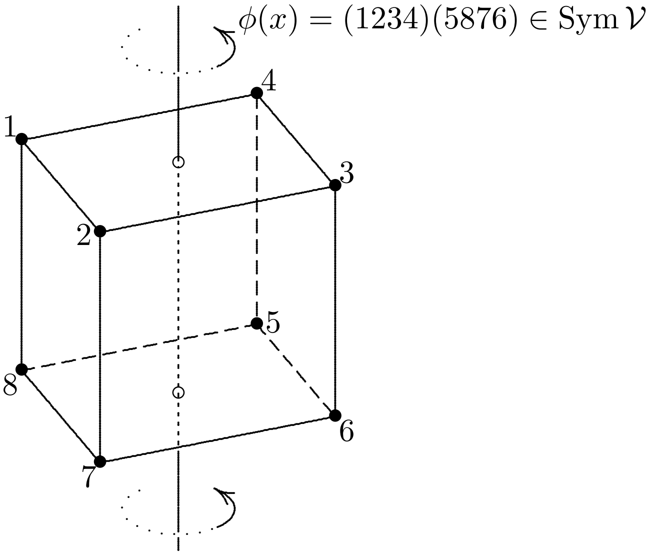
We readily verify that each of these actions *is* a homomorphism. It is natural to require this property; for example the fact that $\mu(gh) = \mu(g)\mu(h)$ says that the action of gh on the vertex set $\mathcal{V}$, is the same as first applying g to the vertices, then applying h to the vertices.

Note that the actions μ and σ are *faithful:* distinct elements of G give distinct permutations of the vertices, and of the edges. The other actions listed are *unfaithful:* distinct elements of G *may* give the same permutations on $\mathcal{D}$, $\mathcal{A}$ and $\mathcal{O}$. It is possible to identify G with its image $\mu(G) \le \operatorname{Sym}\mathcal{V}$; certainly this permutation group is isomorphic to G, via the isomorphism μ. There is nothing special however about vertices; one may equally well identify G with the permutation group $\sigma(G) \le \operatorname{Sym}\mathcal{E}$. However, the groups $\delta(G)$, $\tau(G)$, $\gamma(G)$ have orders $2, 2, 1$ respectively; these groups cannot be identified with G. Indeed, the action γ is *trivial:* every element of G fixes the center.

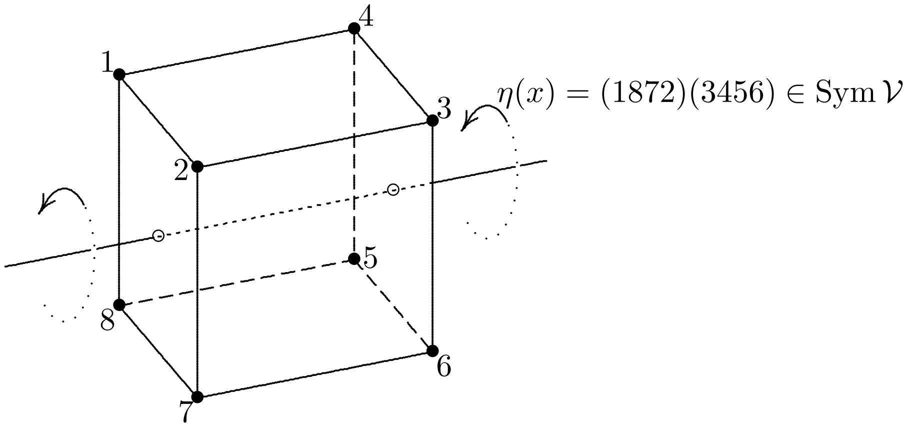
We speak of orbits and stabilizers not only for permutation groups, but more generally for permutation actions. In the example above, G acts transitively on $\mathcal{D}$, giving an orbit $D^G = \{D, D'\}$. The stabilizer of D is the subgroup $G_D = \{I, R^2, RT, R^3T\}$ and we check that $|G_D||D^G| = 4 \cdot 2 = 8 = |G|$.

In our previous example, $\mu, \sigma, \delta, \tau, \gamma$ were five actions of the group G on *distinct* sets $\mathcal{V}, \mathcal{E}, \mathcal{D}, \mathcal{A}, \mathcal{O}$. Next we consider a situation where a single group G can act on *one* set in more than one way.

Let $G = C_4 = \{e, x, x^2, x^3\}$, a cyclic group of order 4 with $x^4 = e$. We can let G act on a cube as a group of rotational symmetries. One natural way to do this would be to let x act as a 90° counter-clockwise rotation about a vertical axis, as shown. We have labeled the vertices of the cube as $\mathcal{V} = \{1, 2, \dots, 8\}$.



Clearly G permutes the vertices intransitively in this case; there are two orbits, $\{1, 2, 3, 4\}$ and $\{5, 6, 7, 8\}$, each of length 4. Each vertex has trivial stabilizer in G, and we check that $|1_G||1^G| = 1 \cdot 4 = 4 = |G|$. However, our choice of action of G on the cube was rather arbitrary; we could instead let x act as a 90° rotation about a different axis, giving rise to the action $\eta : G \to \operatorname{Sym}\mathcal{V}$ indicated:



As an example of an unfaithful representation, consider the action $\psi : G \to \operatorname{Sym}\mathcal{V}$ in which $\psi(x)$ is a reflection in a horizontal plane through the cube's center; in this case

$$\psi(x) = (18)(27)(36)(45).$$

In such cases where more than one action of G on a given set $\mathcal{S}$ is considered, it is ambiguous to write a^g for $a \in \mathcal{S}$ and $g \in G$; we must explicitly denote the action as, for example, $a^{\phi(g)}$. In our example of C_4 acting on $\mathcal{V}$, we have $1^{\phi(x)} = 2$, whereas $1^{\eta(x)} = 1^{\psi(x)} = 8$. The orbit depends on the choice of action; in our example, $1^{\phi(G)} = \{1, 2, 3, 4\}$, $1^{\eta(G)} = \{1, 2, 7, 8\}$, $1^{\psi(x)} = \{1, 8\}$. Moreover the stabilizer of a point $a \in \mathcal{S}$ may in general depend on the choice of action. While these examples may seem somewhat contrived, there are important examples in which a group G can act in two or more ways on the same set. In particular, we will see how every group G can act *on itself* by right-multiplication *or* by conjugation. (Yet another action, by left-multiplication, can be considered.)

Finally, we prove

8.1 Proposition. $|a^G| = [G : G_a] = |G|/|G_a|$.

Proof. Write $G_a\backslash G = \{G_a g : g \in G\}$, which is just the *set* of right cosets of G_a in G. (*Warning:* This is *not* a group unless G_a is normal in G.) Of course $|G_a\backslash G| = [G : G_a] = |G|/|G_a|$, so it is sufficient to find a bijection between $G_a\backslash G$ and a^G. So we will show that, given $g, h \in G$,

$$G_a g = G_a h \iff a^g = a^h.$$

If $G_a g = G_a h$, then $g = xh$ for some $x \in G_a$, i.e. $a^x = a$, in which case $a^g = a^{xh} = (a^x)^h = a^h$. Conversely, if $a^g = a^h$, then $a^{gh^{-1}} = a$, so $gh^{-1} \in G_a$, which implies that $G_a g = G_a h$.

It follows that $G_a g \mapsto a^g$ is a bijection $G_a\backslash G \to a^G$. □

We next show the remarkable fact that *every* finite group may be faithfully represented as a permutation group!

8.2 Cayley Representation Theorem. If G is a finite group, then G is isomorphic to a permutation group (i.e. a subgroup of S_n for some n).

Proof. We let G act on itself by right-multiplication, as follows. For $g \in G$, we define

$$\rho_g : G \to G, \quad x \mapsto xg.$$

Clearly this is a bijection on G, so we have $\rho_g \in \operatorname{Sym} G$. If $g, h \in G$, then

$$(x^{\rho_g})^{\rho_h} = (xg)h = x(gh) = x^{\rho_{gh}}$$

for all $x \in G$, i.e. $\rho_g \rho_h = \rho_{gh}$, and so we have a homomorphism

$$\rho : G \to \operatorname{Sym} G, \quad g \mapsto \rho_g.$$

Thus ρ is a representation (called the **right regular representation**) of G. If ρ_g fixes every element of G, then $e = e^{\rho_g} = eg = g$, so ρ is one-to-one (i.e. the representation is faithful), and so ρ gives an isomorphism from G to a subgroup of $\operatorname{Sym} G$. If $|G| = n$, this is just as good as an isomorphism from G to a subgroup of S_n . □

One nice thing about this result is that permutations are somewhat familiar objects which multiply in a predictable way. They are also quite recognizable to a computer. If we have a computer program designed to accept a permutation group as input data, then such a program can be used to handle *any* finite group G, if G is first represented as a permutation group. However, in order for this permutation representation to be useful in a program, the *degree* of the representation should not be too large. In other words, we want to write G as a group of permutations of $\{1, 2, \dots, n\}$ where n is not too large. The problem with our proof of the Cayley Representation Theorem is that it realizes G as a subgroup of S_n for a typically large value of n, namely $n = |G|$. It is often possible to improve on this, as we shall see later. But just as an example of how significant this point really is, note that Cayley's Theorem manages to express S_5 as a set of 120 permutations of $\{1, 2, 3, \dots, 120\}$; clearly the natural representation of S_5 as a set of 120 permutations of $\{1, 2, 3, 4, 5\}$ is much more concise!

There is a way to generalize the basic idea in the proof above, as follows. Let H be any subgroup of G. Then G permutes the right cosets of H by right multiplication. Writing the set of right cosets of H as $H\backslash G = \{Hx : x \in G\}$ (again, this is merely a set, not a group in general), then $|H\backslash G| = [G : H] = |G|/|H|$. For $g \in G$ we define

$$\rho_g : (H\backslash G) \to (H\backslash G), \quad Hx \mapsto Hxg.$$

Clearly this gives a permutation of the right cosets of H, i.e. $\rho_g \in \mathrm{Sym}(H\backslash G)$. (If we had the *trivial* subgroup $H = \{e\}$, then this would be the same action as in the proof of the Cayley Representation Theorem.) Moreover G permutes these cosets transitively! And we have a permutation representation of G of degree $[G : H]$. If we can choose a *large* subgroup $H < G$, and thereby make $[G : H]$ correspondingly small, then we will have succeeded in finding a smaller degree permutation representation of G than that realized by our proof of Theorem 8.2 above. This is good news. The bad news, however, is that this action might not be faithful. Shortly, however, we will find an explicit expression for $\ker \rho$.

Exercises 8.

1. Let H be the subgroup of S_9 generated by (123)(789) and (345). How many orbits does H have on $\{1, 2, 3, \ldots, 9\}$, and what are their sizes?

2. Suppose that $H \le G \le S_n$ where H acts transitively on $\{1, 2, 3, \ldots, n\}$. Show that $G = G_1H$ where $G_1 = \{g \in G : 1^g = 1\}$.

3. (a) Let G be the rotational symmetry group of a cube. What is the order of G? Is G abelian? What else can you say about G? Can you identify G up to isomorphism, as a group we have studied?
   (b) Do the same for a regular octahedron in place of the cube. A regular octahedron has eight triangular faces, all equilateral of the same size.

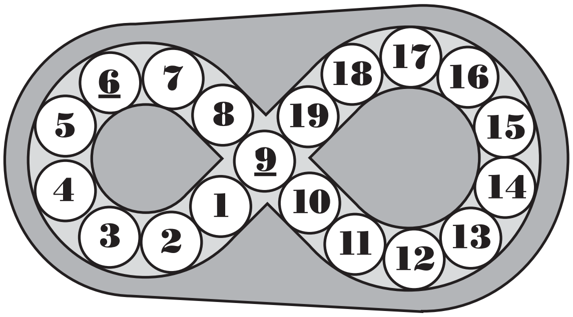
4. Suppose that G acts on $\mathcal{S}$. Let H be any *normal* subgroup of G, and let $\mathcal{S}_H = \{a \in \mathcal{S} : a^h = a$ for all $h \in H\}$, the set of fixed points of H. Prove that G preserves $\mathcal{S}_H$ (i.e. $a^g \in \mathcal{S}_H$ whenever $a \in \mathcal{S}_H$) and so G induces a permutation group on the subset $\mathcal{S}_H \subseteq \mathcal{S}$.

   This latter exercise is so important that we must give an example: Label the vertices of a regular octahedron as 1,2,3,4,5,6 in such a way that vertices 1,2,3,4 form a square. The group of all rotational symmetries of the regular octahedron (Exercise 8.3) has a subgroup $G \cong D_4$ preserving a square formed by the vertices 1,2,3,4. The cyclic subgroup $H \trianglelefteq G$ of order 4 cycles 1,2,3,4 and fixes the remaining two vertices 5,6. Therefore G must preserve $\mathcal{S}_H = \{5, 6\}$. You can see directly that in fact G does permute the pair $\{5, 6\}$ of antipodal vertices.

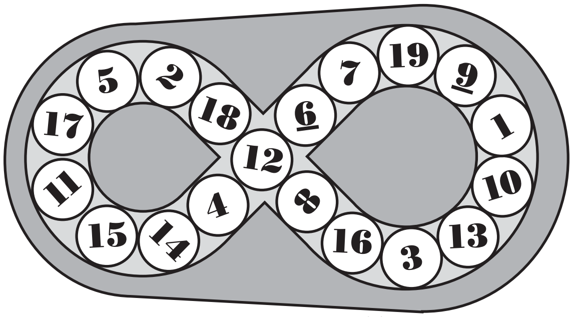
5. Let p be prime, and recall the construction of the group $GL_n(\mathbb{F}_p)$ in Exercise 3.7.
   (a) Show that for every $n \ge 1$, G has a subgroup isomorphic to S_n consisting of permutation matrices (matrices in which every row and column has a single nonzero entry equal to 1).
   (b) Show that every finite group is isomorphic to a subgroup of $GL_n(\mathbb{F}_p)$ for some $n \ge 1$.

   Hint: Apply Theorem 8.2.

6. Consider the puzzle depicted in the illustrations below, in which nineteen circular disks (numbered 1,2,. . . ,19) are free to slide around the two loops of a track shaped in a figure '8'. After sliding disks around one of the loops, one disk must come to rest at the center position (where the two loops meet) before sliding disks around the other loop. The first illustration shows the original configuration, according to the manufacturer's instructions. The second illustration shows the positions of the disks after the puzzle has been played with extensively. I have been trying, with no success, to restore the puzzle to its original configuration, and I am beginning to suspect that my child has dropped the puzzle on the floor, causing some disks to come loose from their tracks, then pushed the loose disks back into their tracks in a configuration which cannot be solved (i.e. restored to the original configuration using only legal moves).



Puzzle: Original Position



Puzzle: Altered Position

(a) Do you think it is possible to restore the puzzle from the altered position shown, to its original position? Explain.

(b) If some disks pop out and are pushed back into their tracks in a random configuration, what do you think is the probability that the resulting configuration is solvable (i.e. can be restored to the original configuration using only legal moves)? Explain.

Hint: Recall Exercise 2.4.

7. Let G be a finite group of order divisible by a prime p. Prove that G has an element of order p. (This is *Cauchy's Theorem;* compare Exercise 6.15 where we settled the case G is abelian.)

Hint: Let $\mathcal{S}$ be the set of all p-tuples $(g_1, g_2, \ldots, g_p)$ of elements of G satisfying $g_1 g_2 \cdots g_p = 1$, and define $\sigma : (g_1, g_2, \ldots, g_p) \mapsto (g_2, g_3, \ldots, g_p, g_1)$. Show that $\mathcal{S}$ is invariant under σ and that σ permutes $\mathcal{S}$ in orbits of size 1 and p. Determine $|\mathcal{S}|$ and observe that $\sigma \in \operatorname{Sym}\mathcal{S}$ fixes $(1, 1, \ldots, 1) \in \mathcal{S}$.

9. Conjugation

Let G be any finite group. Given $g, h \in G$, we say that g is **conjugate** to h in G (written $g \sim h$) if $g = x^{-1}hx$ for some $x \in G$. Note that $g = e^{-1}ge$, so that every group element is conjugate to itself. Also if $g \sim h$, say $g = x^{-1}hx$, then $h = (x^{-1})^{-1}g(x^{-1})$, so $h \sim g$. Furthermore if $g \sim h \sim k$, say $g = x^{-1}hx$ and $h = y^{-1}ky$, then $g = x^{-1}y^{-1}kyx = (xy)^{-1}k(xy)$, so that $g \sim k$. This shows that conjugacy is an equivalence relation on the elements of G. The equivalence classes for this relation are called the **conjugacy classes** of G. In particular, the conjugacy classes partition the elements of G.

For example in S_3, we have $(23) = (123)^{-1}(12)(123)$ and so $(23) \sim (12)$. Indeed the conjugacy classes of S_3 are

$$\{(1)\}, \qquad \{(12), (13), (23)\}, \qquad \{(123), (132)\}.$$

Note that the size of each of these conjugacy classes divides $|S_3| = 6$. This is no accident, as we shall soon see.

Recall that the inner automorphism $\psi_x \in \operatorname{Inn} G$ is defined by $\psi_x : g \mapsto g^{\psi_x} = x^{-1}gx$. Recall also that $\psi_x\psi_y = \psi_{xy}$ for all $x, y \in G$. Thus we have a homomorphism

$$\psi : G \to \operatorname{Inn} G, \quad x \mapsto \psi_x\,,$$

which is in fact a permutation representation of G, called the **conjugation action** of G on itself. Clearly, the orbits under this action are precisely the conjugacy classes of G. The stabilizer of an element $g \in G$ is just the *centralizer* of g in G:

$$\{x \in G : g^{\psi_x} = g\} = \{x \in G : x^{-1}gx = g\} = \{x \in G : gx = xg\} = C_G(g);$$

see Exercise 4.9. So the conjugacy class of an element $g \in G$ is the orbit $g^G = g^{\psi(G)}$, which has length $[G : C_G(g)] = |G|/|C_G(g)|$. In particular, the size of each conjugacy class divides $|G|$. Returning to our S_3 example, we have $C_{S_3}\big((123)\big) = \langle(123)\rangle = \{(1), (123), (132)\}$ of order 3, and the corresponding conjugacy class $\{(123), (132)\}$ has size $[S_3 : C_{S_3}((123))] = \frac{6}{3} = 2$.

So far we have used two actions of G on the elements of G: the right regular representation ρ, and the conjugation action ψ. If we were to subsequently write x^g, you might well wonder whether we meant $x^{\rho_g} = xg$ or $x^{\psi_g} = g^{-1}xg$. In fact we will typically mean the latter:

$$x^g = g^{-1}xg.$$

Just as we talk about conjugate *elements* of G, we can talk about *conjugate subgroups.* If H is any subgroup of G, and $x \in G$, then

$$H^x = x^{-1}Hx = \{x^{-1}hx : h \in H\}$$

is also a subgroup of G (since it is the image of H under ψ_x, and the image of any group homomorphism is a subgroup.) Again, G acts on the set of subgroups of G, by conjugation. The set of conjugates of H will be one orbit under this action. What is the length of this orbit? That is, how many distinct conjugates does H have inside G? The stabilizer of H in this action is otherwise known as the **normalizer** of H in G, written

$$N_G(H) = \{x \in G : H^x = H\} = \{x \in G : x^{-1}Hx = H\} = \{x \in G : Hx = xH\}.$$

Observe that $H \trianglelefteq N_G(H) \leq G$, and that $N_G(H) = G$ iff $H \trianglelefteq G$. Now the number of subgroups of G which are conjugate to H is $[G : N_G(H)] = |G|/|N_G(H)|$, and in particular this number divides $|G|$.

For example consider $\langle(123)\rangle < S_4$. Then S_4 has exactly four subgroups conjugate to $\langle(123)\rangle$, namely

$$\langle(123)\rangle, \quad \langle(124)\rangle, \quad \langle(134)\rangle, \quad \langle(234)\rangle.$$

This agrees with our formula, since $N_{S_4}\big(\langle(123)\rangle\big) = \langle(123), (12)\rangle = S_3$ of order 6, which has index 4 in S_4 .

Next, we show how conjugate subgroups arise naturally in the study of permutation groups: *points in the same orbit have conjugate stabilizers.* For suppose that G acts transitively on a set $\mathcal{S}$, i.e. $\phi : G \to \operatorname{Sym}\mathcal{S}$ is a transitive permutation action. We have seen that $|\mathcal{S}| = |a^G| = [G : G_a]$ for $a \in \mathcal{S}$. If $b \in \mathcal{S}$ also, then by transitivity we have $b = a^g$ for some $g \in G$. We claim that $G_b = (G_a)^g = g^{-1}G_ag$. Firstly, if $x \in G_a$, then

$$b^{g^{-1}xg} = ((b^{g^{-1}})^x)^g = (a^x)^g = a^g = b,$$

which shows that $g^{-1}xg \in G_b$. The reverse inclusion $G_b \subseteq g^{-1}G_ag$ is equivalent to $gG_bg^{-1} \subseteq G_a$, which follows by the same reasoning since $a = b^{g^{-1}}$. Thus $G_b = g^{-1}G_ag$ as claimed. Our superscript notation makes everything easy to remember: the stabilizer of a is G_a, so the stabilizer of a^g is $(G_a)^g$.

Exercises 9.

1. Let $H \le G$, and let $x, y \in H$. If x and y are conjugate in H, must they be conjugate in G? If they are conjugate in G, must they be conjugate in H? In each case, prove the statement or supply a counterexample.

2. (a) List the conjugacy classes of S_4.

   (b) Show that a subgroup $H \le G$ is normal iff H is a union of conjugacy classes of G (i.e. $H = \bigcup \mathcal{S}$ where $\mathcal{S}$ is a collection of conjugacy classes of G).

   (c) Using (b), determine all normal subgroups of S_4.

3. Consider the permutations $\sigma = (1942)(35)(68)$ and $\tau = (137)(28)$ in S_9.

   (a) Determine $\sigma^{-1}\tau\sigma$ and $(1^\sigma\, 3^\sigma\, 7^\sigma)(2^\sigma\, 8^\sigma)$. What do you observe?

   (b) Generalize your observation to show that two permutations in S_n are conjugate iff they have the same **cycle structure**, i.e. the same number of disjoint cycles of each length.

4. Let G be a group of order $2n$ where $n \ge 1$ is odd. Prove that G has a subgroup of order n. (Do *not* use the Sylow Theorems!)

   Hint: Let $\rho : G \to \operatorname{Sym} G$ be the right regular representation of G. By Exercise 1.7, there exists $\tau \in G$ of order 2. Show that the permutation ρ_τ is odd, by considering its cycle structure; then consider the kernel of the homomorphism $G \to \{\pm 1\},\ g \mapsto \operatorname{sgn}(\rho_g)$.

5. Let G be any group, and let $a \in G$. Show that $a \in Z(G)$ iff $\{a\}$ is a conjugacy class of G. Conclude that $Z(G)$ is the union of all singleton conjugacy classes of G.

6. Two elements $A, B \in GL_n(\mathbb{C})$ are conjugate iff they are similar as matrices.

   (a) List all properties that you can think of, shared by two similar matrices. (For example, similar matrices have the same trace.)

   (b) How does one check whether or not two given matrices are similar? Give necessary and sufficient conditions similar to the criterion for conjugacy of permutations, given in Exercise 9.3.

7. Complete the following steps (a)–(f) to show that A_n is a simple group for $n \ge 5$. Assume that $n \ge 5$ and suppose $1 \ne K \trianglelefteq A_n$. We must show that $K = A_n$.

   (a) Show that A_n is generated by its 3-cycles.

   Hint: First express an arbitrary product of two 2-cycles as a product of at most two 3-cycles, not necessarily disjoint.

   (b) Show that if K contains a 3-cycle, then $K = A_n$.

   Hint: Recall Exercise 9.3.

   (c) If $\sigma = (i_1i_2i_3\cdots i_k) \times (\text{cycles disjoint from } i_1, i_2, \ldots, i_k) \in K$ where $k \ge 4$, show that K contains a 3-cycle.

   Hint: Compute the commutator $(i_1i_2i_3)^{-1}\sigma^{-1}(i_1i_2i_3)\sigma$.

(d) If $\sigma = (i_1 i_2)(i_3 i_4) \times$ (cycles disjoint from i_1, i_2, i_3, i_4) $\in K$, show that K contains a 3-cycle.
Hint: Similar to (c).

(e) If $\sigma = (i_1 i_2 i_3)(i_4 i_5 i_6) \times$ (cycles disjoint from $i_1, i_2, i_3, \ldots, i_6$) $\in K$, show that K contains a 3-cycle.

(f) Complete your proof using the steps above. What goes wrong with your proof if $n \leq 4$?

8. Consider the dihedral group $D_n = \langle R, T \rangle$ of order $2n$, using the notation of Section 2. Find all conjugacy classes of D_n .

9. Let G be a group. Define the **left** and **right regular representations** of G by $\lambda, \rho : G \to \operatorname{Sym} G$ where

$$\lambda_a : G \to G, \quad x \mapsto a^{-1}x;$$
$$\rho_a : G \to G, \quad x \mapsto xa.$$

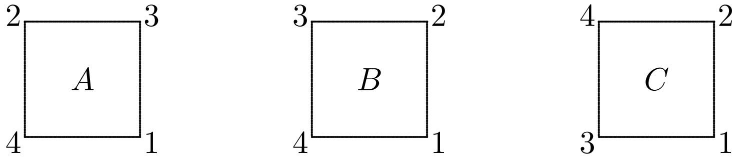
(a) Prove that $\lambda_a \lambda_b = \lambda_{ab}$ and $\rho_a \rho_b = \rho_{ab}$ with left-to-right composition as usual. Thus λ and ρ *are* permutation representations of G. Why did we require the a^{-1} in the definition of λ_a?

(b) Show that $G \cong \langle \lambda_a : a \in G \rangle \cong \langle \rho_a : a \in G \rangle \leq \operatorname{Sym} G$.
Hint: Use the First Isomorphism Theorem.

(c) Let $M = \langle \lambda_a, \rho_a \ : \ a \in G \rangle$. Show that $M \cong (G \times G)/Z$ where $Z = \{(z, z) : z \in Z(G)\}$.
Hint: Use the First Isomorphism Theorem.

10. More on Transitive Permutation Representations

We begin by giving two examples of permutation actions ϕ, ψ of the group $G = S_4$. In the action ϕ, we let G permute the set $\mathcal{S} = \{A, B, C\}$ consisting of the following three squares:



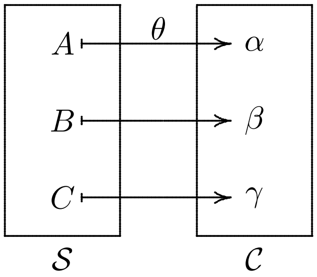
The action of G is defined naturally by its action on the vertex labels $\{1, 2, 3, 4\}$. For example, consider $(12) \in G$, which is a symmetry of square A, so $A^{(12)} = A$; however, it interchanges the other two squares: $B^{(12)} = C$, $C^{(12)} = B$. Thus $\phi\big((12)\big) = (BC) \in \operatorname{Sym} \mathcal{S}$ in the usual cycle notation. One of the conjugacy classes of G is $\mathcal{C} = \{\alpha{=}(12)(34), \beta{=}(13)(24), \gamma{=}(14)(23)\}$ and we denote by ψ the action of G on $\mathcal{C}$ by conjugation. For example, consider the action of $(12) \in G$:

$$\alpha^{(12)} = \alpha, \quad \beta^{(12)} = \gamma, \quad \gamma^{(12)} = \beta$$

so we may write $\psi\big((12)\big) = (\beta\gamma) \in \operatorname{Sym} \mathcal{C}$. Writing () for the identity permutation on both $\mathcal{S}$ and $\mathcal{C}$, a table of values of the two actions ϕ, ψ is as follows:

g	$\phi(g)$	$\psi(g)$	g	$\phi(g)$	$\psi(g)$	g	$\phi(g)$	$\psi(g)$
(1)	()	()	(132)	(ABC)	$(\alpha\beta\gamma)$	(13)(24)	()	()
(12)	(BC)	$(\beta\gamma)$	(124)	(ABC)	$(\alpha\beta\gamma)$	(14)(23)	()	()
(13)	(AC)	$(\alpha\gamma)$	(142)	(ACB)	$(\alpha\gamma\beta)$	(1234)	(AC)	$(\alpha\gamma)$
(14)	(AB)	$(\alpha\beta)$	(134)	(ACB)	$(\alpha\gamma\beta)$	(1243)	(AB)	$(\alpha\beta)$
(23)	(AB)	$(\alpha\beta)$	(143)	(ABC)	$(\alpha\beta\gamma)$	(1324)	(BC)	$(\beta\gamma)$
(24)	(AC)	$(\alpha\gamma)$	(234)	(ABC)	$(\alpha\beta\gamma)$	(1342)	(AB)	$(\alpha\beta)$
(34)	(BC)	$(\beta\gamma)$	(243)	(ACB)	$(\alpha\gamma\beta)$	(1423)	(BC)	$(\beta\gamma)$
(123)	(ACB)	$(\alpha\gamma\beta)$	(12)(34)	()	()	(1432)	(AC)	$(\alpha\gamma)$

It should be clear from this table that the two permutation representations ϕ, ψ are equivalent: every element $g \in G$ permutes the three squares A, B, C in exactly the same way as it permutes the three involutions α, β, γ. In other words, if we simply rename the three points being permuted, each of the permutations $\phi(g)$ becomes $\psi(g)$. This choice of renaming is formally expressed by the bijection



The equivalence of the two permutation representations of G is expressed by the rule that $\phi(g)\theta = \theta\psi(g)$ for all $g \in G$, using left-to-right composition. For example the fact that $B \xmapsto{\phi((12))} C$ gives, after using θ to rename points, the fact that $B^\theta \xmapsto{\psi((12))} C^\theta$. In other words, the effect of the composite map

$$B \xmapsto{\phi((12))} C \xmapsto{\theta} C^\theta$$

is the same as the effect of the composite map

$$B \xmapsto{\theta} B^\theta \xmapsto{\psi((12))} C^\theta.$$

In other words, $\phi((12))\theta = \theta\psi((12))$. We are ready to formally define the notion of equivalence of permutation actions. This is the appropriate equivalence relation for actions, just as isomorphism is the appropriate equivalence relation for groups.

Let $\phi : G \to \operatorname{Sym}\mathcal{S}$ and $\phi' : G \to \operatorname{Sym}\mathcal{S}'$ be two permutation representations of the same group G. Then we say that ϕ is **equivalent** to ϕ' if there exists a bijection $\theta : \mathcal{S} \to \mathcal{S}'$ such that $\phi(g)\theta = \theta\phi'(g)$ for all $g \in G$, i.e. the following diagram commutes:

$$\begin{tikzcd}
\mathcal{S} \arrow[r, "\theta"] \arrow[d, "\phi(g)"] & \mathcal{S}' \arrow[d, "\phi'(g)"] \\
\mathcal{S} \arrow[r, "\theta"] & \mathcal{S}'
\end{tikzcd}$$

The map θ which achieves this equivalence, is called an **intertwining map**. To say that this diagram commutes means that if we compose maps indicated by arrows, in this case from upper left to lower right, it does not matter which of the two possible routes we take (i.e. 'right, then down' gives the same result as 'down, then right').

There is a third permutation representation of G, equivalent to both of the two actions above, that we need to mention: the action of G by right-multiplication on the three right cosets of the subgroup $H = G_A = C_G(\alpha)$. This action is the map $\rho : G \to \mathrm{Sym}(H\backslash G)$ defined by $g \mapsto \rho_g$ where

$$\rho_g : (H\backslash G) \to (H\backslash G), \qquad Hx \mapsto Hxg.$$

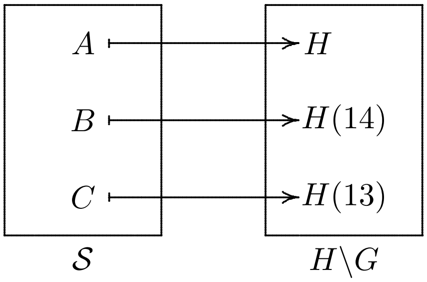
Explicitly, we list the right cosets of H as

$$\begin{aligned} H &= \{(1),\, (1324),\, (12)(34),\, (1423),\, (13)(24),\, (14)(23),\, (12),\, (34)\},\\ H(14) &= \{(14),\, (132),\, (1243),\, (234),\, (1342),\, (23),\, (124),\, (143)\},\\ H(13) &= \{(13),\, (243),\, (1234),\, (142),\, (24),\, (1432),\, (123),\, (134)\} \end{aligned}$$

We determine for example the action of $(12) \in G$:

$$\begin{aligned} H^{\rho_{(12)}} &= H(12) = H,\\ \big(H(14)\big)^{\rho_{(12)}} &= H(14)(12) = H(142) = H(13),\\ \big(H(13)\big)^{\rho_{(12)}} &= H(13)(12) = H(132) = H(14). \end{aligned}$$

We may write $\rho_{(12)} = \big(H(14)\,,\; H(13)\big)$, if we are willing to stretch the customary cycle notation a little. In fact, for every $g \in G$, ρ_g permutes right cosets of H in exactly the same way as $\phi(g)$ permutes squares A, B, C; also the same way as $\psi(g)$ permutes α, β, γ. An intertwining map from ϕ to ρ is the map



This is typical of the more general result

10.1 Theorem. Let G be a group. Every transitive permutation action of G is equivalent to the action of G, by right-multiplication, on the right cosets of a subgroup $H \le G$. This subgroup is actually the stabilizer of one of the points.

Proof. Let $\phi : G \to \mathrm{Sym}\,\mathcal{S}$ be a transitive permutation representation, and let $a \in \mathcal{S}$. We write simply a^g in place of $a^{\phi(a)}$. Let $H = G_a = \{g \in G : a^g = a\}$. In the course of proving Proposition 8.1, we showed that for all $x, y \in G$, $a^x = a^y$ iff $Hx = Hy$. In particular we have a well-defined bijection

$$\theta : \mathcal{S} \to H\backslash G, \quad a^x \mapsto Hx.$$

Let $g \in G$. Then for every $x \in G$ we have

$$(a^x)^{\theta\rho_g} = ((a^x)^\theta)g = (Hx)g = Hxg,$$

whereas

$$(a^x)^{\phi(g)\theta} = ((a^x)^g)^\theta = (a^{xg})^\theta = Hxg.$$

Thus $\theta\rho_g = \phi(g)\theta$ for all $g \in G$ as required. □

Return now to our example $G = S_4$ with action $\phi : G \to \operatorname{Sym}\mathcal{S}$ where $\mathcal{S} = \{A, B, C\}$. We see that $\ker\phi = \{(1), (12)(34), (13)(24), (14)(23)\}$, so ϕ is not faithful. Evidently it is impossible for S_4 to have a faithful permutation representation of degree 3, since the order $|S_4| = 24$ exceeds $3! = 6$. Moreover by the First Isomorphism Theorem, $G/\ker\phi \cong \phi(G) \leq \operatorname{Sym}\mathcal{S}$. We have equality: $\phi(G) = \mathcal{S}$, either by comparing orders ($\frac{24}{4} = 6 = 3!$), or by observing, from our table of values of $\phi(g)$, that ϕ is surjective.

What is the kernel of an arbitrary transitive permutation representation? Let $\phi : G \to \operatorname{Sym}\mathcal{S}$ be an arbitrary transitive permutation representation, and let $a \in \mathcal{S}$. As before, let $H = G_a$. We saw in Section 9 that the stabilizer of an arbitrary point $a^g \in \mathcal{S}$ is the conjugate subgroup $H^g = g^{-1}Hg$. The set of all elements $g \in G$ fixing *every* point of $\mathcal{S}$ is the intersection of *all* these point stabilizers. This intersection is known as the **core** of H in G. We have just proved

10.2 Proposition. Let $\phi : G \to \operatorname{Sym}\mathcal{S}$ be a transitive permutation representation, and let H be the stabilizer of some point of $\mathcal{S}$. Then $\ker\phi = \operatorname{Core}_G(H)$.

Here we define, for an arbitrary group G and arbitrary subgroup $H \leq G$,

$$\operatorname{Core}_G(H) = \bigcap_{g\in G} H^g = \bigcap_{g\in G} g^{-1}Hg.$$

This is a *normal* subgroup of G, since for all $x \in G$, we have

$$\left(\bigcap_{g\in G} H^g\right)^x = \bigcap_{g\in G} H^{gx} = \bigcap_{y\in G} H^y = \bigcap_{g\in G} H^g.$$

Alternatively, $\operatorname{Core}_G(H) \trianglelefteq G$ follows from the fact that $\operatorname{Core}_G(H) = \ker\phi$. In fact, this is the *largest* normal subgroup of G contained in H, in the sense that if $K \trianglelefteq G$ and $K \subseteq H$, then $K \subseteq \operatorname{Core}_G(H)$, since $K = K^g \subseteq H^g$ for all $g \in G$, and $K \subseteq \bigcap_{g\in G} H^g$.

Finally, we are in a position to answer our earlier query about finding faithful permutation representations of small degree for a given group G. (This is equivalent to expressing G itself, not a quotient thereof, as a subgroup of S_n for small n.) Let H be a (fairly large) subgroup of G, so that G acts transitively on the right cosets of H by right-multiplication. Denote by ρ the corresponding permutation representation. In addition to choosing H large (so that the degree $[G:H]$ of the representation will be small), we require H to be **corefree** in G, i.e. $\mathrm{Core}_G(H) = \bigcap_{g\in G} H^g = 1$. This will ensure that the representation is faithful.

Exercises 10.

1. Prove that equivalence of permutation actions is an equivalence relation.

2. Up to equivalence, there are just two transitive subgroups of S_3, namely $\langle(123)\rangle$ and S_3 itself. Up to equivalence, how many transitive permutation groups of degree 4 are there? Of degree 5?

3. Show that every finite group G is isomorphic to a subgroup of A_n for some $n \geq 1$. What is the minimum possible value of n if $G \cong S_3 \times S_3$?

4. Let G be the group of rotational symmetries of a cube, so that $|G| = 24$. Follow steps (a)–(d) to show that $G \cong S_4$.

   (a) Define a *diagonal* of the cube to be a line joining two antipodal vertices. There are four such diagonals, each of which passes through the center of the cube. Label the four diagonals by $\mathcal{D} = \{1,2,3,4\}$. Clearly G acts on $\mathcal{D}$; denote the corresponding action by ϕ. Show that this action is transitive.

   (b) Let H, the subgroup of $S_4 = \mathrm{Sym}\,\mathcal{D}$ generated by the eight 3-cycles. Show that $H = S_4$.

   Hint: Lagrange's Theorem limits the possibilities for a subgroup of S_4.

   (c) By considering 120° rotations about the diagonals, show that the image $\phi(G)$ contains the eight 3-cycles.

   (d) By considering the orders of G and S_4, show that $|\ker\phi| = 1$ and that ϕ is an isomorphism $G \to S_4$.

5. Show that the full isometry group G of a cube is isomorphic to $C_2 \times S_4$.

   Hint: The rotational symmetry group $H < G$ satisfies $H \cong S_4$ by Exercise 10.4. Argue that $|G| = 48$ and so $H \lhd G$. Observe that there is an element $\tau \in G$ mapping each vertex of the cube to its antipode (the vertex furthest away). We may consider every element of G as a linear transformation $\mathbb{R}^3 \to \mathbb{R}^3$, with origin at the center of the cube. By considering the matrix representing τ, show that $\tau \in Z(G)$. Use Exercise 6.13.

6. Consider the dodecahedron pictured on the title page of these notes.

   (a) Show that the rotational symmetry group of the dodecahedron is isomorphic to A_5.

   (b) Show that the full symmetry group of the dodecahedron is isomorphic to $C_2 \times A_5$.

   Hint: Recall Exercises 10.4 and 10.5.

7. Recall (see Exercise 3.7) that the group $G = GL_2(\mathbb{F}_3)$ has order 48.

   (a) Let H be the set of all upper-triangular matrices in G. Show that H is a subgroup of order 12.

   (b) Determine $K = \mathrm{Core}_G(H)$.

   (c) Using the action of G on right cosets of H, conclude that $G/K \cong S_4$. Thus, in notation similar to Exercise 6.11, we have $PGL_2(\mathbb{F}_3) \cong S_4$.

11. p-Groups and Sylow p-Subgroups

For a prime p, a **p-group** is a group whose order is a power of p. Thus, for example, a 2-group has order $\in \{1,2,4,8,\ldots\}$. One property which makes p-groups special is that they have nontrivial centers, as the following proposition shows. It is easy to find non-p-groups with trivial centers, such as S_3.

11.1 Proposition. If P is a nontrivial p-group, then P has a nontrivial center, i.e. $|Z(P)| \neq 1$.

Proof. Express P as a disjoint union of its conjugacy classes, thus:

$$P = \{e\} \cup \mathcal{C}_2 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_r .$$

Each $|\mathcal{C}_i|$ divides $|P|$ and so is a power of p. Then for some $i \in \{2,3,\ldots,r\}$, we must have that $|\mathcal{C}_i|$ has size 1; otherwise $|P| \equiv 1 \mod p$, a contradiction. This shows that $\{e\}$ is not the only singleton conjugacy class in P. By Exercise 9.5, it follows that $Z(P)$ is nontrivial. □

As simple as the latter proof is, it uses an important trick which will be encountered again in the next section, in proving the Sylow Theorems.

Now let G be any finite group. A subgroup $P \leq G$ is called a **p-subgroup** if P is a p-group. In case $|P|$ is the *highest* power of p dividing $|G|$, we call P a **Sylow p-subgroup** of G. Clearly if $p \not\big| \, |G|$ then $\{e\}$ is the only Sylow p-subgroup of G. But if $p \bigm| |G|$, it is not obvious that G has *any* nontrivial p-subgroups. The existence of Sylow p-subgroups is assured by the famous Sylow theorems, which we shall prove in Section 12. Their existence may be regarded as a partial converse to Lagrange's Theorem, in the case of prime-power divisors of $|G|$. Recalling that G does not necessarily have subgroups of every order dividing $|G|$, this shows another respect in which prime power orders are special.

Observe also that if P is a Sylow p-subgroup of G, then $|P^g| = |P|$ for every $g \in G$, so clearly every conjugate of P is also a Sylow p-subgroup. It is less obvious that *every* Sylow p-subgroup of G is conjugate to P, but this is another of Sylow's theorems treated in the next section.

In the remainder of this section we will be content to show how Sylow p-subgroups may be constructed for one very important family of groups:

11.2 Proposition. For every positive integer n and every prime p, the group S_n has a Sylow p-subgroup.

We proceed to construct only a Sylow 3-subgroup of S_n; however you should have no trouble in modifying our construction for a general prime p. Please observe that S_n may have *many* Sylow 3-subgroups, and at present we are interested only in finding one. Note that, for example, $|S_4| = 24 = 2^3 \cdot 3$ has four Sylow 3-subgroups, namely $\langle(123)\rangle$, $\langle(124)\rangle$, $\langle(134)\rangle$ and $\langle(234)\rangle$, each of which is cyclic of order 3.

Since $|S_1| = 1$ and $|S_2| = 1$, these groups have only the trivial Sylow 3-subgroup $\{(1)\}$. So we may suppose that $n \geq 3$. Now $|S_3| = 6 = 2 \cdot 3$, $|S_4| = 24 = 2^3 \cdot 3$ and $|S_5| = 120 = 2^3 \cdot 3 \cdot 5$, so for $3 \leq n \leq 5$, a Sylow 3-subgroup of S_n is given by $\langle(123)\rangle$. We illustrate the action of our chosen Sylow 3-subgroup with a picture in each case, intended to emphasize how it cyclically permutes $1, 2, 3$:

$$\begin{array}{ccc} \underrightarrow{\boxed{1\,2\,3}} & \underrightarrow{\boxed{1\,2\,3}}\ \ 4 & \underrightarrow{\boxed{1\,2\,3}}\ \ 4\,5 \\ n=3 & n=4 & n=5 \end{array}$$

Since $|S_6| = 720 = 2^4 \cdot 3^2 \cdot 5$, a Sylow 3-subgroup of S_6 has order 3^2; we may choose $\langle(123),(456)\rangle$, which permutes $1, 2, 3$ cyclically, and independently permutes $4, 5, 6$ cyclically. This group is isomorphic to $C_3 \times C_3$, and we represent its action by the picture:

$$\begin{array}{c} \underrightarrow{\boxed{1\,2\,3}}\ \underrightarrow{\boxed{4\,5\,6}} \\ n=6 \end{array}$$

Since $|S_7| = 2^4 \cdot 3^3 \cdot 5 \cdot 7$ and $|S_8| = 2^7 \cdot 3^2 \cdot 5 \cdot 7$, we may also choose $\langle(123),(456)\rangle$ as a Sylow 3-subgroup of S_7 and of S_8 :

$$\begin{array}{cc} \underrightarrow{\boxed{1\,2\,3}}\ \underrightarrow{\boxed{4\,5\,6}}\ \ 7 & \underrightarrow{\boxed{1\,2\,3}}\ \underrightarrow{\boxed{4\,5\,6}}\ \ 7\,8 \\ n=7 & n=8 \end{array}$$

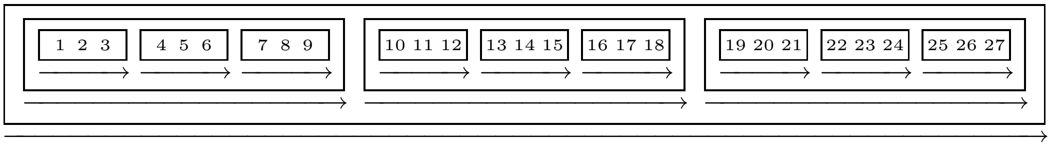
Now $|S_9| = 2^7 \cdot 3^4 \cdot 5 \cdot 7$, so $\langle(123),(456),(789)\rangle$ does not suffice. However, a Sylow 3-subgroup of S_9 is given by the nonabelian subgroup

$$\langle(123),(456),(789),(147)(258)(369)\rangle = \langle(123),(147)(258)(369)\rangle$$

of order 3^4, where (123), (456) and (789) each cyclically permutes the points within a *block* of size three, and (147)(258)(369) cyclically permutes these three blocks themselves:

$$\begin{array}{c} \underrightarrow{\boxed{\underrightarrow{\boxed{1\,2\,3}}\ \underrightarrow{\boxed{3\,4\,5}}\ \underrightarrow{\boxed{7\,8\,9}}}} \\ n=9 \end{array}$$

Let's call the latter pattern of three blocks a *superblock.* Now let's jump ahead and see what happens when n reaches the next power of 3, namely 27. In this case $|S_{27}| = 2^{23} \cdot 3^{13} \cdot 5^6 \cdot 7^3 \cdot 11^2 \cdot 13^2 \cdot 17 \cdot 19 \cdot 23$, so we may construct a Sylow 3-subgroup of S_{27} which cyclically permutes three superblocks of size nine inside one *superduperblock,* and acts individually on each superblock as in the case $n = 9$ above:

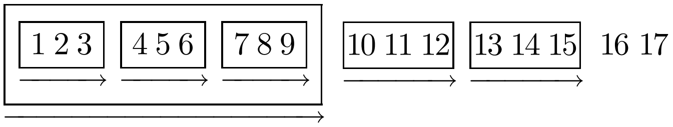


$$n = 27$$

Such a Sylow 3-subgroup has order 3^{13}. Now for general n, consider the ternary (i.e. base 3) expression for n, say

$$n = n_0 + 3n_1 + 3^2 n_2 + 3^3 n_3 + \cdots$$

where each $n_i \in \{0, 1, 2\}$. Then a Sylow 3-subgroup of S_n has n_0 fixed points, plus n_1 blocks, plus n_2 superblocks, plus n_3 superduperblocks, etc. For example, 17 is written as 122 (base 3), so we need 2 fixed points, plus 2 blocks, plus 1 superblock, thus:



$$n = 17$$

Explicitly, our Sylow 3-subgroup of S_{17} is given by

$$\begin{aligned}&\langle (1\,2\,3),\ (4\,5\,6),\ (7\,8\,9),\ (1\,4\,7)(2\,5\,8)(3\,6\,9),\ (10\,11\,12),\ (13\,14\,15)\rangle\\ &\qquad = \langle (1\,2\,3),\ (1\,4\,7)(2\,5\,8)(3\,6\,9),\ (10\,11\,12),\ (13\,14\,15)\rangle\end{aligned}$$

of order 3^6. It may be shown that the highest power of p dividing $|S_n| = n!$ is p^e where

$$e = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots$$

and as usual, $\lfloor x \rfloor$ denotes the greatest integer $\le x$. Note that the latter sum has only finitely many nonzero terms; for example

$$\left\lfloor \frac{17}{3} \right\rfloor + \left\lfloor \frac{17}{9} \right\rfloor + \left\lfloor \frac{17}{27} \right\rfloor + \left\lfloor \frac{17}{81} \right\rfloor + \cdots = 5 + 1 + 0 + 0 + \cdots = 6,$$

so a Sylow 3-subgroup of S_{27} has order 3^6 as above.

Exercises 11.

1. List *all* Sylow 2-subgroups, *all* Sylow 3-subgroups, and *all* Sylow 5-subgroups of S_5.

2. Let H be a proper subgroup of a p-group P, and let $N = N_P(H)$. Prove that $H \lhd N$.

   Hint: The key conclusion here, namely the fact that H is *properly* contained in N, follows by an argument similar to the proof of Proposition 11.1.

3. Let $G = GL_n(\mathbb{F}_p)$, the multiplicative group of invertible $n \times n$ matrices over $\mathbb{F}_p = \{0, 1, 2, \ldots, p-1\}$; see Exercises 3.7 and 8.5. By considering upper-triangular matrices, find an explicit Sylow p-subgroup of G.

12. The Sylow Theorems

The Theorems of Sylow are among the most celebrated in group theory. Their proofs will require a little more work than previous results.

12.1 Sylow Theorems. Let G be a finite group, and let p be a prime. Then the following statements hold.

(i) G has a Sylow p-subgroup.

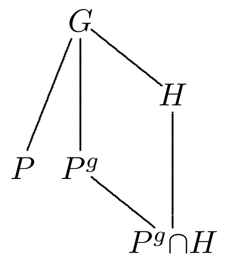
(ii) Any two Sylow p-subgroups of G are conjugate in G.

(iii) Every p-subgroup of G is contained in a Sylow p-subgroup of G.

(iv) The number of Sylow p-subgroups of G is a divisor of $|G|$ of the form $kp+1$ for some integer $k \geq 0$.

Note that (iv) implies (i); however, we shall prove (i) before the remaining conclusions. The first Sylow Theorem (i) is an easy consequence of the following Proposition 12.2. To see why, let H be any finite group. If n is large enough, then by the Cayley Representation Theorem 8.2, we may consider H to be a subgroup of S_n . Now S_n has a Sylow p-subgroup P by Section 11. Then Proposition 12.2 says that we can obtain a Sylow p-subgroup of H by intersecting H with a suitable conjugate of P in S_n . [Alternatively, H is isomorphic to a subgroup of $GL_n(\mathbb{F}_p)$ by Exercise 8.5, and $GL_n(\mathbb{F}_p)$ has a Sylow p-subgroup by Exercise 11.3.] Later we will also see how conclusions (ii) and (iii) also follow from this same Proposition; then we will need another argument to obtain (iv). But let's first verify the following

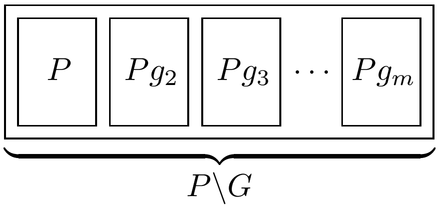
12.2 Proposition. Let G be a finite group with subgroup H. If P is a Sylow p-subgroup of G, then there exists $g \in G$ such that $P^g \cap H$ is a Sylow p-subgroup of H.



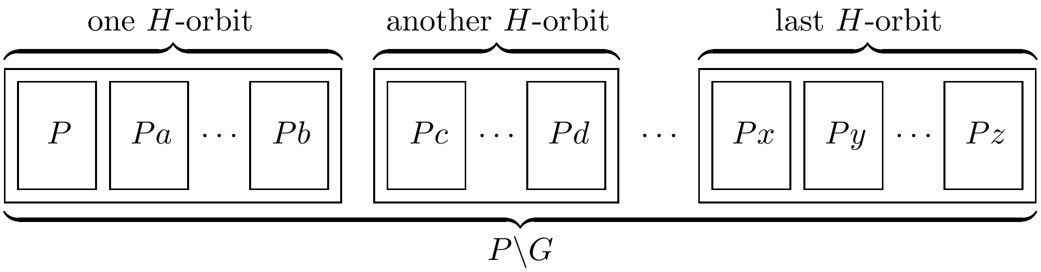
Proof. As usual we denote the set of right cosets of P in G by

$$P\backslash G = \{Pg : g \in G\}.$$

The number of cosets, $m = [G : P] = |G|/|P|$, is not divisible by p since $|P|$ is the highest power of p dividing $|G|$. Recall from Section 8 that G acts transitively on $P\backslash G$ by right multiplication; and that the stabilizer of P in this action is P. Using observations from Section 9, the stabilizer of any right coset $Pg \in P\backslash G$ is the conjugate P^g.



Consider now the action of H on $P\backslash G$ by right-multiplication. In general we cannot expect H (like G) to permute the right cosets of P in G transitively; rather, $P\backslash G$ may split into several H-orbits, as shown:



However, one thing is certain: the H-orbits *do not all* have length divisible by p, since the total of the lengths of H-orbits, namely m, is not divisible by p. So we may choose a coset Pg belonging to an H-orbit of length not divisible by p. Since the stabilizer of Pg in G is P^g (see Section 9), the stabilizer of Pg in H is

$$\{x \in P^g : x \in H\} = P^g \cap H.$$

So by Proposition 8.1, the length of the H-orbit containing Pg is $[H : P^g \cap H] = |H|/|P^g \cap H|$, which is not divisible by p. However, $P^g \cap H \leq P^g$ is a p-group. This means that $|P^g \cap H|$ is the highest power of p dividing H, as required. □

We now complete the proof of the Sylow Theorems. Conclusion (i) has been verified, as noted earlier.

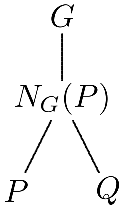
To prove (ii), let P and Q be two Sylow p-subgroups of G. Applying Proposition 12.2 with Q in place of H, there exists $g \in G$ such that $P^g \cap Q$ is a Sylow p-subgroup of Q. Since $|P^g \cap Q|$ is the highest power of p dividing $|Q|$, we must have $P^g \cap Q = Q$, i.e. $Q \subseteq P^g$. But $|Q| = |P^g|$ is the highest power of p dividing $|G|$, so $Q = P^g$, which proves (ii).

To prove (iii), let P be a Sylow p-subgroup of G, and let H be *any* p-subgroup of G. By Proposition 12.2, there exists $g \in G$ such that $P^g \cap H$ is a Sylow p-subgroup of H. As before, this means that $P^g \cap H = H$, so H is contained in P^g, a Sylow p-subgroup of G.

Finally, to prove (iv), we make use of the following:

12.3 Proposition. If $P \neq Q$ are two Sylow p-subgroups of G, then Q does not normalize P, i.e. $Q \not\subseteq N_G(P)$.

Proof of Proposition 12.3. Suppose that $Q \subseteq N_G(P)$. Since $P \subseteq N_G(P)$, we observe the inclusions indicated by the diagram



Since $|P| = |Q|$ is the highest power of p dividing $|G|$, it must also be the highest power of p dividing $|N_G(P)|$. So P and Q are Sylow p-subgroups of $N_G(P)$. By (ii), there exists $g \in N_G(P)$ such that $P^g = Q$. But since $g \in N_G(P)$, by definition we have $P^g = P$. Thus $P = P^g = Q$, contradicting the assumption that P and Q are distinct. □

We now return to prove (iv). Let P be a Sylow p-subgroup of G, whose existence is guaranteed by (ii). Let $\mathcal{S} = \{P_1{=}P, P_2, P_3, \ldots, P_s\}$ be the class of all Sylow p-subgroups of G. If $s = 1$ then we are done; hence we assume that $s \geq 2$. By (ii), every member of $\mathcal{S}$ is a conjugate of P, or in words, G acts transitively on $\mathcal{S}$ by conjugation:

$$\underbrace{\boxed{\quad P \qquad P_2 \qquad P_3 \qquad \cdots \qquad P_s \quad}}_{\mathcal{S}}$$

Recall from Section 9 that $N_G(P)$ is the stabilizer of P in this action, and so $s = [G : N_G(P)] = |G|/|N_G(P)|$ is a divisor of $|G|$. We must show that $s \equiv 1 \mod p$.

Consider the action of P on $\mathcal{S}$ by conjugation. This will no longer be transitive! In particular P normalizes P, so P fixes P in this action. The remaining members of $\mathcal{S}$ are partitioned into P-orbits of undetermined size, thus:

$$\underbrace{\overbrace{\boxed{P}}^{\substack{\text{fixed}\\ \text{by } P}} \; \overbrace{\boxed{P_{i_1} \quad P_{i_2} \quad \cdots \quad P_{i_t}}}^{\text{another } P\text{-orbit}} \; \overbrace{\boxed{P_{j_1} \quad P_{j_2} \quad \cdots \quad P_{j_u}}}^{\text{yet another } P\text{-orbit}} \; \cdots \; \overbrace{\boxed{P_{\ell_1} \quad P_{\ell_2} \quad \cdots \quad P_{\ell_v}}}^{\text{last } P\text{-orbit}}}_{\mathcal{S}}$$

By Proposition 8.1, the length of every P-orbit divides $|P|$, and so is one of $1, p, p^2, \ldots, |P|$. But if $i \neq 1$ then by Proposition 12.3, P does not normalize P_i. This means that P_i is not fixed in the conjugation action of P, so the length of the orbit containing P_i is one of $p, p^2, \ldots, |P|$. In particular $\mathcal{S}$ is a disjoint union of P-orbits, all having length divisible by p, except for the singleton orbit $\{P\}$, which has length 1. Thus $s = |\mathcal{S}| \equiv 1 \mod p$ as required. □

Exercises 12.

1. If G is a group of order 42, how many subgroups of order 7 does G have? Why?
2. Show that every group of order 15 is cyclic.
3. If H is any subgroup of S_{11} of order 110, must H act transitively on $\{1, 2, 3, \ldots, 11\}$? Why or why not?
4. Let G be a finite group, P a Sylow p-subgroup of G for some prime p, and $N = N_G(P)$. If $H \trianglelefteq G$ has index $[G : H]$ not divisible by p, show that $HN = G$.

   Hint: Use Exercise 8.2.

*5. Show that every simple group of order 60 is isomorphic to A_5 .

   Hint: Consider the action of G on its Sylow 2-subgroups by conjugation.

13. Composition Series

A **normal series** for a group G is a chain of subgroups of the form

$$1 = G_0 \lhd G_1 \lhd G_2 \lhd \cdots \lhd G_{k-1} \lhd G_k = G,$$

this example having **length** k. Sometimes we may relax the condition $G_{i-1} \lhd G_i$ to $G_{i-1} \trianglelefteq G_i$; but in such cases the repeated subgroups may be deleted until the remaining inclusions are proper. It does *not* follow (cf. Exercise 5.4) that every G_i is normal in G; rather, we say that G_i is **subnormal** in G. A second normal series for G,

$$1 = H_0 \lhd H_1 \lhd H_2 \lhd \cdots \lhd H_{\ell-1} \lhd H_\ell = G,$$

is called a **refinement** of the former series, if every G_i is an H_j for some J, i.e. if the chain $H_j : 0 \le j \le \ell$ is obtained from the chain $G_i : 0 \le i \le k$ by 'squeezing in' additional subgroups in a way that preserves normality of adjacent terms in the chain, and in particular $\ell \ge k$. Every group $G \neq 1$ has a trivial normal series $1 \lhd G$, of which all other normal series are refinements. Observe that G is simple iff the series $1 \lhd G$ cannot be properly refined.

Assume that $1 < |G| < \infty$. Starting with any normal series for G (such as the trivial series), after a finite number of refinements we arrive at a normal series for G which may not be further refined. Such a series is called a **composition series** for G. Equivalently, any normal series $1 = G_0 \lhd G_1 \lhd G_2 \lhd \cdots \lhd G_{k-1} \lhd G_k = G$ is a composition series iff G_{i-1} is maximal normal in G_i for all $i = 1, 2, \ldots, k$ (see Exercise 6.6), iff each factor G_i/G_{i-1} is simple. Accordingly, the k quotient groups G_i/G_{i-1} are called the **composition factors** of G, and k is the **composition length** of G. The composition factors and composition length of G are independent of the choice of composition series:

13.1 Jordan-Hölder Theorem. Any two composition series for a finite group G have the same length and factor groups, counting multiplicity.

We will not prove this result here, although it is fundamentally important for us to be able to use the result. The Jordan-Hölder Theorem generalizes the Fundamental Theorem of Arithmetic (see Section 0). Many problems in finite group theory are tackled by induction on the composition length of the group, typically by reducing the problem to the case of a simple group.

Consider, for example, the cyclic group of order 12, denoted $C_{12} = \{e, x, x^2, \dots, x^{11}\}$, whose subgroups are described explicitly in Section 4. Since every subgroup is normal, C_{12} has exactly three composition series, each of length three:

$$\begin{array}{c} 1 \lhd \langle x^6\rangle \lhd \langle x^3\rangle \lhd C_{12} \\ \underbrace{C_2 \qquad C_2 \qquad C_3}_{\text{composition factors}} \end{array} \qquad \begin{array}{c} 1 \lhd \langle x^6\rangle \lhd \langle x^2\rangle \lhd C_{12} \\ \underbrace{C_2 \qquad C_3 \qquad C_2}_{\text{composition factors}} \end{array} \qquad \begin{array}{c} 1 \lhd \langle x^4\rangle \lhd \langle x^2\rangle \lhd C_{12} \\ \underbrace{C_3 \qquad C_2 \qquad C_2}_{\text{composition factors}} \end{array}$$

Whichever composition series we choose for C_{12}, the composition factors are C_2, C_2 and C_3. As another example, a composition series of length 4 for S_4 is shown:

$$\begin{array}{c} 1 \lhd \langle (12)(34)\rangle \lhd \langle (12)(34), (13)(24)\rangle \lhd A_4 \lhd S_4 \\ \underbrace{C_2 \qquad\qquad\quad C_2 \qquad\qquad\qquad\qquad\quad C_3 \quad C_2}_{\text{composition factors}} \end{array}$$

Only two other composition series for S_4 exist, found by replacing $\langle (12)(34)\rangle$ by either $\langle (13)(24)\rangle$ or $\langle (14)(23)\rangle$. Observe that from the composition series above, A_4 has the same composition factors as C_{12}. So unlike a positive integer, which is uniquely determined by its prime factors (counting multiplicity), a group is *not* uniquely determined by its composition factors. A different sort of example is S_5, whose unique composition series is

$$\begin{array}{c} 1 \lhd A_5 \lhd S_5 \\ \underbrace{A_5 \quad C_2}_{\text{composition factors}} \end{array}$$

We say that G is **solvable** if all its composition factors are cyclic of prime order. Conversely, G is **nonsolvable** if at least one nonabelian simple group is among the composition factors of G. The examples above show that C_{12} and S_4 are solvable, whereas S_5 is nonsolvable. Solvability is an important property in the study of groups, and especially so for us since it relates directly to the question of whether a given polynomial equation is solvable by radicals ... more about this later in the course!

13.2 Proposition. If G is solvable, then so is every subgroup of G.

Proof. Let $H \le G$, where G has a composition series

$$1 = G_0 \lhd G_1 \lhd G_2 \lhd \cdots \lhd G_{k-1} \lhd G_k = G$$

in which each factor G_i/G_{i-1} is cyclic of prime order. Intersect every term in the latter series with H to obtain $H_i = H \cap G_i$. It is easily checked that $H_{i-1} \trianglelefteq H_i$ for $i = 1, 2, \ldots, k$. Furthermore

$$H_i/H_{i-1} = H_i/(H_i \cap G_{i-1}) \cong H_i G_{i-1}/G_{i-1} \leq G_i/G_{i-1}$$

using the Second Isomorphism Theorem (Exercise 6.8). Since G_i/G_{i-1} is cyclic of prime order, H_i/H_{i-1} is either trivial or cyclic of prime order. So after removing duplications from the sequence

$$1 = H_0 \trianglelefteq H_1 \trianglelefteq H_2 \trianglelefteq \cdots \trianglelefteq H_{k-1} \trianglelefteq H_k = H,$$

we obtain a composition series for H, in which all composition factors are cyclic of prime order. □

It is also true that homomorphic images of solvable groups are solvable. A stronger statement is the following

13.3 Proposition. Suppose that $H \lhd G$. Then G is solvable iff both H and G/H are solvable.

Proof. We may assume that $1 \neq H \neq G$. Refine the normal series $1 \lhd H \lhd G$ to obtain a composition series for G which includes H:

$$1 = G_0 \lhd G_1 \lhd G_2 \lhd \cdots \lhd G_{k-1} \lhd G_k = H \lhd G_{k+1} \lhd \cdots \lhd G_{\ell-1} \lhd G_\ell = G.$$

From this we can immediately read off a composition series for H:

$$1 = G_0 \lhd G_1 \lhd G_2 \lhd \cdots \lhd G_{k-1} \lhd G_k = H,$$

and the composition factors for H are the first k of the composition series for G, namely G_i/G_{i-1} for $i = 1, 2, \ldots, k$. Also by Exercise 6.6, we obtain a normal series for G/H:

$$1 = H/H = G_k/H \lhd G_{k+1}/H \lhd \cdots \lhd G_{\ell-1}/H \lhd G_\ell/H = G/H.$$

By the Third Isomorphism Theorem (Exercise 6.9), the factors in this series are

$$(G_i/H)\big/(G_{i-1}/H) \cong G_i/G_{i-1} \quad \text{for } k < i \leq \ell,$$

all of which are simple, so we have a composition series for G/H. Thus the composition factors of H, together with those of G/H, make up those of G, counting multiplicity. The result is then immediate. □

Exercises 13.

1. Prove that every abelian group is solvable.
2. Prove that every dihedral group is solvable.
3. Prove that every p-group is solvable.
4. Prove that S_n is solvable iff $n \leq 4$.

Chapter II

Rings

14. Definitions and Examples

Unfortunately there are many definitions to absorb in this new topic! But fortunately, the definitions are reasonable, and are supported by examples, which we shall supply as soon as possible.

A **ring** is a set R with *two* binary operations, usually addition (denoted by '+') and multiplication (denoted by juxtaposition of elements), such that

(i) R is an abelian group under addition, with additive identity denoted by 0;
(ii) R has associative multiplication, i.e. $(ab)c = a(bc)$ for all $a, b, c \in R$; and
(iii) multiplication is distributive over addition, i.e. $a(b+c) = ab+ac$ and $(a+b)c = ac+bc$ for all $a, b, c \in R$.

If $ab = ba$ for all $a, b \in R$, we say that R is **commutative**. If R has a two-sided multiplicative identity $1 \in R$ with $1 \neq 0$, then R is a **ring with unity**, or a **ring with identity**. (This is often called a *ring with unit,* which is somewhat misleading, since *units,* as we are about to define shortly, include more than just 1.) We require $1 \neq 0$ in this case, in order to eliminate the **trivial ring** $R = 0$. It is easy to see that if a two-sided identity exists, then it is unique (by the same trick as in Section 1 for groups). *Warning:* Many authors use the term 'ring' to refer to a commutative ring with identity. When referring to a book for results on 'rings', be sure to check carefully at the beginning of the section, chapter or book to see what the author means by this terminology.

The additive inverse of $a \in R$ is denoted by $-a$, and $b - a$ means $b + (-a)$.

14.1 Proposition. For all $a, b \in R$, we have
(i) $0a = a0 = 0$, and
(ii) $(-a)b = a(-b) = -(ab)$, so we denote this simply by $-ab$.

Proof. We have $0a = (0+0)a = 0a+0a$, and similarly on the other side, which proves (i). Also $0 = 0b = (a + (-a))b = ab + (-a)b$, which proves that $(-a)b = -(ab)$; similarly, $a(-b) = -(ab)$. □

More generally, we define integer multiples of ring elements as follows: for each positive integer k and each $a \in R$, we define

$$ka = \overbrace{a + a + \cdots + a}^{k \text{ times}} \in R;$$

and for each negative integer k, define $ka = -(|k|a)$ where $|k|a$ is defined as above. It is easy to check that

$$k(\ell a) = (k\ell)a, \qquad (ka)b = k(ab),$$

$$(k+\ell)a = ka + \ell a, \qquad k(a+b) = ka + kb$$

for all $k, \ell \in \mathbb{Z}$ and $a, b \in R$. If it happens that the integer k is in R as well as in $\mathbb{Z}$, there is no need to worry about whether the product ka should be regarded as a product in R or as defined above, for in such cases, both answers must agree. For each $a \in R$ we write $\mathbb{Z}a = \{ka : k \in \mathbb{Z}\}$; clearly this is the additive subgroup of R generated by a.

Let R be a ring. If $a, b \in R$ such that $a, b \neq 0$ but $ab = 0$, then a and b are called **zero divisors**. If R is a ring with unity, we say an element $a \in R$ is **invertible** (or that a is a **unit**) if $ab = ba = 1$ for some $b \in R$. If such an inverse exists, it is easy to see that it is unique, and so we write it as b^{-1}. Note that no unit is a zero divisor; if $bc = 0$ where b is a unit, then left-multiplication by b^{-1} gives $c = 0$. It is easy to show that in any ring with unity, the set of units forms a multiplicative group. This group is called the **unit group** of R, and is denoted $R^\times$ (or sometimes R^*).

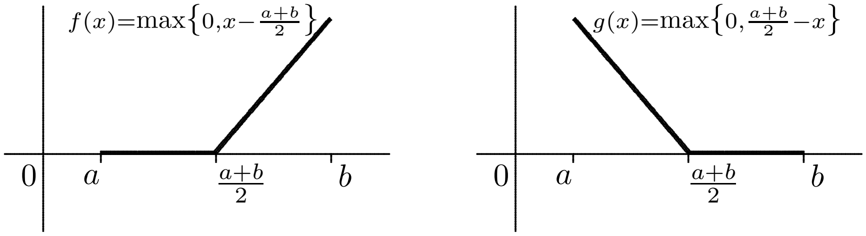
An **integral domain** is a commutative ring with unity and no zero divisors. A **skewfield** (or **division ring**) is a ring in which the nonzero elements form a multiplicative group. A **field** is a ring in which the nonzero elements form an *abelian* group. In other words, a commutative ring with unity is a field if *every* nonzero element is a unit; if we remove the requirement of commutativity, we get a skewfield. (*Warning:* Some older books use the term 'field' for skewfield.) Since units are not zero divisors, every field is an integral domain. Here is a picture showing all implications between the classes of rings we have defined:

$$\begin{tikzcd}
& \text{integral domain} \arrow[r, Rightarrow] & \text{commutative ring with unity} \arrow[r, Rightarrow] \arrow[dd, Rightarrow] & \text{commutative ring} \arrow[dd, Rightarrow] \\
\text{field} \arrow[ur, Rightarrow] \arrow[dr, Rightarrow] & & & \\
& \text{skewfield} \arrow[r, Rightarrow] & \text{ring with unity} \arrow[r, Rightarrow] & \text{ring}
\end{tikzcd}$$

None of these implications is reversible.

Two rings R and S are **isomorphic** (written $R \cong S$) if there exists an **isomorphism** $\phi : R \to S$, i.e. a bijection $\phi : R \to S$ such that $\phi(x+y) = \phi(x) + \phi(y)$ and $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in R$. Isomorphism is clearly an equivalence relation on the class of all rings, and two rings in the same isomorphism class will be considered essentially the same.

Examples. Some well-known fields include $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{Q}$. The ring of integers, $\mathbb{Z}$, is an integral domain, whose group of units is $\{1, -1\}$. The set of all continuous real-valued functions defined on the interval $[a, b]$, under the operations of pointwise addition and multiplication of functions, is a commutative ring with unity, denoted $C([a, b])$. Clearly $C([a, b])$ has zero divisors, for example $fg = 0$ where f and g are nonzero elements of $C([a, b])$ with graphs as shown:



The group of units of $C([a, b])$ is $\{f \in C([a, b]) : f(x) \neq 0 \text{ for all } x \in [a, b]\}$. Note that if pointwise multiplication is replaced by composition, then we don't get a ring since $f \circ (g + h) \neq f \circ g + f \circ h$.

If R is any ring, then we have the ring of $n \times n$ matrices over R,

$$R^{n\times n} = \left\{ \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} : a_{ij} \in R \right\},$$

under the usual matrix addition and multiplication. (Sometimes $R^{n\times n}$ is denoted $M_n(R)$.) If R has unity 1, then $R^{n\times n}$ has identity given by the usual $n \times n$ identity matrix I. Typically $R^{n\times n}$ is noncommutative; for example $F^{n\times n}$ is noncommutative for every field F and all $n \geq 1$. Also if $n \geq 2$ and $R \neq 0$, then $R^{n\times n}$ has zero divisors, for example

$$\left(\begin{smallmatrix} a & 0 \\ 0 & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & 0 \\ 0 & a \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$$

where $0 \neq a \in R$. The group of units of $R^{n\times n}$ is $GL_n(R)$; see Exercise 3.7.

The ring of integers modulo n is

$$\mathbb{Z}/n\mathbb{Z} = \{0, 1, 2, \ldots, n{-}1\}$$

with addition and multiplication modulo n. This is a commutative ring with unity 1. For example in $\mathbb{Z}/6\mathbb{Z}$, we have $2+3 = 5$, $4+3 = 1$, $5\cdot 2 = 4$, $2\cdot 3 = 0 = 3\cdot 4$. In particular, we have zero divisors $2, 3, 4 \in \mathbb{Z}/6\mathbb{Z}$. The group of units is $(\mathbb{Z}/6\mathbb{Z})^\times = \{1, 5\}$. Note that our symbols for elements of $\mathbb{Z}/n\mathbb{Z}$ are ambiguous unless the value of n is clear from context. (*Warning:* Some authors write $\mathbb{Z}_n$ in place of $\mathbb{Z}/n\mathbb{Z}$. This unfortunate practice conflicts with the usage of $\mathbb{Z}_p$ for the ring of p-adic integers, which we will not define here. The notation $\mathbb{Z}/n\mathbb{Z}$ is standard notation for a quotient ring, which we will define in Section 15.)

Every finite field has size equal to a power of some prime p; see Exercise 18.5. Moreover, for every $q = p^e$ where p is prime and $e \geq 1$, there exists (up to isomorphism) a unique field with exactly q elements, called the **finite field** (or **Galois field**) of **order** q, denoted by $\mathbb{F}_q$ or $GF(q)$. When $q = p$ is prime, this is nothing other than $\mathbb{Z}/p\mathbb{Z}$; however for $q = p^2, p^3, \ldots$, the ring $\mathbb{Z}/q\mathbb{Z}$ has zero divisors, and so it is quite different from the field $\mathbb{F}_q$.

If R is any ring and X is an indeterminate (i.e. a symbol with no numerical value), then we have the **polynomial ring** $R[X]$ consisting of all polynomials $p(X) = a_0 + a_1X + a_2X^2 + \cdots + a_kX^k$ where $a_0, a_1, \ldots, a_k \in R$, $k \geq 0$, with the usual addition and multiplication of polynomials. In this case R is usually commutative with unity 1, in which case $R[X]$ is commutative with unity 1 (= the constant polynomial 1). Note that X commutes with every element of R. The additive identity of $R[X]$ is the **zero polynomial**, which is the polynomial all of whose coefficients are zero. The **degree** of any nonzero polynomial $p(X)$, denoted $\deg p(X)$, is the largest k such that the coefficient of X^k in $p(X)$ is nonzero. We often extend this definition to say that the zero polynomial has degree $-\infty$; this convenience allows us to state Theorem 14.2 below, for *all* polynomials, including the zero polynomial. We also define a **constant polynomial** to be a polynomial of degree ≤ 0, i.e. an element of R interpreted as a polynomial with no X^k terms for $k \geq 1$.

14.2 Theorem. If R is an integral domain, then so is $R[X]$, and $\deg\big(f(X)g(X)\big) = \deg f(X) + \deg g(X)$ for all $f(X), g(X) \in R[X]$.

Proof. We may suppose that $f(X)$ and $g(X)$ are not both zero; otherwise the result holds with the usual convention that $-\infty + k = -\infty$ whenever $k \in \mathbb{Z} \cup \{-\infty\}$. Thus we may write $f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_kX^k$ and $g(X) = b_0 + b_1X + b_2X^2 + \cdots + b_\ell X^\ell$ where $a_k\,, b_\ell \neq 0$. Then

$$f(X)g(X) = a_0b_0 + (a_0b_1 + a_1b_0)X + \cdots + a_kb_\ell X^{k+\ell}.$$

If R is an integral domain then $a_kb_\ell \neq 0$, in which case $\deg\big(f(X)g(X)\big) = k + \ell$ and $f(X)g(X) \neq 0$, so that $R[X]$ has no zero divisors. □

It is important to realize that each element of $R[X]$ is simply a *formal expression* of the form $p(X) = a_0 + a_1X + a_2X^2 + \cdots + a_kX^k$, not to be confused with a function $R \to R$. To make this point clear, consider the case $R = \mathbb{F}_2 = \{0, 1\}$, which is the field of integers modulo 2. The polynomial $p(X) = X + X^2 \in \mathbb{F}_2[X]$ represents the zero function $R \to R$, and yet $f(X)$ is by definition *not* the zero polynomial, since not all its coefficients are zero! So in the proof above, the conclusion $f(X)g(X) \neq 0$ does *not* mean that $f(X)g(X)$ has no zeroes in R; *nor* does it mean that the function $R \to R$ is not identically zero. Rather, it means that $f(X)g(X)$ is not the zero polynomial, i.e. that $f(X)g(X)$ has at least one nonzero coefficient. Moreover, $1+X+X^2 \in \mathbb{F}_2[X]$ is by definition not a constant polynomial, although it represents a constant function $\mathbb{F}_2 \to \mathbb{F}_2$. And finally, when we write $f(X) = g(X)$, we mean simply that they are the same polynomial, i.e. that all their corresponding coefficients are equal, which is stronger than saying that they represent the same function $R \to R$. For example, $X \neq X^2$, even though they represent the same function $\mathbb{F}_2 \to \mathbb{F}_2$.

If you are used to working only with the fields $\mathbb{R}$ and $\mathbb{C}$, you might be a little shaken up, or at least a little confused, by these examples, as I once was. If so, I hope you will sleep better tonight if I assure you that two polynomials $f(X), g(X) \in \mathbb{R}[X]$ represent the same function (i.e. $f(a) = g(a)$ for all $a \in \mathbb{R}$), then they are the same polynomial (i.e. all corresponding coefficients are the same). This is more generally true with any infinite coefficient field in place of $\mathbb{R}$. Yet this is not obvious, and needs to be proved! (See Exercise 15.9.) Until it is clear to you how to think of a polynomial as distinct from a function, or how to think of $a_0+a_1X+a_2X^2+\cdots+a_kX^k$ as a formal expression, you would be better to treat elements of $R[X]$ as finite sequences over R of the form $(a_0, a_1, a_2, \ldots, a_k)$, of arbitrary length. Addition is componentwise (after the shorter sequence has been padded with 0's on the right until the sequences have the same length). And multiplication is defined by

$$(a_0, a_1, a_2, \ldots, a_k)(b_0, b_1, b_2, \ldots, b_\ell) = (a_0b_0, a_0b_1+a_1b_0, \ldots, a_{k-1}b_\ell + a_kb_{\ell-1}, a_kb_\ell).$$

Padding 0's on the right of any sequence doesn't change its meaning, so $(a_0, a_1, \ldots, a_k) = (a_0, a_1, \ldots, a_k, 0, 0, \ldots, 0)$. Of course the use of an indeterminate X makes the notation much more readable and the rule for multiplication much easier to remember; yet the X serves merely as a placeholder, and the expressions involving X have exactly the same meaning as these sequences we have just introduced. And that's what polynomials *really are,* as distinct from whatever functions they might represent.

Now let's consider the ring of polynomials in n indeterminates $X_1, X_2, \ldots, X_n$ and coefficients in R, denoted $R[X_1, X_2, \ldots, X_n]$. Of course the indeterminates $X_1, X_2, \ldots, X_n$ commute with each other, as well as with every element of R. Clearly, every $f(X_1, X_2, \ldots, X_n) \in R[X_1, X_2, \ldots, X_n]$ may be uniquely expressed as a polynomial in X_n with coefficients in $R[X_1, X_2, \ldots, X_{n-1}]$; this gives an isomorphism

$$R[X_1, X_2, \ldots, X_n] \cong (R[X_1, X_2, \ldots, X_{n-1}])[X_n].$$

So by induction, using Theorem 14.2 we obtain

14.3 Theorem. If R is an integral domain, then so is $R[X_1, X_2, \ldots, X_n]$.

I still owe you an example of a noncommutative skewfield. It is a fact (Wedderburn's Theorem) that every finite skewfield is a field, so our example will necessarily be infinite. Everyone's favorite example is the set $\mathbb{H}$ of **real quaternions**, which are expressions of the form

$$x = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}, \qquad a, b, c, d \in \mathbb{R},$$

so that $\mathbb{H}$ is a 4-dimensional vector space over $\mathbb{R}$ with basis $\{1, \mathbf{i}, \mathbf{j}, \mathbf{k}\}$, and the usual componentwise vector addition. Multiplication of real quaternions is associative, and works according to the famous rules of Hamilton:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1.$$

Furthermore every real number commutes with $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$. Then $\mathbb{H}$ is a ring with unity 1, but $\mathbb{H}$ is not commutative, since for example, $\mathbf{ij} = \mathbf{k} = -\mathbf{ji}$, as you should verify using the rules above. Define the **conjugate** of an arbitrary $x = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \in \mathbb{H}$ by

$$\overline{x} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k},$$

and the **norm** of x by

$$\|x\| = \sqrt{x\overline{x}} = \sqrt{a^2 + b^2 + c^2 + d^2}\,.$$

This is just the usual Euclidean norm when $\mathbb{H}$ is identified naturally with the real vector space $\mathbb{R}^4$. One checks that $\overline{xy} = \overline{y}\,\overline{x}$ for all $x, y \in \mathbb{H}$, and from this it follows that $\|xy\| = \|x\| \cdot \|y\|$ for all $x, y \in \mathbb{H}$. From the identity $x\overline{x} = \|x\|^2$ we see that every nonzero real quaternion is a unit; indeed if $x = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \neq 0$, then

$$x^{-1} = \frac{\overline{x}}{\|x\|^2} = \frac{a}{\|x\|^2} - \frac{b}{\|x\|^2}\mathbf{i} - \frac{c}{\|x\|^2}\mathbf{j} - \frac{d}{\|x\|^2}\mathbf{k}\,.$$

A subset S of a ring R which is itself a ring with respect to the binary operations of R restricted to S, is called a **subring** of R. It is easy to check that a nonempty subset $S \subseteq R$ is a subring iff S is closed under products and differences, i.e. if $ab,\ a - b \in S$ whenever $a, b \in S$. For example, $\mathbb{Z}$ is a subring of $\mathbb{Q}$, which is a subring of $\mathbb{R}$, which is a subring of $\mathbb{C}$. It is possible to see $\mathbb{C}$ as a subring of $\mathbb{H}$ in several different ways; for example,

$$\{a + b\mathbf{i} : a, b \in \mathbb{R}\} \cong \{a + c\mathbf{j} : a, c \in \mathbb{R}\} \cong \{a + d\mathbf{k} : a, d \in \mathbb{R}\} \cong \mathbb{C}.$$

The set of all constant polynomials in $R[X]$ is a subring isomorphic to R, so we naturally identify R with this subring. The matrix ring $R^{n\times n}$ has many subrings; for example the set of diagonal matrices; or the set of upper-triangular matrices. The set of all **scalar matrices** (matrices of the form $aI,\ a \in R$) is a subring isomorphic to R.

If a field E has a subring $F \subseteq E$ which is also a field, then we say that F is a **subfield** of E, or that E is an **extension** of F. It follows directly from the axioms in this case that E is a vector space over F. The dimension of this vector space is known as the **degree** of the extension $E \supseteq F$, and this number (finite or infinite) is denoted $[E : F]$. We have for example that $[\mathbb{R} : \mathbb{Q}] = \infty$; see Exercise 14.11. Extensions of degree 2, 3, 4, 5 are called **quadratic**, **cubic**, **quartic**, **quintic** respectively. For example the extension $\mathbb{C} \supset \mathbb{R}$ is quadratic since $\{1, i\}$ is a basis for $\mathbb{C}$ over $\mathbb{R}$. A **tower** of extension fields is a sequence of extensions $F_1 \supseteq F_2 \supseteq \cdots \supseteq F_k$.

If $\alpha \in E \supseteq F$, then we define

$$F[\alpha] = \{f(\alpha) : f(X) \in F[X]\}.$$

Clearly $F[\alpha]$ is the smallest subring of E containing both F and α. We have $E \supseteq F[\alpha] \supseteq F$, but $F[\alpha]$ may not be a field. The smallest *subfield* of E containing both F and α is the subfield is

$$F(\alpha) = \left\{ \frac{g(\alpha)}{h(\alpha)} \;:\; g(X), h(X) \in F[X],\ h(\alpha) \neq 0 \right\}.$$

We now have $E \supseteq F(\alpha) \supseteq F[\alpha] \supseteq F$ in which $E \supseteq F(\alpha) \supseteq F$ is a tower of extension fields. Notice that the construction of the field $F(\alpha)$ from the ring $F[\alpha]$ closely resembles the construction of the field $\mathbb{Q}$ from the ring $\mathbb{Z}$. We call $F(\alpha)$ the *quotient field* of $F[\alpha]$. In a similar way, every integral domain is extendible to its quotient field; see Exercise 14.13. We similarly define the ring $F[\alpha_1, \ldots, \alpha_k]$ and its quotient field $F(\alpha_1, \ldots, \alpha_k)$.

It frequently happens that the ring $F[\alpha]$ is *already* a field, in which case its quotient field $F(\alpha) = F[\alpha]$. For example, we have $\mathbb{R} \supseteq \mathbb{Q}(\sqrt{5}) \supseteq \mathbb{Q}[\sqrt{5}] \supseteq \mathbb{Q}$, where $\mathbb{Q}[\sqrt{5}] = \{a + b\sqrt{5} : a, b \in \mathbb{Q}\}$. We note that $\mathbb{Q}[\sqrt{5}]$ is a field. To verify this, we only need to check that it is closed under division. This follows from the familiar process of *rationalizing the denominator;* for example, here we divide two typical elements of $\mathbb{Q}[\sqrt{5}]$:

$$\frac{\frac{3}{2} - \frac{1}{3}\sqrt{5}}{\frac{3}{4} + \frac{1}{2}\sqrt{5}} = \frac{18 - 4\sqrt{5}}{9 + 6\sqrt{5}} \cdot \frac{9 - 6\sqrt{5}}{9 - 6\sqrt{5}} = \frac{282 - 144\sqrt{5}}{81 - 180} = -\tfrac{94}{33} + \tfrac{16}{11}\sqrt{5}.$$

Thus $\mathbb{Q}(\sqrt{5}) = \mathbb{Q}[\sqrt{5}] = \{a + b\sqrt{5} : a, b \in \mathbb{Q}\}$, is a quadratic extension of $\mathbb{Q}$, having basis $\{1, \sqrt{5}\}$, i.e. $[\mathbb{Q}[\sqrt{5}] : \mathbb{Q}] = 2$.

Finally, for an example to show that $F[\alpha]$ need not be a field, see Exercise 14.12. The precise conditions under which $F(\alpha) = F[\alpha]$ will be presented in Theorem 18.2.

Exercises 14.

1. Let R be a ring with unity. Show that the units of $R[X]$ are just the units of R.

2. Let R and S be rings. Define $R \oplus S = \{(r, s) : r \in R,\ s \in S\}$ with addition and multiplication defined componentwise by

$$(r, s) + (r', s') = (r{+}r', s{+}s'); \qquad (r, s)(r', s') = (rr', ss').$$

   (a) Show that $R \oplus S$ is a ring. This is known as the **direct sum** of R and S.

   (b) If R and S are rings with unity, show that $R \oplus S$ is also a ring with unity, and that its unit group is given by $(R \oplus S)^\times = R^\times \times S^\times$.

3. Let R be a ring. Define the **center** of R by

$$Z(R) = \{a \in R : ra = ar \text{ for all } r \in R\},$$

and the **centralizer** of an element $r \in R$ by

$$C_R(r) = \{a \in R : ar = ra\}.$$

Show that $Z(R)$ and $C_R(r)$ are subrings of R.

4. Show that the unit group of $\mathbb{Z}^{n\times n}$ (the ring of all $n \times n$ matrices with integer entries) is the group

$$GL_n(\mathbb{Z}) = \{A \in \mathbb{Z}^{n\times n} : \det(A) = \pm 1\}.$$

5. Can a ring equal the union of two of its proper subrings? Give an easy example, or prove that this is not possible.

6. Let R be any ring, and let $S = R^{2\times 2}$. Show that the ring $S^{2\times 2}$ is naturally isomorphic to $R^{4\times 4}$. Try to generalize this result.

7. (a) Show that
$$\left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}$$
is a subring of $\mathbb{R}^{2\times 2}$ isomorphic to $\mathbb{C}$.

   (b) Use similar ideas to show that $\mathbb{H}$ is isomorphic to a subring of $\mathbb{R}^{4\times 4}$, and isomorphic to a subring of $\mathbb{C}^{2\times 2}$.

8. Let R be a nontrivial ring. Show that
$$S = \left\{ \begin{pmatrix} a & b & c \\ c & a & b \\ b & c & a \end{pmatrix} : a, b, c \in R \right\}$$
is a subring of $R^{3\times 3}$, and that S has zero divisors.

9. Prove that no two of the rings $\mathbb{Z}$, $\mathbb{Z}[X]$, $\mathbb{Z}[X, Y]$ are isomorphic. Is there a ring R such that $R[X] \cong R$? Justify your answer.

10. Consider the ring $R = \{a + bI + cJ + dK : a, b, c, d \in \mathbb{C}\}$ where $I^2 = J^2 = K^2 = IJK = -1$. Show that R is *not* a skewfield.

11. Prove that $[\mathbb{R} : \mathbb{Q}] = \infty$.

    Hint: Show that for every integer $n \geq 1$, the vector space $\mathbb{Q}^n$ is countably infinite.

12. Show that $\frac{1}{\pi} \notin \mathbb{Q}[\pi]$ and conclude that $\mathbb{Q}(\pi) \supsetneq \mathbb{Q}[\pi]$.

    Hint: You man assume the fact that π is not a zero of any nonzero $f(X) \in \mathbb{Q}[X]$. This is the statement that π is *transcendental;* more about this in Section 18.

13. The following exercise shows that every integral domain R is a subring of some field. In examples we have seen so far, this was clear since R was chosen as a subring of a known field. But given *any* integral domain R, we can nevertheless construct a field containing R, and whose elements are simply the quotients of the original ring R. The construction follows:

    Let R be an integral domain. For $a, b, c, d \in R$, we say that $(a, b) \sim (c, d)$ iff $ad = bc$. This defines a relation on $R \times R$.

    (a) Prove that $\sim$ gives an equivalence relation on the set $S = \{(x, y) \in R \times R : y \neq 0\}$.

    (b) Let K denote the set of equivalence classes, and denote by $\frac{a}{b}$ the equivalence class of (a, b). Define addition and multiplication on K by
    $$\tfrac{a}{b} + \tfrac{c}{d} = \tfrac{ad+bc}{bc}, \qquad \tfrac{a}{b} \cdot \tfrac{c}{d} = \tfrac{ac}{bd}\,.$$
    Prove that K is a field. We call K the **quotient field** of R, and this construction naturally generalizes the construction of $\mathbb{Q}$ from $\mathbb{Z}$.

    (c) Prove that the set of all $\frac{a}{1}$ with $a \in R$ is a subring of K, naturally isomorphic to R. Thus every integral domain embeds naturally in its quotient field.

    Example: For any integral domain R, the polynomial ring $R[X]$ is also an integral domain by Theorem 14.2. The quotient field of $R[X]$, denoted $R(X)$, is the set of all **rational functions** $f(X)/g(X)$, where $f(X), g(X) \in R[X]$ with $g(X) \neq 0$.

14. Let F be a field, and define a **(formal) power series** in X with coefficients in F, to be an expression of the form $a_0 + a_1X + a_2X^2 + \cdots$ where $a_i \in F$ for all $i \geq 0$. The set of all such power series forms a ring, denoted $F[[X]]$, having $F[X]$ as a subring. Like polynomials, these are strictly formal objects, not functions (and so questions of convergence need not arise). More generally, a **(formal) Laurent series** in X with coefficients in F, is an expression of the form $a_kX^k + a_{k+1}X^{k+1} + a_{k+2}X^{k+2} + \cdots$ where $k \in \mathbb{Z}$ and $a_i \in F$ for all $i \geq k$. The set of all such Laurent series forms a ring, denoted $F((X))$.
   (a) Show that $F((X))$ is a field, and that it is in fact (up to isomorphism) the quotient field of $F[[X]]$; see Exercise 14.13. Thus we have the following subrings of $F((X))$:

$$\begin{tikzcd}
 & F((X)) \arrow[dl, dash] \arrow[dr, dash] & \\
F(X) \arrow[dr, dash] & & F[[X]] \arrow[dl, dash] \\
 & F[X] &
\end{tikzcd}$$

   For example, we have the rational function $X/(1+X) = X - X^2 + X^3 - X^4 + \cdots \in F(X)$.
   (b) Let $F = \mathbb{F}_2$ and consider the rational function $f(X) = (1+X+X^3)/(X^2+X^3+X^4) \in \mathbb{F}_2(X)$. Expand $f(X)$ in its Laurent series, showing all terms up to and including the X^5 term.

15. Ideals and Quotients

Let R be a ring. An **ideal** of R is a subring which is invariant under left- and right-multiplication by elements of R. That is, a subset $A \subseteq R$ is an ideal iff

(i) $A \neq \emptyset$;
(ii) $a - b \in A$ for all $a, b \in R$, and
(iii) $ra,\ ar \in A$ for all $a \in A$ and $r \in R$.

For example, every ring is an ideal of itself, and every ring has the trivial ideal $\{0\}$. The ring $\mathbb{Z}$ of integers has ideals of the form $m\mathbb{Z} = \{ma : a \in \mathbb{Z}\}$ for each $m \in \mathbb{Z}$.

An ideal A of a ring R is **proper** if $A \subsetneq R$. It is clear (but worthy of explicit mention) that if R has unity 1, then no proper ideal of R contains 1 (or for that matter, any units).

Given two ideals $A, B \subseteq R$, we define their **sum** as

$$A + B = \{a + b : a \in A,\ b \in B\}$$

and their product AB as the set of all sums $a_1b_1 + a_2b_2 + \cdots + a_kb_k$ where all $a_i \in A$ and $b_i \in B$.

15.1 Proposition. (i) If A and B are ideals of a ring R, then so are AB, $A+B$ and $A \cap B$.
(ii) If $\{A_\alpha\}_\alpha$ is a nonempty collection of ideals of R, then their intersection $\bigcap_\alpha A_\alpha$ is an ideal of R.

Warning: To define AB, in general it does not suffice to take just products ab with $a \in A$ and $b \in B$, for this does not always give an ideal. For example consider $R = \mathbb{Z}[X]$, which has ideals $A = \{f(X) \in \mathbb{Z}[X] : f(0) \text{ is even}\}$ and $B = \{f(X) \in \mathbb{Z}[X] : f(1) \text{ is even}\}$. Now AB contains $X(X+1) + (2-X)(X-1) = 4X-2$, which is *not* of the form $f(X)g(X)$ with $f(X) \in A$ and $g(X) \in B$.

Proof of Proposition 15.1. Suppose that A and B are ideals of R, and let $x = \sum_i a_i b_i \in AB$, $y = \sum_j a'_j b'_j \in AB$. Then $x - y = \sum_i a_i b_i + \sum_j (-a'_j) b'_j \in AB$ by definition. If also $r \in R$, then $rx = \sum (ra_i) b_i \in AB$ since every $ra_i \in A$; and similarly, $xr \in AB$. This proves that AB is an ideal of R. The proof that $A + B$ is an ideal, is left as an exercise. The fact that $A \cap B$ is an ideal is a special case of (ii), whose proof follows.

Let $\{A_\alpha\}_\alpha$ be a nonempty collection of ideals of R, and let $\mathcal{I} = \bigcap_\alpha A_\alpha$. Since 0 belongs to each A_α, we have $0 \in \mathcal{I}$, and so $\mathcal{I} \neq \emptyset$. If $x, y \in \mathcal{I}$ then x and y belong to each A_α, so $x - y$ belongs to each A_α, so $x - y \in \bigcap_\alpha A_\alpha = \mathcal{I}$. Similarly if $r \in R$ and $x \in \mathcal{I}$, then $rx, xr \in \mathcal{I}$. □

For example, given two ideals $m\mathbb{Z}$ and $n\mathbb{Z}$ in $\mathbb{Z}$, the sum, product and intersection are given by $m\mathbb{Z} + n\mathbb{Z} = gcd(m,n)\mathbb{Z}$, $(m\mathbb{Z})(n\mathbb{Z}) = (mn)\mathbb{Z}$, and $m\mathbb{Z} \cap n\mathbb{Z} = lcm(m,n)\mathbb{Z}$; see Exercise 15.2.

You should immediately observe the strong analogy between group theory and ring theory results: Many statements of ring theory results are obtained from group theory results simply by replacing the word ‘group’ by ’ring’; replacing ‘subgroup’ by ‘subring’; and replacing ‘normal subgroup’ by ‘ideal’. This should help considerably in absorbing these new results! Already you can see this pattern arising in Proposition 15.1, since the class of normal subgroups of a given group G is similarly closed under product and intersection (see Exercise 5.1).

Just as we formed quotients of groups modulo normal subgroups, we form quotients of rings modulo ideals. We have already encountered an example of this: the quotient ring $\mathbb{Z}/n\mathbb{Z}$. More generally, given a ring R and an ideal $A \subseteq R$, the **quotient ring**, denoted R/A, is the set of all cosets $r + A = \{r + a : a \in A\}$, for $r \in R$. (Because addition is commutative, it does not matter whether we write left cosets or right cosets.) Addition and multiplication of cosets is defined by

$$(r + A) + (s + A) = (r + s) + A, \qquad (r + A)(s + A) = rs + A.$$

Why are these operations well-defined? It is easy to see that addition of cosets is well-defined. (These are, after all, cosets of an additive subgroup of the abelian additive group of R, so this statement has already been proved.) But for the multiplication of cosets to be well-defined, we really need to use the fact that A is an *ideal* of R, not just an arbitrary subring. Suppose that $r + A = r' + A$ and $s + A = s' + A$. Then $r + 0 = r' + a$ for some $a \in A$, which says that $r - r' \in A$. Similarly, $s - s' \in A$. Now we have ‘defined’ $(r + A)(s + A)$ as $rs + A$, which equals $r's' + (r - r')s' + r(s - s') + A = r's' + A$ since $(r - r')s' \in A$ and $r(s - s') \in A$.

Are we justified in calling the set of cosets of A a quotient *ring*? Yes:

15.2 Theorem. If A is an ideal of R, then R/A is a ring with respect to the addition and multiplication defined above.

Proof. Let $r, s, t \in R$. Then

$$\begin{aligned}
&(r+A)+(s+A) = (r+s)+A = (s+r)+A = (s+A)+(r+A),\\
&(r+A)+A = (r+A)+(0+A) = (r+0)+A = r+A,\\
&(r+A)+((-r)+A) = (r+(-r))+A = 0+A = A,
\end{aligned}$$

which shows that R/A is an additive abelian group with identity $A = 0 + A$. Also

$$\begin{aligned}
(r+A)\big((s+A)+(t+A)\big) &= (r+A)\big((s+t)+A\big)\\
&= r(s+t)+A\\
&= (rs+rt)+A\\
&= (rs+A)+(rt+A)\\
&= (r+A)(s+A)+(r+A)(t+A),
\end{aligned}$$

which proves one of the distributive laws. The other distributive law, and the associativity of multiplication, are proved similarly. □

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

(i) $\phi(x+y) = \phi(x) + \phi(y)$, and
(ii) $\phi(xy) = \phi(x)\phi(y)$

for all $x, y \in R$. (We say that ϕ *preserves* both addition and multiplication.) The **kernel** and **image** of ϕ are

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

Given an ideal $A \subseteq R$, the **canonical homomorphism** from R to R/A is defined by

$$\pi : R \to R/A, \qquad r \mapsto r + A.$$

It is easy to check that π is in fact a homomorphism, and that $\ker\pi = A$, which shows that every ideal is the kernel of some homomorphism. Conversely, the kernel of any ring homomorphism is an ideal; this is part of Theorem 15.3 below. Note that these statements are the ring-theoretic analogues of statements from Section 6.

Of course, a bijective homomorphism is the same thing as an isomorphism. We now translate the three isomorphism theorems for groups (Theorem 6.1 and Exercises 6.7,8) into the language of ring theory.

15.3 First Isomorphism Theorem. If $\phi : R \to S$ is a homomorphism of rings, then $\ker\phi$ is an ideal of R, and $R/\ker\phi \cong \phi(R)$.

Proof. Let $K = \ker\phi$. We have $\phi(0) = \phi(0+0) = \phi(0)+\phi(0)$, which implies that $\phi(0) = 0$. This shows that $0 \in K$ and so $K \neq \emptyset$. If $x, y \in K$ then $\phi(x-y) = \phi(x)-\phi(y) = 0-0 = 0$, and so $x - y \in K$. If $x \in K$ and $r \in R$ then $\phi(rx) = \phi(r)\phi(x) = \phi(r)0 = 0$ and so $rx \in K$; similarly, $xr \in K$. This proves that K is an ideal of R.

Define $\overline{\phi} : R/K \to \phi(R)$ by $\overline{\phi}(r+K) = \phi(r)$. First we check that $\overline{\phi}$ is well-defined: If $r + K = r' + K$ then $r - r' \in K$, which says that $\phi(r - r') = 0$, and so $\overline{\phi}(r+K) = \phi(r) = \phi(r') = \overline{\phi}(r'+K)$. Thus $\overline{\phi}$ is well-defined.

Let $r, s \in R$. Then $\overline{\phi}\big((r+K)(s+K)\big) = \overline{\phi}(rs+K) = \phi(rs) = \phi(r)\phi(s) = \overline{\phi}(r+K)\overline{\phi}(s+K)$. Similarly, $\overline{\phi}\big((r+K)+(s+K)\big) = \overline{\phi}(r+K)+\overline{\phi}(s+K)$, and so $\overline{\phi}$ is a homomorphism.

Clearly $\overline{\phi}$ is onto $\phi(R)$, since each element of $\phi(R)$ has the form $\phi(r) = \overline{\phi}(r+K)$ for some $r \in R$. Also, if $\overline{\phi}(r+K) = 0$ then $\phi(r) = 0$, whence $r \in \ker\phi = K$ and $r+K = K$, so that $\ker\overline{\phi}$ contains only the identity $K \in R/K$, i.e. $\overline{\phi}$ is one-to-one. Thus $\overline{\phi}$ is an isomorphism. □

The proofs of the following two theorems are left as exercises.

15.4 Second Isomorphism Theorem. Let R be a ring with ideal A and subring S. Then $A \cap S$ is an ideal of S, and $S/(A\cap S) \cong (S+A)/A$.

15.5 Third Isomorphism Theorem. Let R be a ring with ideals $A \subseteq B$. Then B/A is an ideal of R/A, and ${}^{(R/A)}/_{(B/A)} \cong R/B$.

Given a subset $X \subseteq R$, we speak of the **ideal generated by** X, denoted by (X); this is the unique smallest ideal of R containing X. This ideal exists by Proposition 15.1(ii); it is the intersection of all ideals of R which contain X. Usually X is a finite subset, and if $X = \{x_1, x_2, \ldots, x_k\}$, we alternatively write $(X) = (x_1, x_2, \ldots, x_k)$. In case $X = \{x\}$ is a singleton subset, this gives an ideal (x) generated by the single element $x \in R$, known as a **principal ideal**. If R is a commutative ring with unity, then $(x) = Rx = \{rx : r \in R\}$ by Exercise 15.1. As an example, the ideal of $\mathbb{Z}$ generated by integers $m_1, m_2, \ldots, m_k$ is the same thing as the principal ideal generated by $m = \gcd(m_1, m_2, \ldots, m_k)$, so we have $(m_1, m_2, \ldots, m_k) = (m) = m\mathbb{Z}$.

The familiar 'modulus' notation for integers extends to arbitrary rings with ideals. Every ideal A of a ring R determines an equivalence relation on the elements of R: we say that $x \equiv y \mod A$ if $x - y \in A$. Note that the equivalence classes of this relation are just the elements of R/A. Also observe that for integers, we have $x \equiv y \mod (m)$ iff $x \equiv y \mod m$ (in the usual notation).

Every ideal of $\mathbb{Z}$ is principal. For suppose that $A \neq (0)$ is an ideal of $\mathbb{Z}$. Since A contains a nonzero integer, and since A is closed under multiplication by $-1 \in \mathbb{Z}$, there must be a positive integer in A. Let m be the smallest positive integer in A. By definition of an ideal, $(m) = m\mathbb{Z} \subseteq A$. This must be equality, for if $n \in A$, then we may write $n = qm + r$ for some integers q, r with $0 \leq r < m$; now $0 \leq r = n - qm \in A$, and since m is the smallest positive element of A, we obtain $r = 0$, i.e. $n = qm \in (m)$.

Define a **principal ideal ring** to be a ring in which every ideal is principal. Also a **principal ideal domain** (or **P.I.D.**) is a principal ideal ring which is an integral domain. We have just proved:

15.6 Proposition. $\mathbb{Z}$ is a principal ideal domain.

We wish to generalize Proposition 15.6 to a larger class of rings which includes polynomial rings $F[X]$ where F is an arbitrary field. Define a **Euclidean ring** as a commutative ring $R \neq 0$ together with a function $\deg : R \to \{-\infty, 0, 1, 2, 3, \ldots\}$ such that for all $a, b \in R$,

(i) $\deg(a) = -\infty$ iff $a = 0$;
(ii) $\deg(a) \leq \deg(ab)$; and
(iii) if $b \neq 0$, then there exist $q, r \in R$ such that $a = qb + r$ and $\deg(r) < \deg(b)$.

Property (iii) (and the usual algorithm for computing the *quotient* q and *remainder* r) are together known as the **Division Algorithm** for R. A Euclidean ring which is an integral domain is called a **Euclidean domain**. For example, $\mathbb{Z}$ is a Euclidean domain with $\deg(a) = |a|$ for nonzero integers a; and $\deg(0) = -\infty$. Also if F is any field, then the polynomial ring $F[X]$ is a Euclidean domain with the usual degree function. However, $\mathbb{Z}[X]$ is *not* a Euclidean domain; the usual degree function for $\mathbb{Z}[X]$ does not satisfy the Division Algorithm (iii). Of course this is not enough to prove that $\mathbb{Z}[X]$ is not a Euclidean ring, since it is still conceivable that a *different* degree function might work; see however Exercise 15.4.

15.7 Theorem. Every Euclidean ring is a principal ideal ring with unity.

15.8 Corollary. If F is a field, then $F[X]$ is a principal ideal domain.

Proof of Theorem 15.7. Let A be an ideal of a Euclidean ring R; we must show that A is principal. If $A = (0)$ then we are done. So assume that A has a nonzero element, and choose a nonzero element $a \in A$ for which $\deg(a)$ is minimal. Clearly $Ra \subseteq (a) \subseteq A$. Now let $x \in A$. We may write $x = qa + r$ for some $q, r \in R$ such that $\deg(r) < \deg(a)$. Then $r = x - qa \in A$, so by minimality of the degree of a, we must have $r = 0$, which says that $x = qa \in Ra$. Now $A \subseteq Ra \subseteq (a) \subseteq A$, so equality must hold: $A = Ra = (a)$. In particular, R is a principal ideal ring.

Since R is an ideal of itself, we have $R = Ru = (u)$ for some nonzero $u \in R$. In particular, we have $u = eu = ue$ for some $e \in R$. For every $x \in R$, we have $x = bu$ for some $b \in R$, and so $xe = bue = bu = x$, so e is a multiplicative identity for R. Finally, $e \neq 0$ since $u \neq 0$, so R is a ring with unity. □

Exercises 15.

1. (a) Let R be a commutative ring, and let $a \in R$. Show that $(a) = \mathbb{Z}a + Ra$, i.e. $(a) = \{ka + ra : k \in \mathbb{Z},\ r \in R\}$.

   (b) Conclude that if R is a commutative ring with unity, then $(a) = Ra$.

2. Let R be an integral domain. Given $a, b \in R$, we say a **divides** b (or a is a **divisor** of b, denoted $a\,|\,b$) if $b = ca$ for some $c \in R$. Prove that the following three conditions are equivalent:

   (i) $a\,|\,b$ and $b\,|\,a$;

   (ii) $a = ub$ for some unit $u \in R^\times$;

   (iii) $(a) = (b)$.

   Elements a and b having the above relation are called **associates**. This relation is an equivalence relation on R, as is clear from (iii). For example, the associates of $10 \in \mathbb{Z}$ are 10 and -10.

3. Let R be a Euclidean domain, and let $a, b \in R$. An element $g \in R$ is a **common divisor** of a and b if g divides both a and b. We say g is a **gcd (greatest common divisor)** of a and b if (i) g is a common divisor of a and b, and (ii) $\deg g \geq \deg h$ for every common divisor h of a and b. In general, the gcd is *not* unique.

   (a) Prove that if $a, b \in R$ are not both zero, then a and b have a gcd g; that the associates of g are *all* the gcd's of a and b; and that $(a) + (b) = (g)$. (See Exercise 15.2.)

   Hint: Since R is a Euclidean domain, $(a) + (b) = (g)$ for some $g \in R$. Prove that g has the required properties.

   (b) In the notation of (a), conclude that $g = as + bt$ for some $s, t \in R$. This fact (and the well-known algorithm for determining such $s, t \in R$, generalizing the example of Section 0) are together known as **Euclid's Algorithm** or the **Euclidean Algorithm**.

   (c) With a, b, g as above, prove that the common divisors of a and b are precisely the divisors of g.

4. Show that $\mathbb{Z}[X]$ is not a P.I.D., and hence by Theorem 15.7, $\mathbb{Z}[X]$ is not a Euclidean ring.

5. Prove the Second Isomorphism Theorem 15.4.

   Hint: It may help to imitate Exercise 6.7.

6. Prove the Third Isomorphism Theorem 15.5.

   Hint: It may help to imitate Exercise 6.8.

7. Let F be a field. We say that a constant $a \in F$ is a **zero** of a polynomial $f(X) \in F[X]$ if $f(a) = 0$. Given $a \in F$, show that a is a zero of $f(X)$ iff $f(X) = (X - a)q(X)$ for some $q(X) \in F[X]$.

   Hint: In order to show that $f(X)$ is divisible by $X - a$, first divide $f(X)$ by $X - a$ and obtain a quotient and remainder using the Division Algorithm.

   Remark: More generally, a **zero** of $f(X) \in F[X]$ may refer to an element a in an *extension* field $E \supseteq F$ such that $f(a) = 0$.

8. Let $f(X) \in F[X]$ be a polynomial of degree n where F is a field, and let $E \supseteq F$ be an extension field. Show that $f(X)$ has at most n zeroes in E.

   Hint: Use Exercise 15.7.

9. Let $f(X) \in F[X]$ where F is an infinite field. Show that $f(X)$ represents the zero function $F \to F$ iff $f(X) = 0$. This means that the only polynomial vanishing at every field element is the zero polynomial.

10. Let F be a field, and let G be a finite subgroup of the multiplicative group of $F^\times = F \smallsetminus \{0\}$. Prove that G is cyclic.

    Hint: By Exercise 6.15, $G \cong C_{n_1} \times C_{n_2} \times \cdots \times C_{n_r}$ for some positive integers $n_1, n_2, \ldots, n_r$. If $\gcd(n_i, n_j) \neq 1$ for some $i \neq j$, then choose a prime p dividing both n_i and n_j and count the number of zeroes of $X^p - 1 \in F[X]$ to obtain a contradiction. Then use Exercise 3.3.

16. Maximal and Prime Ideals

An ideal P of a ring R is **prime** if

(i) $P \subsetneq R$, and
(ii) for any two ideals $A, B \subseteq R$, if $AB \subseteq P$ then either $A \subseteq P$ or $B \subseteq P$.

The prime ideals of $\mathbb{Z}$ are (0) and (p) for p prime (in the usual sense of 'prime'). This follows easily from Proposition 15.6, and the fact that $(m)(n) = (mn)$ for ideals of $\mathbb{Z}$.

A useful characterization of prime ideals is the following

16.1 Proposition. Let P be an ideal of a ring R.

(i) Suppose that $ab \in P$ implies that either $a \in P$ or $b \in P$. Then P is prime.
(ii) If R is commutative, then the converse of (i) holds.

Proof. (i) Suppose that A and B are ideals of R such that $AB \subseteq P$ with $A \not\subseteq P$. Choose $a \in A \smallsetminus P$. For every $b \in B$ we have $ab \in AB \subseteq P$, so that $b \in P$; hence $B \subseteq P$. Thus P is prime.
(ii) Suppose that R is commutative with a prime ideal $P \subset R$, and that $ab \in P$ where $a, b \in R$. By Exercise 15.1, $(a) = \mathbb{Z}a + Ra = \{na + ra : n \in \mathbb{Z},\ r \in R\}$. Similarly, $(b) = \mathbb{Z}b + Rb$. From this it is not hard to see that $(a)(b) \subseteq (ab) \subseteq P$. Thus either $(a) \subseteq P$ (in which case $a \in P$), or $(b) \subseteq P$ (in which case $b \in P$). □

An ideal M of a ring R is **maximal** if

(i) $M \subsetneq R$, and
(ii) there is no ideal A such that $M \subsetneq A \subsetneq R$.

16.2 Proposition. Let R be a ring with unity. Then every maximal ideal of R is prime.

Proof. Let M be a maximal ideal of R, and suppose that $AB \subseteq M$, $A \not\subseteq M$, $B \not\subseteq M$ for certain ideals A, B of R. Then $M \subsetneq A + M$, so by maximality of M, we have $A + M = R$, and similarly $B + M = R$. Thus $1 = a + m = b + m'$ for some $a \in A$, $b \in B$ and $m, m' \in M$. Now $1 = (a + m)(b + m') = ab + am' + mb + mm' \in M$, contradicting $M \subsetneq R$. □

Every maximal ideal of $\mathbb{Z}$ is of the form (p) for p prime. So every *nonzero* prime ideal of $\mathbb{Z}$ is maximal. It is *not* generally true, however, that nonzero prime ideals are necessarily maximal, even for commutative rings with identity. For example, in the ring $\mathbb{Z} \oplus \mathbb{Z}$ with componentwise addition and multiplication, the nonzero ideal $\mathbb{Z} \oplus 0 = \{(x, 0) : x \in \mathbb{Z}\}$ is prime but not maximal, since it lies inside the larger proper ideal $\mathbb{Z}\oplus 2\mathbb{Z}$, which is maximal. The most important characterization of prime and maximal ideals is

16.3 Theorem. Let R be a commutative ring with unity, and let $A \subsetneq R$ be an ideal. Then

(i) A is prime iff R/A is an integral domain.
(ii) A is maximal iff R/A is a field.

Proof. (i) Suppose that A is a prime ideal of R. If $(x+A)(y+A) = 0 + A = A$, then by definition $xy + A = A$, so either $x \in A$ or $y \in A$, i.e. either $x + A = A$ or $y + A = A$. Thus R/A has no zero divisors. Since R has an identity 1, an identity for R/A is $1 + A$; and since R is commutative, so is R/A. Thus R/A is an integral domain. The converse is just as easy.
(ii) Suppose that A is a maximal ideal of R. By Proposition 16.2, A is a prime ideal, so R/A is an integral domain. To show that R/A is in fact a field, we must show that every nonzero element $x + A \in R/A$ is invertible. Since $x + A$ is a nonzero element of R/A, we have $x \notin A$, so by maximality of A, we have $(x) + A = R$. This means that $1 = rx + a$ for some $r \in R$, $a \in A$. But then $(r + A)(x + A) = 1 + A$, so the inverse of $x + A \in R/A$ is the element $r + A \in R/A$. □

For example, for each prime p, the ideal $(p) = p\mathbb{Z}$ is maximal, and $\mathbb{Z}/p\mathbb{Z} = \{0, 1, 2, \ldots, p-1\}$ is the finite field $\mathbb{F}_p$. In the next section, we will see more examples of fields constructed as quotient rings.

Exercises 16.

1. Give an example of a ring $R \supseteq \mathbb{Z}$ such that R has a *unique* maximal ideal $M \neq 0$. Justify your answer.
2. Let $a < b$ be real numbers. For each $c \in [a, b]$, define M_c to be the set of all $f \in C([a, b])$ such that $f(c) = 0$.
   (i) Prove that M_c is a maximal ideal of $C([a, b])$.
   Hint: Consider the map $C([a, b]) \to \mathbb{R}$, $f \mapsto f(c)$.
   (ii) Is *every* maximal ideal of $C([a, b])$ of the form M_c for some $c \in [a, b]$? Justify your answer.
3. Use Zorn's Lemma (see the Appendix) to prove that every proper ideal of a ring R is contained in a maximal ideal.

17. Irreducibility

Let R be a commutative ring with unity. An element $f \in R$ is **irreducible in** R if

(i) f is not a unit of R, and
(ii) whenever $f = gh$ for two elements $g, h \in R$, at least one of g, h is a unit of R.

Note that an irreducible element is always nonzero. Also observe that the property of irreducibility is always relative to a particular choice of ring R; for example X^2+1 is irreducible in the polynomial ring $\mathbb{Z}[X]$ (also in $\mathbb{Q}[X]$ and in $\mathbb{R}[X]$), but is reducible in $\mathbb{C}[X]$ since $X^2+1=(X+i)(X-i)$. Also, $2X+2=2(X+1)$ is reducible in $\mathbb{Z}[X]$, but irreducible in $\mathbb{Q}[X]$ (also in $\mathbb{R}[X]$ and in $\mathbb{C}[X]$). We are interested in irreducibility primarily in the case of polynomial rings.

Note that the irreducible elements of the ring $\mathbb{Z}$ are simply the numbers $\pm p$ where $p\in\mathbb{Z}$ is an ordinary prime. The fact that an ordinary prime $p\in\mathbb{Z}$ is irreducible in $\mathbb{Z}$ is actually *by definition;* to conclude from this that the ideal $(p)\subset\mathbb{Z}$ is prime requires Euclid's Lemma or something more general, such as Theorem 17.2 below.

17.1 Theorem. Let R be an integral domain, and let $0\ne f\in R$. If the ideal (f) is prime, then f is irreducible in R.

Proof. Suppose that the ideal $(f)\subseteq R$ is prime, and that $f=gh$ for some $g,h\in R$. By Proposition 16.1(ii), at least one of g,h lies in (f). We may suppose that $g=qf$ for some $q\in R$. Then $(1-qh)f=0$, and since R is an integral domain, this implies that $qh=1$. This means that h is a unit of R, and so f is irreducible in R. □

17.2 Theorem. Let R be a P.I.D., and let $0\ne f\in R$. Then the following three statements are equivalent.

(i) The element $f\in R$ is irreducible.

(ii) The ideal $(f)\subseteq R$ is maximal.

(iii) The ideal $(f)\subseteq R$ is prime.

Proof. First suppose (i) holds, so that $f\in R$ is irreducible. We will show that (ii) holds. By definition, $f\notin R^\times$, so it generates a proper ideal $(f)\subsetneq R$. Suppose that (f) is properly contained in an ideal $(b)\subseteq R$. (There is no loss of generality in calling this ideal (b) for some $b\in R$ since R is a P.I.D.). Now $f\in(b)$ implies that $f=qb$ for some $q\in R$. Clearly q is not a unit, since $(f)\subsetneq(b)$. Since f is irreducible in R, we must have that b is a unit of R, so $(b)=R$ is the unique ideal properly containing (f). This proves that (f) is maximal.

If (ii) holds, then (iii) follows by Proposition 16.2.

Finally, suppose that (iii) holds, i.e. the ideal $(f)\subset R$ is prime. We must verify (i). Since the ideal $(f)\subset R$ is proper, f is not a unit of R. Suppose that $f=gh$ for some $g,h\in R$. By Proposition 16.1, at least one of g,h lies in (f); without loss of generality, assume $g=qf$ for some $q\in R$. Thus $(1-qh)f=0$ so h is a unit, proving (i). □

As an example, consider $R = \mathbb{Z}[X]$. The ideal $(X) \subset R$ is prime, with quotient $\mathbb{Z}[X]/(X) \cong \mathbb{Z}$. This isomorphism follows from the First Isomorphism Theorem using the homomorphism $\mathbb{Z}[X] \to \mathbb{Z}$, $f(X) \mapsto f(0)$. The ideal $(X, 2) \subset R$ is maximal; it is the kernel of the homomorphism $\mathbb{Z}[X] \to \mathbb{F}_2$, $f(X) \mapsto f(0) \bmod 2$. Again, the First Isomorphism Theorem gives $\mathbb{Z}[X]/(X, 2) \cong \mathbb{F}_2$.

We are interested in irreducible polynomials largely as a means of producing extension fields:

17.3 Theorem. Let F, and let $f(X) \in F[X]$ be irreducible in $F[X]$. Let $n = \deg f(X)$. Then

(i) $E = F[X]/(f(X))$ is a field.

(ii) E contains a subfield $\{a + (f(X)) : a \in F\}$ naturally isomorphic to F (and therefore we identify this subfield with F).

(iii) $[E : F] = n$.

(iv) $f(X)$ has a zero $\alpha = X + (f(X)) \in E$, and $E = F[\alpha] = F(\alpha)$.

Proof. (i) follows from Theorems 16.3 and 17.2.

(ii) Let $\pi : F[X] \to F[X]/(f(X))$ denote the canonical homomorphism $g(X) \mapsto g(X) + (f(X))$. Since $n \geq 1$, the restriction of π to F is one-to-one, thereby giving an isomorphism $F \to \pi(F) = \{a + (f(X)) : a \in F\}$.

(iii) Using the Division Algorithm, every element of E looks like $a_0 + a_1X + a_2X^2 + \cdots + a_{n-1}X^{n-1} + (f(X))$ for some $a_0, a_1, \ldots, a_{n-1} \in F$. This shows that the elements $\pi(X^i) = X^i + (f(X))$ for $i = 0, 1, 2, \ldots, n-1$ span E over F. Moreover this set is linearly independent, for if $a_0\pi(1) + a_1\pi(X) + a_2\pi(X^2) + \cdots + a_{n-1}\pi(X^{n-1}) = (f(X))$, then $a_0 + a_1X + a_2X^2 + \cdots + a_{n-1}X^{n-1} \in (f(X))$, and comparing degrees, this gives $a_0 = a_1 = \cdots = a_{n-1} = 0$. So E is an n-dimensional vector space over $\pi(F) \cong F$.

(iv) We have $f(\alpha) = f(X) + (f(X)) = (f(X))$, which is the zero element of E. Also every element of E is of the form $g(X) + (f(X)) = g(\alpha)$ for some $g(X) \in F[X]$, so that $E \subseteq F[\alpha] \subseteq F(\alpha) \subseteq E$. □

For example, suppose F is a field and $d \in F$ has no square root in F. Then $X^2-d \in F[X]$ is irreducible in F, and $E = F[X]/(X^2-d) \cong F[\sqrt{d}]$ is a quadratic extension of E. We have $E = \{a + bX + (X^2-d) : a, b \in F\} \cong \{a + b\sqrt{d} : a, b \in F\}$. The map $a + bX + (X^2-d) \mapsto a + b\sqrt{d}$ is an isomorphism in this case. Particular cases give $\mathbb{C} = \mathbb{R}[i] \cong \mathbb{R}[X]/(X^2+1)$ and $\mathbb{Q}[\sqrt{5}] \cong \mathbb{Q}[X]/(X^2-5)$.

In view of the utility of Theorem 17.3 for constructing extension fields, it is helpful to have available quick methods for determining irreducibility of polynomials. Unfortunately, in general it is difficult to prove irreducibility. However, some criteria do exist, especially the following.

17.4 Theorem. Let $f(X) \in \mathbb{Z}[X]$. If $f(X)$ is irreducible in $\mathbb{Z}[X]$, then $f(X)$ is irreducible in $\mathbb{Q}[X]$.

Proof. Suppose that $f(X)$ is reducible in $\mathbb{Q}[X]$. Then by factoring out all integer common factors, we have $f(X) = \frac{a}{b} g(X)h(X)$ for some fraction $\frac{a}{b} \in \mathbb{Q}$ in lowest terms, and some $g(X), h(X) \in \mathbb{Z}[X]$ such that the coefficients of $g(X)$ have no common integer divisor greater than 1, and similarly for $h(X)$. We must show that $b = \pm 1$. If not, then we may choose some prime p dividing b. Reducing the polynomial equation $bf(X) = ag(X)h(X)$ modulo p, we obtain $\overline{g}(X)\overline{h}(X) = 0$, where the polynomials $\overline{g}(X), \overline{h}(X) \in \mathbb{F}_p[X]$ are obtained by reducing the coefficients of $g(X), h(X) \in \mathbb{Z}[X]$ modulo p. By construction, both $\overline{g}(X)$ and $\overline{h}(X)$ are nonzero polynomials, and this violates the fact that $\mathbb{F}_p[X]$ has no zero divisors (see Theorem 14.2). □

As an example, the polynomial $X^3 + X + 2$ is irreducible in $\mathbb{Q}[X]$. For otherwise, by Theorem 17.3, it is irreducible in $\mathbb{Z}[X]$, so that $X^3 + X + 2 = (aX + b)(cX^2 + dX + e)$ for some $a, b, c, d, e \in \mathbb{Z}$; but then $a \in \{-1, 1\}$ and $b \in \{-2, -1, 1, 2\}$, so that either $X - 2$ or $X + 2$ divides $X^3 + X + 2$, a contradiction. So by Theorem 17.3, we have a cubic extension of $\mathbb{Q}$ given by $\mathbb{Q}(X)/(X^3 + X + 2)$.

Checking irreducibility of $f(X) \in F[X]$ is usually easier when F is a finite field, since in this case, there are only finitely many polynomials to check as possible factors of $f(X)$. For example, we easily see that the polynomial $f(X) = X^3 + X + 1 \in \mathbb{F}_2[X]$ is irreducible in $\mathbb{F}_2[X]$, since the only polynomials of degree 1 are X and $X + 1$, neither of which divides $f(X)$. So we have a cubic extension $\mathbb{F}_8 = \mathbb{F}_2[X]/(f(X)) \subseteq \mathbb{F}_2$. The elements $a + bX + cX^2 + (f(X)) \in E$ may be abbreviated 'abc', so the eight elements of E are '000' $= (f(X))$, '001' $= X^2 + (f(X))$, '010' $= X + (f(X))$, $\ldots$, '111' $= 1 + X + X^2 + (f(X))$. Addition of elements of E is addition of the coordinates modulo 2; for example, '011' + '110' = '101'. Multiplication is not too hard, e.g.

$$\text{`011'}\cdot\text{`110'} = (X + X^2)(1 + X) + (f(X)) = X + X^3 + (f(X)) = 1 + (f(X)) = \text{`100'}.$$

Another very useful irreducibility criterion is the following.

17.5 Eisenstein Irreducibility Criterion. Let $f(X) = a_0 + a_1X + \cdots + a_nX^n \in \mathbb{Z}[X]$. Suppose there exists a prime p such that

(i) $p \nmid a_0$,
(ii) $p \mid a_1, a_2, \ldots, a_n$, and
(iii) $p^2 \nmid a_n$.

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

Proof. We may suppose that the gcd of the coefficients in $f(X)$ is 1; otherwise factor out the gcd, and this does not affect irreducibility in $\mathbb{Q}[X]$. We will show that $f(X)$ is irreducible in $\mathbb{Z}[X]$, and then by Theorem 17.4 the result will follow. Suppose on the contrary that $f(X)$ is reducible in $\mathbb{Z}[X]$. Then there exist polynomials $g(X), h(X) \in \mathbb{Z}[X]$, neither of which is a unit (i.e. $g(X), h(X) \notin \{1, -1\}$), such that $f(X) = g(X)h(X)$. And neither of the factors $g(X), h(X)$ is an integer constant, since the gcd of the coefficients in $f(X)$ is 1. So $g(X)$ and $h(X)$ are polynomials of degree ≥ 1, say

$$\begin{aligned} g(X) &= b_0 + b_1X + b_2X^2 + \cdots + b_kX^k, \\ h(X) &= c_0 + c_1X + c_2X^2 + \cdots + c_{n-k}X^{n-k} \end{aligned}$$

in $\mathbb{Z}[X]$, where $1 \leq k \leq n-1$.

Let $\overline{}$ denote the reduction modulo p, so that $\overline{\mathbb{Z}} = \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$. By hypothesis (ii) we have $\overline{f}(X) = \overline{a_0} \in \mathbb{F}_p[X]$, a constant polynomial, and by (i), this is not the zero polynomial, so $\deg(\overline{f}(X)) = 0$. Since $\overline{f}(X) = \overline{g}(X)\overline{h}(X)$, by Theorem 14.2 we conclude that both the polynomials $\overline{g}(X), \overline{h}(X) \in \mathbb{F}_p[X]$ are constant (degree 0). In particular, $\overline{b_k} = \overline{c_{n-k}} = 0$, i.e. p divides both b_k and c_{n-k}. Then p^2 divides $a_n = b_kc_{n-k}$, contrary to (iii). □

Observe that

$$a_0 + a_1X + \cdots + a_nX^n = (b_0 + b_1X + b_2X^2 + \cdots + b_kX^k)(c_0 + c_1X + c_2X^2 + \cdots + c_{n-k}X^{n-k})$$

if and only if

$$a_n + a_{n-1}X + \cdots + a_0X^n = (b_k + b_{k-1}X + \cdots + b_0X^k)(c_{n-k} + c_{n-k-1}X + \cdots + c_0X^{n-k}).$$

(This is easily verified by replacing $X \mapsto \frac{1}{X}$, then finding a common denominator.) Therefore the irreducibility of a polynomial is unaffected by *reversing* its list of coefficients. For this reason, the Eisenstein Criterion 17.5 is sometimes stated with the alternative hypotheses

(i′) $p \nmid a_n$,
(ii′) $p \mid a_0, a_1, \ldots, a_{n-1}$, and
(iii′) $p^2 \nmid a_0$.

At first sight, the hypotheses (i)–(iii) [or (i′)–(iii′)] might seem to be too restrictive to be widely applicable. However, we use a trick: although a polynomial $f(X)$ itself might not satisfy the hypotheses (i)–(iii), it may nevertheless be possible to perform a change of variable $X \mapsto X + m$ for some integer m, to obtain a new polynomial which does satisfy the criteria. As an example, consider $f(X) = X^3 - 3X + 5 \in \mathbb{Z}[X]$. There is no prime p satisfying conditions (i)–(iii) or (i′)–(iii′) for $f(X)$. However, substitute $X = Y + 1$ to obtain $f(X) = (Y+1)^3 - 3(Y+1) + 5 = Y^3 + 3Y^2 + 3$. The latter polynomial $g(Y) \in \mathbb{Z}[Y]$ satisfies (i′)–(iii′) for the prime 3, and so $g(Y)$ is irreducible in $\mathbb{Q}[Y]$. It clearly follows that $f(X)$ is irreducible in $\mathbb{Q}[X]$, since any factorization of $f(X)$ in $\mathbb{Z}[X]$ gives, by change of variable, a factorization of $g(Y)$ in $\mathbb{Z}[Y]$. This same trick is useful in proving

17.6 Theorem. Let p be a prime. Then the polynomial

$$\Phi_p(X) = 1 + X + X^2 + \cdots + X^{p-1}$$

is irreducible in $\mathbb{Q}[X]$.

Proof. Note that $(X-1)\Phi_p(X) = X^p - 1$. Substitute $X = Y+1$ to obtain

$$g(Y) = \Phi_p(Y+1) = \frac{(Y+1)^p - 1}{Y} = \frac{1}{Y}\sum_{j=1}^{p}\binom{p}{j}Y^j = Y^{p-1} + \sum_{j=1}^{p-1}\binom{p}{j}Y^{j-1}.$$

Observe that for $1 \le j \le p{-}1$, the coefficient $\binom{p}{j} = \frac{p!}{j!(p-j)!}$ is divisible by p. Also, the constant term in $g(Y)$ is p, which is not divisible by p^2. Thus $g(Y)$ satisfies (i′)–(iii′), so $g(Y)$ is irreducible in $\mathbb{Q}[Y]$. This means that $\Phi_p(X) \in \mathbb{Z}[X]$ is irreducible in $\mathbb{Q}[X]$. □

The polynomial $\Phi_p(X)$ above is an important example of a *cyclotomic polynomial,* which will be useful in Section 23. Exercise 17.9 gives a slight generalization of Theorem 17.6.

Exercises 17.

1. Let $f(X) = X^4 + X^2 + 1$. Factor $f(X)$ into irreducible factors
   (a) in $\mathbb{Q}[X]$; (b) in $\mathbb{R}[X]$; (c) in $\mathbb{C}[X]$; (d) in $\mathbb{F}_2[X]$;
   (e) in $\mathbb{F}_3[X]$; (f) in $\mathbb{F}_5[X]$; (g) in $\mathbb{F}_7[X]$.
2. Let $f(X) \in \mathbb{Q}[X]$. Show that if $f(X^2)$ is irreducible in $\mathbb{Q}[X]$, then so is $f(X)$. Does the converse hold? Justify your answers.
3. Show that $X^3 - 3X - 1$ is irreducible in $\mathbb{Q}[X]$.
4. Construct a field of order 27.
5. Show that $X^5 - X + 1$ is irreducible in $\mathbb{Q}[X]$.
   Hint: If $X^5 - X + 1$ is reducible in $\mathbb{Q}[X]$, then use Theorem 17.4 to obtain a factorization in $\mathbb{Z}[X]$. Now reduce this factorization modulo 5, to obtain a factorization in $\mathbb{F}_5[X]$.
6. Show that for every prime p, there exists an irreducible polynomial $f(X) \in \mathbb{F}_p[X]$ of degree 2. Conclude that for every prime p, there exists a field of order p^2.
7. Let p be prime. As in Exercise 17.6, prove that there exist irreducible polynomials in $\mathbb{F}_p[X]$ of degree 3 and 4. Conclude that there exist fields of order p^3 and p^4.
8. Show that *every* quadratic extension of a given field F may be obtained by Theorem 17.3. In other words, if $[E:F] = 2$, show that $E \cong F[X]/(f(X))$ for some irreducible $f(X) \in F[X]$ of degree 2.
9. Let $q = p^n$ be a power of a prime p, where $n \ge 1$. Define
   $$\Phi_q(X) = \frac{X^{p^n} - 1}{X^{p^{n-1}} - 1} = 1 + X^{p^{n-1}} + X^{2p^{n-1}} + \cdots + X^{(p-1)p^{n-1}} \in \mathbb{Z}[X].$$
   Prove that $\Phi_q(X)$ is irreducible in $\mathbb{Q}[X]$.
   Hint: Imitate the proof of Theorem 17.6.

Chapter III

Fields

18. Algebraic Extensions

We begin with some standard definitions. Let F be a field. Since F contains 1, for every positive integer it also contains an element $1+1+\cdots+1$ (n times), abbreviated simply as n. However, it may happen that $n = 1+1+\cdots+1 = 0$ in F, for some positive integer n. If so, then the smallest positive integer n for which this happens must be prime; otherwise by the distributive law,

$$n = n_1 n_2 = \overbrace{(1+1+\cdots+1)}^{n_1 \text{ times}}\overbrace{(1+1+\cdots+1)}^{n_2 \text{ times}} = \overbrace{(1+1+\cdots+1)}^{n \text{ times}} = 0$$

where $n_1, n_2 < n$, so that $n_1 = 0$ or $n_2 = 0$, contradicting the minimality of n. This minimum n is called the **characteristic** of F. If there is *no* positive integer n such that $n = 0$ in F, we say that F has **characteristic zero**. We denote the characteristic of an arbitrary field F by $\operatorname{char} F$. So either $\operatorname{char} F = p$, a prime, in which case $pa = a + a + \cdots + a = (1+1+\cdots+1)a = 0$ for all $a \in F$; or $\operatorname{char} F = 0$, in which case $na = a + a + \cdots + a$ is *never* zero for any $a \in F^\times$ and any positive integer n.

The unique smallest subfield of F is called the **prime field** of F. This is the subfield of F generated by 1, and it is isomorphic to $\mathbb{F}_p \cong \mathbb{Z}/p\mathbb{Z}$ if $\operatorname{char} F = p$ is prime; the prime field is isomorphic to $\mathbb{Q}$ if $\operatorname{char} F = 0$. In particular every finite field has prime characteristic, since it cannot have an infinite subfield $\mathbb{Q}$. We will devote more time, however, to studying fields of characteristic zero.

First, a result showing the multiplicativity of degrees of extensions:

18.1 Theorem. For every tower $K \supseteq E \supseteq F$ of fields, $[K : F] = [K : E][E : F]$.

Proof. Let $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$ be a basis for K over E, and let $\{\beta_1, \beta_2, \ldots, \beta_n\}$ be a basis for E over F. Then it is easy to show (Exercise 18.1) that $\{\alpha_i\beta_j : 1 \le i \le m,\ 1 \le j \le n\}$ is a basis for K over F, so that $[K : F] = mn = [K : E][E : F]$. □

Let $E \supseteq F$ be any extension of fields. For each $\alpha \in E$, we say that α is **algebraic** over F if $f(\alpha) = 0$ for some nonzero polynomial $f(X) \in F[X]$. If no such nonzero polynomial exists, we say that α is **transcendental** over F. Thus, for example, $\sqrt{2} \in \mathbb{R}$ is algebraic over $\mathbb{Q}$. Also $\pi \in \mathbb{R}$ is transcendental over $\mathbb{Q}$, but algebraic over $\mathbb{R}$, as it is a zero of the polynomial $X - \pi \in \mathbb{R}[X]$. If $\alpha \in E$ is algebraic over F, then $\{f(X) \in F[X] : f(\alpha) = 0\}$ is a nonzero ideal in $F[X]$, and since $F[X]$ is a P.I.D., this ideal is principal, say

$$\{f(X) \in F[X] : f(\alpha) = 0\} = \big(f_0(X)\big)$$

where $f_0(X) \in F[X]$. Moreover we may assume that $f_0(X)$ is **monic**, i.e. that its leading coefficient (the coefficient of the highest power of X) is 1, for otherwise we may adjust $f_0(X)$ by multiplying by the appropriate constant. Clearly $f_0(X)$ is irreducible in $F[X]$; otherwise one of the factors of $f_0(X)$ would be a polynomial of degree less than $\deg f_0(X)$ having α as a zero. Since $f_0(X)$ is the unique lowest degree monic polynomial in $F[X]$ having α as a zero, we rename it as $\mathrm{Irr}_{\alpha,F}(X)$. The **degree** of α over F is by definition the degree of $\mathrm{Irr}_{\alpha,F}(X)$. Thus, for example, $\mathrm{Irr}_{\sqrt{5},\mathbb{Q}}(X) = X^2 - 5$, whereas $\mathrm{Irr}_{\sqrt{5},E}(X) = X - \sqrt{5}$ if $E = \mathbb{Q}[\sqrt{5}]$; so $\sqrt{5}$ is algebraic of degree 2 over $\mathbb{Q}$, but $\sqrt{5}$ is algebraic of degree 1 over $E = \mathbb{Q}[\sqrt{5}]$.

18.2 Theorem. Let $E \supseteq F$ be an extension of fields, and let $\alpha \in E$. Then the following conditions are equivalent.

(i) α is algebraic over F.

(ii) $[F(\alpha) : F] < \infty$.

(iii) $F(\alpha) = F[\alpha]$.

Proof. (i)⇒(iii) Suppose α is algebraic over F, so that $f(\alpha) = 0$ where $f(X) = \mathrm{Irr}_{\alpha,F}(X)$. We must show that $F[\alpha]$ is a field. Consider a nonzero element of $F[\alpha]$; this has the form $g(\alpha)$ where $g(X) \in F[X]$ is not divisible by $f(X)$. Then by Euclid's Algorithm (see Exercise 15.3) there exist polynomials $u(X), v(X) \in F[X]$ such that

$$u(X)f(X) + v(X)g(X) = 1.$$

Substituting for α in place of X gives $v(\alpha)g(\alpha) = 1$, i.e. $g(\alpha)$ is a unit of $F[\alpha]$.

(iii)⇒(i) We may assume that $\alpha \neq 0$. If $F(\alpha) = F[\alpha]$ then

$$\alpha^{-1} = a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n$$

for some $n \geq 0$, $a_i \in F$. This means that

$$1 - a_0\alpha - a_1\alpha^2 - \cdots - a_n\alpha^{n+1} = 0,$$

where the coefficients of the powers α^i are not all zero (in fact the coefficient of α^0 is 1), so α is algebraic over F.

(ii)⇒(i) If $[F(\alpha) : F] = n < \infty$, then the set $\{1, \alpha, \alpha^2, \ldots, \alpha^n\}$ is linearly dependent over F, and so there exist coefficients $a_i \in F$, not all zero, such that

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} + a_n\alpha^n = 0,$$

i.e. α is algebraic over F.

{(i),(iii)}⇒(ii) If α is algebraic over F, then $f(\alpha)=0$ where $\mathrm{Irr}_{\alpha,F}(X)$ is of degree n, say. Each element of $F(\alpha)=F[\alpha]$ is of the form $g(\alpha)$ for some $g(X)\in F[X]$. Since $g(X)=u(X)f(X)+r(X)$ for some $u(X),r(X)\in F[X]$ with $\deg r(X)<n$, we have

$$F(\alpha)=F[\alpha]=\{r(\alpha):r(X)\in F[X],\ \deg r(X)<n\}.$$

It follows that $\{1,\alpha,\alpha^2,\ldots,\alpha^{n-1}\}$ is a basis for $F(\alpha)=F[\alpha]$ over F, so $[F(\alpha):F]=n<\infty$. □

18.3 Corollary. Let $E\supseteq F$ be an extension of fields, and let $\alpha\in E$ be algebraic over F. Then the degree of α over F equals $[F[\alpha]:F]$.

Proof. The key ideas in proving this result appear in the proof of Theorem 18.2; the details are left as an exercise. □

18.4 Corollary. Suppose that $E\supseteq F$ is a field extension such that $E=F[\alpha]$ for some $\alpha\in E$. Then $E\cong F[X]\big/\big(\mathrm{Irr}_{\alpha,F}(X)\big)$.

Proof. Clearly the map $\pi:F[X]\to E,\ g(X)\mapsto g(\alpha)$ is a homomorphism. By definition, $\ker\pi=\big(\mathrm{Irr}_{\alpha,F}(X)\big)$. Also the image is $\pi(E)=F[\alpha]=E$, so the result follows from the First Isomorphism Theorem 15.3. □

A field extension $E\supseteq F$ is called **algebraic** if every element of E is algebraic over F. If $[E:F]<\infty$ then the extension is algebraic, since for each $\alpha\in E$ we have $E\supseteq F(\alpha)\supseteq F$ and $[E:F]=[E:F(\alpha)][F(\alpha):F]<\infty$, so that $[F(\alpha):F]<\infty$, i.e. α is algebraic over F. We restate this fact formally:

18.5 Corollary. Every finite field extension $E\supseteq F$ is algebraic. In particular if $\alpha\in E$ is algebraic over F, then the extension $F(\alpha)\supseteq F$ is algebraic.

Let $E\supseteq F$ be an extension of fields. If $\alpha,\beta\in E$ are both algebraic over F, then so are $\alpha+\beta$ and $\alpha\beta$. To see this, first note that $[E:F]<\infty$ where $E=F(\alpha)=F[\alpha]$. Now β is a zero of some nonzero polynomial $f(X)\in F[X]\subseteq E[X]$, and so β is algebraic over E, i.e.

$$[E(\beta):F]=[E(\beta):E][E:F]<\infty.$$

This means that $E(\beta)=F(\alpha,\beta)$ is an algebraic extension of F, and in particular the elements $\alpha+\beta$, $\alpha\beta\in F(\alpha,\beta)$ are algebraic over F. So define the **algebraic closure of** F **in** E as

$$\{\alpha\in E:\alpha\text{ is algebraic over }F\}.$$

This is a field by the preceding arguments, and so is an algebraic extension of F. (Of course this is nothing other than E, if E is already algebraic over F. More generally, the algebraic closure of F in E is the largest subfield of E containing F, which is algebraic over F.)

A field is **algebraically closed** if, for every extension $E \supseteq F$, the algebraic closure of F in E is F itself. For example, $\mathbb{C}$ is algebraically closed. Also, there exist algebraically closed fields of every possible characteristic.

Warning: The terminology now starts to sound slightly ambiguous unless one is careful to observe the prepositions! An **algebraic closure** of a field F is an extension $\overline{F} \supseteq F$ such that (i) $\overline{F}$ is algebraically closed, and (ii) the extension $\overline{F} \supseteq F$ is algebraic. It may be shown that any two algebraic closures of the same field F are isomorphic. If one accepts Zorn's Lemma (or, equivalently, the Axiom of Choice) then every field has an (unique) algebraic closure. For example, the algebraic closure of $\mathbb{R}$ is $\mathbb{C}$. It will be convenient for us to assume the existence of algebraic closures, although this is not strictly necessary since all our Galois theory can be accomplished using finite extensions.

Exercises 18.

1. Fill in the missing details in the proof of Theorem 18.1.

2. Let $E \supseteq F$ be an extension of degree 3, and suppose $\alpha \in E$ but $\alpha \notin F$. Show that $F[\alpha] = E$.

3. Recall that if A is a linear transformation $V \to V$ where V is an n-dimensional vector space over a field F, and if $g(X) \in F[X]$ is its characteristic polynomial, then $g(A) = 0$ (the zero matrix) by the Cayley-Hamilton Theorem; and that $g(X)$ is divisible by $m(X) \in F[X]$, the **minimal polynomial** of A over F, i.e. the (unique) minimum degree monic polynomial having A as a zero.

   (a) Let F be a field, and let
   $$f(X) = a_0 + a_1X + a_2X^2 + \cdots + a_{n-1}X^{n-1} + X^n \in F[X].$$
   The **companion matrix** of $f(X)$ is the $n \times n$ matrix
   $$A_f = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix} \in F^{n\times n}.$$
   Find the characteristic polynomial of A_f .

   (b) Let $E \supseteq F$ be an extension of degree n, and let $\alpha \in E$. The map
   $$E \to E, \quad x \mapsto \alpha x$$
   is a linear transformation A_α of E over F. Show that the minimal polynomial of A_α equals $\mathrm{Irr}_{\alpha,F}(X)$. If $E = F(\alpha)$, show that the characteristic polynomial of A_α equals $\mathrm{Irr}_{\alpha,F}(X)$.

4. Prove Corollary 18.3.

5. Let F be a finite field. Show that $|F| = p^r$ for some prime p and some integer $r \geq 1$.

   Hint: Let $p = \operatorname{char} F$. Then F is a vector space of finite dimension r, say, over its prime field $\mathbb{F}_p$.

6. Since $\sqrt{2}$ and $\sqrt{3}$ are algebraic over $\mathbb{Q}$, so is $\alpha = \sqrt{2} + \sqrt{3}$ by the comments following Corollary 18.5. Determine $\mathrm{Irr}_{\alpha,\mathbb{Q}}(X)$.

7. Use Zorn's Lemma to show that every field has an algebraic closure.

8. Assume the remarks at the end of Section 18 regarding existence of algebraic closures. Show that there exist algebraically closed fields F which are proper extensions of $\mathbb{C}$.

9. By Exercise 17.9, the polynomial $\Phi_9(X) = 1 + X^3 + X^6$ is irreducible in $\mathbb{Q}[X]$. Let $\zeta = e^{2\pi i/9}$.
   (a) Show that ζ is a zero of $\Phi_9(X)$. Conclude that $E = \mathbb{Q}(\zeta) \supset \mathbb{Q}$ is an extension of degree 6.
   (b) Let $\alpha = \zeta + \zeta^{-1} = 2\cos(\frac{2\pi}{9}) \in \mathbb{R}$ and let $F = \mathbb{Q}(\alpha)$. Show that α is a *proper* subfield of E.
   (c) Show that ζ is a zero of $X^2 - \alpha X + 1 \in F[X]$. Conclude that $[E : F] = 2$ and $[F : \mathbb{Q}] = 3$.
   (d) Determine $\mathrm{Irr}_{\alpha,\mathbb{Q}}(X)$.

10. **(Straightedge-and-Compass Constructions)** Consider the points $P_0 = (0, 0)$ and $P_1 = (1, 0)$ in the Euclidean plane $\mathbb{R}^2$. Suppose that, for each $n \geq 2$, the point $P_n = (x_n, y_n) \in \mathbb{R}^2$ is constructible from $\{P_0, P_1, \ldots, P_{n-1}\}$ by an *elementary straightedge-and-compass construction,* i.e. one of the following holds:
    (i) P_n is the intersection of two distinct straight lines, each joining two points of $\{P_0, P_1, \ldots, P_{n-1}\}$;
    (ii) P_n is a point of intersection of a straight line joining two points of $\{P_0, P_1, \ldots, P_{n-1}\}$, and a circle with center P_i and radius P_jP_k , $0 \leq i, j, k < n$, $j \neq k$; or
    (iii) P_n is a point of intersection of two distinct circles, each having as center a point in $\{P_0, P_1, \ldots, P_{n-1}\}$, and radius equal in length to a segment joining two of $\{P_0, P_1, \ldots, P_{n-1}\}$.

    A point $P \in \mathbb{R}^2$ is *constructible* if it belongs to some sequence $P_0, P_1, P_2, \ldots$ of points, each point obtainable by an elementary construction from the previous points. Let $F_n = \mathbb{Q}(x_0, y_0, x_1, y_1, \ldots, x_n, y_n)$, which is the smallest subfield of $\mathbb{R}$ containing the coordinates of $P_0, P_1, \ldots, P_n$.
    (a) Show that $[F_n : F_{n-1}] = 1$ or 2.
    (b) Show that $[F_n : \mathbb{Q}] = 2^t$ for some $t \leq n - 1$.
    (c) Show that the point $P = (\frac{1}{2}, \frac{\sqrt{3}}{2})$ is constructible, and hence the angle $\frac{\pi}{3}$ (formed by P_1, P_0, P) is "constructible" by straightedge and compass.
    (d) Show that the point $Q = (\cos\frac{\pi}{9}, \sin\frac{\pi}{9})$ is not constructible, and hence the angle $\frac{\pi}{3}$ may not be trisected using straightedge and compass.

    Hint: Use Exercise 18.9.
    (e) Is the point $(\cos\frac{\pi}{5}, \sin\frac{\pi}{5})$ constructible? Explain.

19. Normal Extensions

Let $f(X) \in F[X]$. We wish to find an extension $E \supseteq F$ in which $f(X)$ splits completely into linear factors. Clearly it is sufficient to be able to do this when $f(X)$ is irreducible in $F[X]$. In this case, $E = F[X]/(f(X))$ has at least one zero $\alpha = X + (f(X))$. However, $f(X)$ might not split into linear factors in $E[X]$.

For example, $X^3 - 2$ is irreducible in $\mathbb{Q}[X]$, and $F = \mathbb{Q}[X]/(X^3 - 2)$ is a cubic extension of $\mathbb{Q}$ in which $f(X)$ has a zero $\alpha = X + (f(X)) \in F$, and $f(X) = (X - \alpha)(X^2 + \alpha X + \alpha^2)$. But the latter quadratic factor is irreducible in $F[X]$. (This is not hard to see. For if $\beta \in F$ is a zero of $X^2 + \alpha X + \alpha^2$, then $[\mathbb{Q}[\beta] : \mathbb{Q}]$ divides $[F : \mathbb{Q}] = 3$. But clearly $\beta \notin \mathbb{Q}$, so the degree of β over $\mathbb{Q}$ is 2, whence $[\mathbb{Q}[\beta] : \mathbb{Q}] = 2$ by Corollary 18.3, a contradiction.) So we take $E = F[T]/(T^2 + \alpha T + \alpha^2)$, which is a quadratic extension of F in which $f(X) = (X - \alpha)(X - \omega\alpha)(X - \omega^2\alpha)$, where $\omega \in E$ is a zero of $T^2 + T + 1$. Altogether we have $[E : \mathbb{Q}] = [E : F][F : \mathbb{Q}] = 2 \cdot 3 = 6$. This extension E is a *splitting field* for the polynomial $X^3 - 2$ over $\mathbb{Q}$. In general, an extension $E \supseteq F$ is called a **splitting field** for a polynomial $f(X) \in F[X]$, if

(i) $f(X)$ splits into linear factors in $E[X]$, i.e. $f(X) = (X - \alpha_1)(X - \alpha_2)\cdots(X - \alpha_n)$, where $\alpha_i \in E$; and

(ii) $E = F(\alpha_1, \alpha_2, \ldots, \alpha_n)$.

The reason for the last restriction is that $f(X)$ splits into linear factors in $F(\alpha_1, \alpha_2, \ldots, \alpha_n) \subseteq E$, and we want E to be the *smallest* possible extension of F in which $f(X)$ splits into linear factors. For example, while $X^3 - 2$ splits into linear factors in $\mathbb{C}$, we would not call $\mathbb{C}$ a splitting field for $X^3 - 2$. However, the field $E = \mathbb{Q}(\alpha, \omega\alpha, \omega^2\alpha) = \mathbb{Q}(\alpha, \omega)$ is a splitting field for $X^3 - 2$.

19.1 Theorem. Any two splitting fields for the same polynomial $f(X) \in F[X]$ are isomorphic.

We shall not prove this fact, but it says for example that *every* splitting field for $X^3 - 2$ over $\mathbb{Q}$, is isomorphic to the extension $E = \mathbb{Q}(\alpha, \omega)$ constructed above. Therefore we are justified in calling E *the* splitting field for $X^3 - 2$ over $\mathbb{Q}$, rather than merely *a* splitting field for $X^3 - 2$ over $\mathbb{Q}$.

In the example above, the extension $F \supset \mathbb{Q}$ has the somewhat unfortunate property that $X^3 - 2$ has a zero in F, yet it does not split completely into linear factors in $F[X]$. The extension $E \supset \mathbb{Q}$ is nicer in this respect: it is a fact that *every* irreducible polynomial in $\mathbb{Q}[X]$ having a zero in E, splits into linear factors in $E[X]$. (This fact follows from Theorem 19.2 below, which we also state without proof.)

We say that an extension $L \supseteq K$ is **normal** if every polynomial $f(X) \in K[X]$ which is irreducible in $K[X]$ and has a zero in L, splits completely into linear factors in $L[X]$. The choice of terminology 'normal' for this property is directly related to the property of a subgroup being normal, as we shall see later. The finite normal extensions are characterized as the splitting fields of polynomials, thus:

19.2 Theorem. A finite extension $L \supseteq K$ of fields is normal iff L is the splitting field of some polynomial $f(X) \in K[X]$ over K.

Note that the polynomial $f(X)$ in Theorem 19.2 is not required to be irreducible; for example

$$\mathbb{Q}[\sqrt{2}, \sqrt{5}] = \{a + b\sqrt{2} + c\sqrt{5} + \sqrt{10} \,:\, a, b, c, d \in \mathbb{Q}\}$$

is a normal extension of $\mathbb{Q}$ of degree 4 (with basis $\{1, \sqrt{2}, \sqrt{5}, \sqrt{10}\}$), since it is the splitting field of $(X^2 - 2)(X^2 - 5)$ over $\mathbb{Q}$.

Exercises 19.

1. Find an extension $E \supset F$ of degree 4 which is not normal. Explain.

2. Give an example of normal extensions $E \supset F$ and $L \supset E$ such that the extension $L \supset F$ is not normal. Justify your answer. (This is just like the situation for groups; see Exercise 5.4).

3. Prove that every quadratic extension of fields is normal. This is the analogue of which result in group theory? (see Section 5).

4. If $L \supseteq E \supseteq F$ are fields such that the extension $L \supseteq F$ is normal, does it follow that the extension $L \supseteq E$ is normal? or that the extension $E \supseteq F$ is normal? Explain.

5. Prove Theorem 19.2.

20. Separable Extensions

Another 'unfortunate' situation that may occur in an algebraic extension $E \supseteq F$, is that if $\alpha \in E$ and $f(X) = \mathrm{Irr}_{\alpha,F}(X)$ is the minimal monic polynomial of α over F, then $f(X)$ may have repeated roots. This never happens in extensions of $\mathbb{Q}$ as we shall see, but consider the field

$$E = \mathbb{F}_2(T) = \left\{ \frac{f(T)}{g(T)} \;:\; f(T), g(T) \in \mathbb{F}_2[T],\ g(T) \neq 0 \right\},$$

i.e. E is the field consisting of all rational functions in a single indeterminate T, with coefficients in the field $\mathbb{F}_2 = \{0,1\}$; see Exercise 14.13. Now E has a subfield $F = \mathbb{F}_2(T^2)$ consisting of all rational functions of T^2 with coefficients in $\mathbb{F}_2$, i.e. F is the quotient field of $\mathbb{F}_2[T^2]$. (Actually $F \cong E$, but that is irrelevant here; F is a *proper* subfield of E.) The element $T \in E$ is algebraic of degree 2 over F, since $T \notin F$ is a zero of the polynomial $f(X) = X^2 - T^2 \in F[X]$. Since $E = F(T)$, it follows from Corollary 18.5 that the extension $E \supset F$ is algebraic of degree 2. (In fact every quadratic extension is algebraic; see Exercise 19.2.) Note that $f(X) = X^2 - T^2 = (X-T)^2$ is irreducible in $F[X]$, and has a double zero $T \in E$.

We wish to avoid such situations, so whenever $E \supseteq F$ is an extension and $\alpha \in E$ is algebraic over F, we say that α is **separable** over F if α is a *simple* zero of $\mathrm{Irr}_{\alpha,F}(X)$.

20.1 Theorem. Let $E \supseteq F$ be an extension of fields of characteristic zero, i.e. $E \supseteq F \supseteq \mathbb{Q}$. Then every element $\alpha \in E$ which is algebraic over F, is separable over F.

Proof. Let $f(X) = \mathrm{Irr}_{\alpha,F}(X)$ have degree n. Since $f(X)$ is monic, its leading term is X^n. If $(X-\alpha)^2$ divides $f(X)$ in $E[X]$, then $(X-\alpha)$ divides $f'(X)$ in $E[X]$. But the leading term of $f'(X)$ is $nX^{n-1} \neq 0$, and so α is a zero of the nonzero polynomial $f'(X) \in F[X]$ of degree $n-1 < \deg f(X)$, contradicting $f(X) = \mathrm{Irr}_{\alpha,F}(X)$. □

Note why it is that the proof above fails in the situation $f(X) = X^2 - T^2 = (X-T)^2 \in \mathbb{F}_2[X]$: in this case T is a zero of $f'(X)$ alright, but $f'(X) = 0$ (the zero polynomial).

An extension $E \supseteq F$ is called **separable** if every element of E is separable over F. (This definition implicitly requires that the extension $E \supseteq F$ is algebraic.) We have just shown that every algebraic extension of a field of characteristic zero, is separable. It may also be shown that if $E \supseteq F$ are finite fields, then the extension $E \supseteq F$ is separable.

Exercises 20.

1. Suppose that $L \supseteq E \supseteq F$ is a tower of fields such that the extension $L \supseteq F$ is separable. Prove that the extensions $L \supseteq E$ and $E \supseteq F$ are separable.

21. Field Isomorphisms

Let F be any field. An **automorphism** of F is a ring isomorphism $\sigma : F \to F$, i.e. a bijection $F \to F$ such that $(x+y)^\sigma = x^\sigma + y^\sigma$ and $(xy)^\sigma = x^\sigma y^\sigma$ for all $x, y \in F$. The set of all automorphisms of F is a group under composition, called the **automorphism group** of F, denoted $\operatorname{Aut} F$. For every $\sigma \in \operatorname{Aut} F$, let $F_\sigma = \{x \in F : x^\sigma = x\}$. It is clear that F_σ contains both 0 and 1, and is closed under addition, subtraction, multiplication and division by nonzero elements of F_σ; hence F_σ is a subfield of F which contains the prime field. We call F_σ the **fixed field** of σ. For example, the usual complex conjugation $\tau : z \mapsto \overline{z}$ is an automorphism of $\mathbb{C}$, with fixed field $\mathbb{C}_\tau = \mathbb{R}$.

Now suppose that $E \supseteq F$ is an extension of fields. An automorphism of E which fixes every element of F, is called an F-automorphism of E. The set of all F-automorphisms of E is a subgroup of $\operatorname{Aut} E$, denoted $G(E/F)$. For each $\sigma \in G(E/F)$, the fixed field E_σ is an intermediate field, i.e. $E \supseteq E_\sigma \supseteq F$. More generally, for each subgroup $H \leq G$, the fixed field $E_H = \{x \in E : x^\sigma = x \text{ for all } x \in H\}$ is an intermediate field, $E \supseteq E_H \supseteq F$. Conversely, given any **intermediate field** L, i.e. $E \supseteq L \supseteq F$, we have a subgroup $G(E/L) \leq G(E/F)$. Our goal in Sections 21 and 22, is to develop the rudiments of *Galois theory,* which is concerned with the special class of extensions $E \supseteq F$ (namely, finite normal separable extensions) for which this correspondence $H \leftrightarrow E_H$ between subgroups of $G(E/F)$ and intermediate fields, is bijective.

Let $E \supseteq F$ and $E' \supseteq F$ be extensions of the same field F. An **F-isomorphism** from E to E' is a field isomorphism $E \to E'$ (i.e. ring isomorphism of fields) which fixes every element of F. An **F-monomorphism** from E to E' is a field monomorphism (i.e. one-to-one ring homomorphism of fields) which fixes every element of F.

21.1 Theorem. Let $E \supseteq F$ be a separable extension of degree n, and let C be an algebraically closed field containing F. Then there exist exactly n distinct F-monomorphisms from E into C.

$$\begin{tikzcd} E \arrow[rr, dotted] & & C \\ & F \arrow[ul, dash] \arrow[ur, dash] & \end{tikzcd}$$

Proof. First consider the special case that $E = F[\alpha]$ for some $\alpha \in E$. Let $f(X) = \operatorname{Irr}_{\alpha,X}(X)$. Since C is algebraically closed, $f(X)$ splits into linear factors in $C[X]$, say $f(X) = (X-\alpha_1)(X-\alpha_2)\cdots(X-\alpha_n)$ for some $\alpha_i \in C$. For each i, observe that $\operatorname{Irr}_{\alpha_i,F}(X) = f(X)$ since $f(X)$ is monic irreducible in $F[X]$ and has α_i as a zero.

For each $i = 1, 2, \ldots, n$, define $\sigma_i : F[\alpha] \to C$ by $g(\alpha) \mapsto g(\alpha_i)$ where $g(X) \in F[X]$. Then σ_i is well-defined, since if $g(\alpha) = h(\alpha)$, then $g(X) \equiv h(X) \mod (f(X))$, in which case $g(\alpha_i) = h(\alpha_i)$. Clearly $\sigma_i : E \to C$ is a ring homomorphism, fixing every element of F. Also σ_i is one-to-one, for if $g(\alpha)^{\sigma_i} = g(\alpha_i) = 0$, then $f(X)$ divides $g(X)$, so that $g(\alpha) = 0$. So each $\sigma_i : E \to C$ is an F-monomorphism. The image of σ_i is the subfield $E^{\sigma_i} = F[\alpha_i] \subseteq C$.

Now $F[\alpha_i] \cong F[\alpha] = E$ is separable over F, so $\alpha_1, \alpha_2, \ldots, \alpha_n$ are distinct. Since $\alpha^{\sigma_i} = \alpha_i$, the monomorphisms $\sigma_1, \sigma_2, \ldots, \sigma_n$ are distinct.

Finally, let σ be any F-monomorphism from E into C. Then $f(\alpha^\sigma) = f(\alpha)^\sigma = 0^\sigma = 0$, so that $\alpha^\sigma \in \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$. Let us say that $\alpha^\sigma = \alpha_i$. Since the ring homomorphisms σ and σ_i agree on F and on α, they must agree on $F[\alpha] = E$, i.e. $\sigma = \sigma_i$. Thus $\sigma_1, \sigma_2, \ldots, \sigma_n$ are the *only* F-monomorphisms from E into C.

Consider now the general case $E \supsetneqq F$, and let $\alpha \in E \smallsetminus F$. We may assume that $F[\alpha] \subsetneqq E$; otherwise we are done by the previous case. We have $E \supsetneqq F[\alpha] \supsetneqq F$ and $n = mt$ where $m = [E : F[\alpha]]$ and $t = [F[\alpha] : F]$. By induction on the degree of extension, there exist t distinct monomorphisms $\sigma_1, \sigma_2, \ldots, \sigma_t : F[\alpha] \to C$. Let $\alpha_i = \alpha^{\sigma_i}$.

$$\begin{tikzcd}
& & & C \arrow[dl, no head] \arrow[dr, no head] & \\
E \arrow[rr, dotted, "\sigma_i"] \arrow[dr, no head] & & E^{\sigma_i} \arrow[rr, dotted, "\theta_{ij}"] \arrow[dl, no head] \arrow[dr, no head] & & E^{\sigma_i\theta_{ij}} \arrow[dl, no head] \\
& F[\alpha] \arrow[rr, dotted, "\sigma_i"] \arrow[dr, no head] & & F[\alpha_i] \arrow[dl, no head] & \\
& & F & &
\end{tikzcd}$$

Since $\sigma_i : E \to E^{\sigma_i}$ is an F-isomorphism, the extension $E^{\sigma_i} \supseteq F$ is separable; hence by Exercise 20.1, the extension $E^{\sigma_i} \supseteq F[\alpha_i]$ is separable. By induction on the degree of extension, for each i there exist m distinct $F[\alpha_i]$-monomorphisms $\theta_{i1}, \theta_{i2}, \ldots, \theta_{im} : E^{\sigma_i} \to C$. The composite maps $\sigma_i\theta_{ij} : E \to C$ (with left-to-right composition) constitute $mt = n$ distinct F-monomorphisms. To see that these are the *only* F-monomorphisms $E \to C$, suppose that $\sigma : E \to C$ is an F-monomorphism. As before, σ must take α to some α_i. Then $\sigma_i^{-1}\sigma : E^{\sigma_i} \to C$ is an $F[\alpha_i]$-monomorphism, so by induction, $\sigma_i^{-1}\sigma = \theta_{ij}$ for some j, whence $\sigma = \sigma_i\theta_{ij}$ as required. □

It is often useful to have a single generator for an extension field. The following result guarantees that such a generator exists for all finite separable extensions. We present a proof, however, only in the special case $\operatorname{char} F = 0$. This case is slightly easier, and it is the primary case we are interested in. For a proof in the general case, see e.g. Garling [1].

21.2 Theorem of the Primitive Element. *Let $E \supseteq F$ be a finite separable extension of fields. Then $E = F[\alpha]$ for some $\alpha \in E$.*

Proof in the case $\operatorname{char} F = 0$. Let C be an algebraically closed field containing F. By Theorem 21.1, there exist distinct F-monomorphisms $\sigma_1, \sigma_2, \ldots, \sigma_n : E \to C$ where $n = [E : F]$.

We claim that there exists $\alpha \in E$ such that the images $\alpha^{\sigma_1}, \alpha^{\sigma_2}, \ldots, \alpha^{\sigma_n} \in C$ are distinct. To see this, we apply Exercise 21.1 as follows. Whenever $1 \le i < j \le n$, the set $V_{ij} = \{x \in E : x^{\sigma_i} = x^{\sigma_j}\}$ is a proper subspace of the vector space E over F. Also $|F| = \infty$ since $\operatorname{char} F = 0$. Since E cannot be covered by finitely many proper subspaces V_{ij}, there exists $\alpha \in E \smallsetminus \left(\bigcup_{1 \le i < j \le n} V_{ij}\right)$, and this α has the required property: $\alpha^{\sigma_i} \ne \alpha^{\sigma_j}$ whenever $i \ne j$.

Since $[E : F] < \infty$, we have $F(\alpha) = F[\alpha]$ by Corollary 18.5 and Theorem 18.2. So we have a tower of extensions $E \supseteq F[\alpha] \supseteq F$ and $n = [E : F[\alpha]][F[\alpha] : F]$. Since the restrictions $\sigma_1, \ldots, \sigma_n : F[\alpha] \to C$ are distinct F-monomorphisms, we have $n \le [F[\alpha] : F]$ by Theorem 21.1. Therefore $[F[\alpha] : F] = n$ and $E = F[\alpha]$. □

Remark: The use of an algebraically closed extension C in the proof of Theorem 21.2 was simply a convenient crutch, and was not really necessary. All that is really required is a splitting field for $f(X)$, which is a finite extension. This releases us from having to assume the Axiom of Choice (see comments at the end of Section 18).

Exercises 21.

1. Let V be a vector space over an infinite field F, and let $V_1, V_2, \ldots, V_n$ be finitely many proper subspaces of V. Prove that the union $V_1 \cup V_2 \cup \cdots \cup V_n$ is a *proper* subset of V. (This problem was used in the proof of Theorem 21.2. The proof is slightly tricky, but requires only elementary linear algebra.)

2. Show that $\mathbb{Q}[\sqrt{2}, \sqrt{3}] \supset \mathbb{Q}$ is a finite normal extension. By Theorem 21.2 there exists α such that $\mathbb{Q}[\sqrt{2}, \sqrt{3}] = \mathbb{Q}[\alpha]$. Find an explicit choice for such an α. Justify your answer.

3. Let $E \supseteq F$ be a separable extension of degree n, and let C be an algebraically closed field containing F. By Theorem 21.1, there exist exactly n distinct F-monomorphisms $\sigma_i : E \to C$. Prove that the functions $\sigma_1, \sigma_2, \ldots, \sigma_n$ are linearly independent over F.

   Hint: If not, consider a subset of $\{\sigma_1, \ldots, \sigma_n\}$ which is linearly dependent, but having no linearly dependent proper subset. We may assume $\{\sigma_1, \ldots, \sigma_k\}$ is such a subset, where $1 \le k \le n$. Now there exist constants $a_1, \ldots, a_k \in C$, none of which are zero, such that $a_1 x^{\sigma_1} + \cdots + a_k x^{\sigma_k} = 0$ for all $x \in E$. Then $k \ge 2$, and since $\sigma_1 \ne \sigma_2$, there exists $c \in E$ such that $c^{\sigma_1} \ne c^{\sigma_2}$. Replacing x by cx in the relation above, we obtain another relation $a_1 c^{\sigma_1} x^{\sigma_1} + \cdots + a_k c^{\sigma_k} x^{\sigma_k} = 0$ for all $x \in E$. Eliminating x^{σ_1} from these two relations, obtain a relation involving only $x^{\sigma_2}, \ldots, x^{\sigma_k}$.

4. Let F be a field. Prove that every automorphism of F fixes every element of the prime subfield $K \subseteq F$.

5. Let F be a field of characteristic p. Show that the map $\sigma : F \to F$ defined by $x \mapsto x^p$ is an $\mathbb{F}_p$-monomorphism.

   Hint: $(x + y)^p = x^p + y^p$ follows from the Binomial Theorem, since each of the binomial coefficients $\frac{p!}{j!(p-j)!}$ is divisible by p for $j = 1, 2, \ldots, p-1$. If $x^p = y^p$ then $(x - y)^p = x^p - y^p = 0$ by a similar argument.

22. Galois Extensions

The following gives an effective upper bound for the number of F-automorphisms of E for any finite separable extension $E \supseteq F$.

22.1 Theorem. Let $E \supseteq F$ be a finite separable extension. Then $|G(E/F)| \leq [E : F]$.

Proof. Let C be an algebraically closed field such that $C \supseteq E \supseteq F$. Then every element of $G(E/F)$ is an F-monomorphism $E \to C$, so the result follows by Theorem 21.1. □

Again, the proofs of Theorem 22.1, and of Theorem 22.2 below, appeal to the existence of an algebraically closed field C only as a convenience; it is possible to prove all these results using finite extensions. When does equality occur in the upper bound of Theorem 22.1? This is answered by

22.2 Theorem. Let $E \supseteq F$ be a finite separable extension of fields. Then the following three conditions are equivalent.

(i) The extension $E \supseteq F$ is normal.

(ii) $|G(E/F)| = [E : F]$.

(iii) F is the fixed field of $G(E/F)$.

Proof. Let $n = [E : F]$, and let C be an algebraically closed field such that $C \supseteq E \supseteq F$. Denote $G = G(E/F)$.

(i)⇒(ii) Suppose that the extension $E \supseteq F$ is normal. So E is the splitting field of some polynomial $f(X) = (X - \alpha_1)(X - \alpha_2) \cdots (X - \alpha_r) \in F[X]$, i.e. $E = F(\alpha_1, \alpha_2, \ldots, \alpha_r)$. Every F-monomorphism $\sigma : E \to C$ permutes the zeroes $\alpha_1, \alpha_2, \ldots, \alpha_r$ of $f(X)$, and so $E^\sigma = E$, i.e. $\sigma \in G$. Conversely, if $\sigma \in G$ then $\sigma : E \to E \subseteq C$ is an F-monomorphism. So G is exactly the set of all F-monomorphisms $E \to C$. By Theorem 21.1, we have $|G| = n = [E : F]$.

(ii)⇒(iii) Suppose that $|G| = n = [E : F]$. Consider any element $\alpha \in E$ which is fixed by every $\sigma \in G$. We have $E \supseteq F[\alpha] \supseteq F$, and by Theorem 21.1, there are only $[E : F[\alpha]]$ many $F[\alpha]$-monomorphisms $E \to C$. However, G contains n such F-monomorphisms, so that $[E : F] = n \leq [E : F[\alpha]]$, which implies that $F[\alpha] = F$, i.e. $\alpha \in F$.

(iii)⇒(i) Suppose that F is the fixed field of G. Let $\alpha \in E$, and define $f(X) = \prod_{\sigma \in G}(X - \alpha^\sigma) \in E[X]$. Then every $\tau \in G$ permutes the factors of $f(X)$, so that τ fixes the polynomial $f(X)$. Since every coefficient in $f(X)$ is fixed by every $\tau \in G$, by assumption we have $f(X) \in F[X]$. Thus $\mathrm{Irr}_{\alpha,F}(X)$ divides $f(X)$, which has all of its zeroes in E, so that the extension $E \supseteq F$ is normal. □

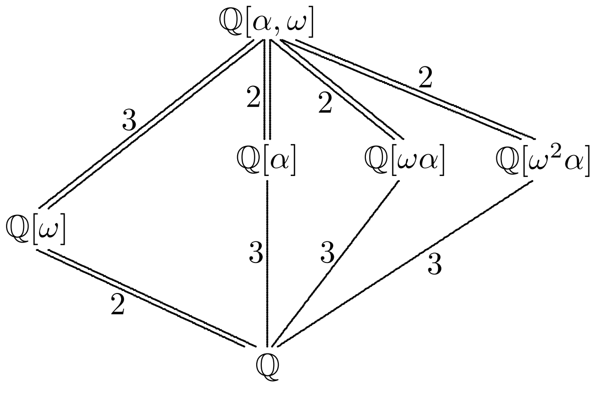
A **Galois extension** is a finite normal separable extension. Such an extension achieves equality in the upper bound of Theorem 22.1; or equivalently, satisfies any (and all) of the conditions of Theorem 22.2. For any Galois extension $E \supseteq F$, the group $G(E/F)$ is called the **Galois group** of the extension, and is usually denoted $\mathrm{Gal}(E/F)$.

Our next theorem is the central result in this course. It shows that for a Galois extension $E \supseteq F$, there is a one-to-one correspondence between intermediate fields of the extension, and subgroups of the Galois group $\mathrm{Gal}(E/F)$. This is known as the **Galois correspondence**. Before proving this result, we give a small example which illustrates the highlights of this correspondence. Consider the splitting field E of the polynomial $f(X) = X^3 - 2$ over $\mathbb{Q}$. We may factor $f(X) = (X - \alpha)(X - \omega\alpha)(X - \omega^2\alpha)$ in $\mathbb{C}[X]$, where $\alpha = \sqrt[3]{2}$, the unique real cube root of 2, and $\omega = e^{2\pi i/3}$, a primitive cube root of 1, satisfying $\omega^2 + \omega + 1 = 0$. We saw in Section 19 that the extension $E = \mathbb{Q}[\alpha, \omega] \supset \mathbb{Q}$ has degree 6. This is a Galois extension, so its Galois group $G = \mathrm{Gal}(E/\mathbb{Q})$ has order 6. What is G? First observe that since G fixes the polynomial $f(X) \in \mathbb{Q}[X]$, it must permute the three zeroes of $f(X)$. Also, any automorphism of E is determined by its action on the three zeroes $\alpha, \omega\alpha, \omega^2\alpha$, since these generate E over $\mathbb{Q}$. Therefore G may be identified as the set of all six permutations of $\{\alpha, \omega\alpha, \omega^2\alpha\}$, i.e. $G \cong S_3$. Indexing the three zeroes of $f(X)$ by

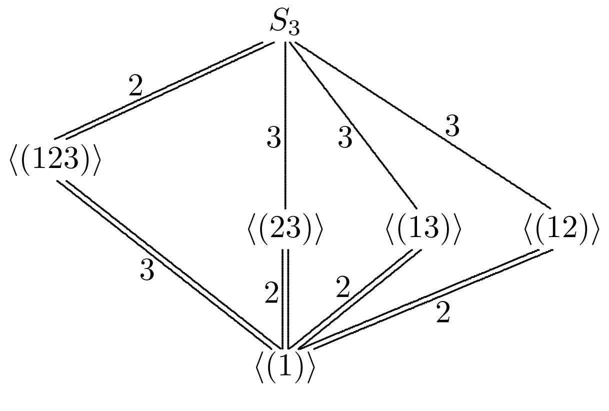
$$\alpha \leftrightarrow \text{‘1’}, \qquad \omega\alpha \leftrightarrow \text{‘2’}, \qquad \omega^2\alpha \leftrightarrow \text{‘3’},$$

and identifying G as S_3, it is clear how G acts on E. For example, $(23) \in G$ fixes the first zero α and interchanges $\omega\alpha \leftrightarrow \omega^2\alpha$. Therefore (23) interchanges $\omega \leftrightarrow \overline{\omega} = \omega^2$ and fixes all real elements of E. In other words, (23) acts on E just the same as complex conjugation! The remaining automorphisms of E are a little more subtle. For example, $(123) \in G$ cycles $\alpha \mapsto \omega\alpha \mapsto \omega^2\alpha \mapsto \alpha$. What does (123) do to ω? We have $\omega = \frac{\omega\alpha}{\alpha} \mapsto \frac{\omega^2\alpha}{\omega\alpha} = \omega$. So (123) fixes everything in the quadratic subfield $\mathbb{Q}[\omega]$. Since (123) does not fix everything in E, the fixed subfield of (123) must be just $\mathbb{Q}[\omega]$ and nothing more.

We illustrate all subgroups of $G = S_3$ and all intermediate fields of the extension $E \supset \mathbb{Q}$ in the diagrams:



$\mathbb{Q}[\alpha, \omega]$
3
2
2
2
$\mathbb{Q}[\alpha]$
$\mathbb{Q}[\omega\alpha]$
$\mathbb{Q}[\omega^2\alpha]$
$\mathbb{Q}[\omega]$
3
3
3
2
$\mathbb{Q}$



Here a double line represents a normal inclusion, and a single line represents an inclusion which is not normal. Each integer label on a line represents the corresponding index or degree. Notice that the two pictures are almost the same, except each is an upside-down image of the other. To each intermediate subfield $L \subseteq E$ on the left, there corresponds the subgroup $G_L \le G$ on the right, the set of all $\sigma \in G$ fixing every element of L. And to each subgroup $H \le G$ on the right, there corresponds the subfield $E_H \subseteq E$ on the left, the fixed field of H. If $L \supseteq L'$ on the left, then on the right this inclusion is reversed as $G_L \le G_{L'}$, and the degree $[L : L']$ equals the index $[G_{L'} : G_L]$. Moreover the extension $L \supseteq L'$ is normal iff $G_L \trianglelefteq G_{L'}$. There is a nontrivial example of this in our picture: the extension $\mathbb{Q}[\omega] \supset \mathbb{Q}$ is normal of degree 2, and the corresponding subgroups are $\langle(123)\rangle \lhd G$, of index 2. Finally, since the extension $\mathbb{Q}[\omega] \supset \mathbb{Q}$ is normal, it is Galois, and its Galois group has order 2. This may be identified with the quotient group $G/\langle(123)\rangle \cong C_2$. Why? The group G acts on $\mathbb{Q}[\omega]$, giving rise to an action $\phi : G \to \mathrm{Gal}(\mathbb{Q}[\omega]/\mathbb{Q})$ which takes each $\sigma \in G$ to its restriction to $\mathbb{Q}[\omega]$. This map ϕ is a homomorphism, and by definition, its kernel is the set of all $\sigma \in G$ which act trivially on $\mathbb{Q}[\omega]$, i.e. the set of all $\sigma \in G_{\mathbb{Q}[\omega]} = \langle(123)\rangle$. The fact that ϕ is onto $\mathrm{Gal}(\mathbb{Q}[\omega]/\mathbb{Q})$ follows by comparing orders. So $G/\langle(123)\rangle \cong \mathrm{Gal}(\mathbb{Q}[\omega]/\mathbb{Q})$ follows from the First Isomorphism Theorem for groups. All these observations we have made in this special example are consequences of the following theorem.

22.3 Fundamental Theorem of Galois Theory. Let $E \supseteq F$ be a Galois extension, with Galois group $G = \mathrm{Gal}(E/F)$. Let $\mathcal{L}$ be the class of all intermediate fields L (i.e. $E \supseteq L \supseteq F$), and let $\mathcal{H}$ be the class of all subgroups of G.

(i) The map which associates to each intermediate field L the subgroup $G_L \le G$ fixing L, is a bijection $\mathcal{L} \to \mathcal{H}$.

(ii) The map which associates to each subgroup $H \le G$ its fixed field $E_H \subseteq E$, is a bijection $\mathcal{H} \to \mathcal{L}$. This is the inverse of the bijection in (i).

(iii) We have containment $L \supseteq L'$ for intermediate fields, iff we have containment $G_{L'} \ge G_L$ of the corresponding subgroups. Moreover, in this case, $[L : L'] = [G_{L'} : G_L]$.

(iv) Suppose that $L \supseteq L'$ for intermediate fields. Then the extension $L \supseteq L'$ is normal (and hence Galois) iff $G_L \trianglelefteq G_{L'}$. If these conditions are satisfied, then $\mathrm{Gal}(L/L') \cong G_{L'}/G_L = \mathrm{Gal}(E/L')/\mathrm{Gal}(E/L)$.

Proof. Let $\phi : \mathcal{L} \to \mathcal{H}$, $L \mapsto G_L$, the map in (i); and let $\psi : \mathcal{H} \to \mathcal{L}$, $H \mapsto E_H$, the map in (ii). For each $L \in \mathcal{L}$, the extension $E \supseteq L$ is Galois with Galois group $G_L = \mathrm{Gal}(E/L)$, and by Theorem 22.2, the fixed field of G_L is L. This proves that the composite map $\mathcal{L} \xrightarrow{\phi} \mathcal{H} \xrightarrow{\psi} \mathcal{L}$ is the identity.

Consider an arbitrary subgroup $H \le G$, and let $L = E_H$, the fixed field of H. We must show that $G_L = H$. By definition, we have $G_L \ge H$. By Theorem 21.2, $E = L[\alpha]$ for some $\alpha \in L$. Define $f(X) = \prod_{\sigma \in H}(X - \alpha^\sigma) \in E[X]$. Clearly every $\tau \in H$ permutes the factors of $f(X)$, and so fixes $f(X)$. Since the coefficients in $f(X)$ are all fixed by

H, they lie in the fixed field L, i.e. $f(X) \in L[X]$. So the degree of α over L satisfies $[L[\alpha] : L] \le \deg f(X) = |H|$. Using Theorem 22.2, we have $|G_L| = [E : L] \le |H| \le |G_L|$. Therefore $G_L = H$; in other words, the composite map $\mathcal{H} \xrightarrow{\psi} \mathcal{L} \xrightarrow{\phi} \mathcal{H}$ is the identity. Therefore the maps ϕ and ψ are bijections, each the inverse of the other.

If $L \supseteq L'$ are intermediate fields, then by definition every L-automorphism of E fixes the smaller field L', i.e. $G_L \le G_{L'}$. Conversely, suppose $G_L \le G_{L'}$. Since $\mathrm{Gal}(E/L) = G_L \le G_{L'}$ fixes every element of L', Theorem 22.2 forces $L' \subseteq L$. In this case $[G_{L'} : G_L] = |G_{L'}|/|G_L| = [E : L']/[E : L] = [L : L']$, which proves (iii).

Suppose that $L \supseteq L'$ is a normal extension of intermediate fields. We will show that every $\sigma \in G_{L'}$ preserves L. Given $\alpha \in L$, let $f(X) = \mathrm{Irr}_{\alpha,L'}(X)$. Then α^σ is also a zero of $f(X)$, and since the extension $L \supseteq L'$ is normal, we have $\alpha^\sigma \in L$, Thus $L^\sigma = L$ as required. The restriction $G_{L'} \to \mathrm{Gal}(L/L')$, $\sigma \mapsto \sigma\big|_L$ is clearly a homomorphism. By definition its kernel is G_L, so we have $G_L \trianglelefteq G_{L'}$ and by the First Isomorphism Theorem for groups, $G_{L'}/G_L$ is isomorphic to a subgroup of $\mathrm{Gal}(L/L')$. Comparing orders, we get equality: $\mathrm{Gal}(L/L') \cong G_{L'}/G_L = \mathrm{Gal}(E/L')/\mathrm{Gal}(E/L)$.

Conversely, suppose that $G_L \trianglelefteq G_{L'}$. Then $G_{L'}$ must preserve L, the fixed field of G_L, by Exercise 8.4. So every $\sigma \in G_{L'}$ induces an L'-automorphism of L, and the restriction $G_{L'} \to G(L/L')$, $\sigma \mapsto \sigma\big|_L$ is a homomorphism with kernel G_L. Therefore $G_{L'}/G_L$ is isomorphic to a subgroup of $G(L/L')$. But comparing orders, using Theorem 22.1 we have $|G_{L'}/G_L| \le |G(L/L')| \le [L : L'] = [G_{L'} : G_L]$. Therefore equality holds in the upper bound of Theorem 22.1; in other words, the extension $L \supseteq L'$ is normal. □

The **Galois group** of a polynomial $f(X) \in F[X]$ over F, is by definition $\mathrm{Gal}(E/F)$ where E is a splitting field for $f(X)$ over F.

22.4 Theorem. Suppose that $f(X) \in F[X]$ is a polynomial of degree $n \ge 1$, and let G be the Galois group of $f(X)$ over F. Then G is isomorphic to a subgroup of S_n. Moreover if $f(X)$ is irreducible over F, then G is isomorphic to a transitive subgroup of S_n.

Proof. Let E be a splitting field for $f(X)$. We may assume $f(X) = (X - \alpha_1)(X - \alpha_2)\cdots(X - \alpha_n)$, $E = F(\alpha_1, \alpha_2, \dots, \alpha_n)$, and $G = \mathrm{Gal}(E/F)$. Every $\sigma \in G$ permutes the zeroes $\alpha_1, \alpha_2, \dots, \alpha_n$ of $f(X)$, and since $\alpha_1, \alpha_2, \dots, \alpha_n$ generate E over F, σ is determined by its action on $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$. Thus G may be identified as a subgroup of $\mathrm{Sym}\{\alpha_1, \alpha_2, \dots, \alpha_n\} \cong S_n$.

Suppose that $f(X)$ is irreducible in $F[X]$. Let $g(X) = \prod_{\sigma\in G}(X - \alpha_1^\sigma)$. Then every $\tau \in G$ permutes the factors of $g(X)$, and so G fixes $g(X)$. By Theorem 22.2, we have $g(X) \in F[X]$. Since α_1 is a zero of $g(X)$, it follows that $\mathrm{Irr}_{\alpha_1,F}(X) = f(X)$ divides $g(X)$. Therefore every α_i equals α_1^σ for some $\sigma \in G$. That is, G acts transitively on $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$. □

We provide two illustrations of Theorem 22.4 using the notation of the previous example. The polynomial $f(X) = X^3 - 2$ is irreducible over $\mathbb{Q}$. Its Galois group over $\mathbb{Q}$ is S_3, permuting the three zeroes $\alpha, \omega\alpha, \omega^2\alpha$ in all six possible ways.

Also, $f(X) = X^3 - 2$ is irreducible over $\mathbb{Q}[\omega]$. The Galois group of $f(X)$ over $\mathbb{Q}[\omega]$ is $G = \mathrm{Gal}(\mathbb{Q}[\alpha,\omega]/\mathbb{Q}[\omega]) \cong \langle(123)\rangle$, which cyclically permutes the three zeroes $\alpha, \omega\alpha, \omega^2\alpha$ of $f(X)$. Once again, G acts transitively on the zeroes of $f(X)$.

Recall Cayley's Representation Theorem 8.2: Every finite group G is isomorphic to a subgroup of S_n for some n. Now it is natural to ask: Is every finite group isomorphic to the Galois group of some polynomial over $\mathbb{Q}$? This is the so-called *inverse problem of Galois theory,* which has occupied the minds of many brilliant mathematicians. Despite great advances in this area, to date there is no complete solution known to this problem.

Exercises 22.

1. Show that the extension $E = \mathbb{Q}[\sqrt{2}, \sqrt{5}] \supset \mathbb{Q}$ is Galois. Compute the Galois group $G = \mathrm{Gal}(E/\mathbb{Q})$. Give diagrams illustrating the intermediate fields, and the subgroups of G, similar to those given for the example above.

2. Let F be a finite field of characteristic p. By Exercise 18.5, $|F| = q = p^r$ where p is prime and $r \geq 1$; moreover, $F \supseteq \mathbb{F}_p$ is an extension of degree r. Define $\sigma : F \to F,\ x \mapsto x^p$.

   (a) Prove that $\sigma \in \mathrm{Aut}\, F$.

   Hint: Use Exercise 21.5.

   (b) Prove that the extension $F \supseteq \mathbb{F}_p$ is separable.

   Hint: Suppose $\alpha \in F$ is a multiple zero of $f(X) = \mathrm{Irr}_{\alpha,\mathbb{F}_p}(X)$. Write $n = \deg f(X)$. Since α is a zero of the polynomial $f'(X) \in F[X]$ whose degree is less than n, we must have $f'(X) = 0$. Deduce that every nonzero term $a_j X^j$ appearing in $f(X)$ must have j divisible by p, and that $f(X) = g(X)^p$ where $g(X) \in F[X]$.

   (c) Prove that $\mathrm{Aut}\, F = \{1, \sigma, \sigma^2, \ldots, \sigma^{r-1}\}$.

   Hint: If $\sigma^i = \sigma^j$ where $1 \leq i < j < r$, then the polynomial $X^{p^j} - X^{p^i}$ of degree $p^j < q$ has q zeroes $a \in F$, which is impossible. Therefore the automorphisms $1, \sigma, \sigma^2, \ldots, \sigma^{r-1} \in \mathrm{Aut}\, F$ are distinct. Apply Theorem 22.1.

   Remark: This shows that every finite field is a Galois extension of its prime field, with cyclic Galois group; and so by the Fundamental Theorem of Galois Theory that every extension $E \supseteq F$ where E and F are finite fields, is Galois with cyclic Galois group.

23. Cyclotomic Extensions

For every positive integer n, the set of all complex n-th roots of unity, i.e. the set of all complex solutions of $z^n = 1$, is a cyclic group of order n. A **primitive n-th root of unity** is a generator of this group. Let ζ_n denote your favorite n-th root of unity; mine is $e^{2\pi i/n}$, but the choice is not actually relevant. Then the primitive n-th roots of unity are just the powers ζ_n^k for those values of $k \in \{1, 2, \ldots, n\}$ such that $\gcd(k, n) = 1$. Therefore the number of such primitive roots is just given by Euler's function

$$\varphi(n) = |\{k \in \mathbb{Z} : 1 \leq k \leq n,\ \gcd(k,n) = 1\}|.$$

A **cyclotomic extension** is an extension of the form $F(\zeta_n) \supseteq F$, where F is some extension of $\mathbb{Q}$. Clearly $F(\zeta_n) = F(1, \zeta_n, \zeta_n^2, \dots, \zeta_n^{n-1})$ is the splitting field of $X^n - 1$, so cyclotomic extensions are normal, and hence Galois. Let $G = \mathrm{Gal}(F(\zeta_n)/F)$. Then $|G| = [F(\zeta_n) : F] \leq n$, since the minimal polynomial $\mathrm{Irr}_{\zeta_n,F}(X)$ divides $X^n - 1$. But we can do much better. Define the **n-th cyclotomic polynomial** by

$$\Phi_n(X) \;=\; \prod_{\substack{1\leq k\leq n\\ \gcd(k,n)=1}} (X - \zeta_n^k).$$

This polynomial has degree $\varphi(n)$. (We have already encountered this polynomial in special cases; see Theorem 17.6 and Exercise 17.9.) Every automorphism of the Galois extension $\mathbb{Q}(\zeta_n) \supseteq \mathbb{Q}$ permutes the primitive n-th roots of unity, and so fixes the polynomial $\Phi_n(X)$; therefore by Theorem 22.2, we have $\Phi_n(X) \in \mathbb{Q}[X]$. (In fact one can show that $\Phi_n(X) \in \mathbb{Z}[X]$, although we will not need this.) So in an arbitrary cyclotomic extension $F(\zeta_n) \supseteq F$, the minimal polynomial $\mathrm{Irr}_{\zeta_n,F}(X)$ divides $\Phi_n(X)$, and so $[F(\zeta_n) : F] \leq \varphi(n)$.

23.1 Theorem. Let $F(\zeta_n) \supseteq F$ be a cyclotomic extension. Then the Galois group $G = \mathrm{Gal}(F(\zeta_n)/F)$ is abelian. Moreover, $|G| = [F(\zeta_n) : F] \leq \varphi(n)$, and equality holds iff $\Phi_n(X)$ is irreducible in $F[X]$.

Proof. Let $\sigma \in \mathrm{Gal}(F(\zeta_n)/F)$. Then σ maps ζ_n to another primitive n-th root of unity, say $\zeta_n^\sigma = \zeta_n^k$, where $\gcd(k,n) = 1$. If also $\tau \in \mathrm{Gal}(F(\zeta_n)/F)$ then $\zeta_n^\tau = \zeta_n^\ell$ for some integer ℓ with $\gcd(\ell, n) = 1$. Then both $\sigma\tau$ and $\tau\sigma$ map $\zeta_n \mapsto \zeta_n^{k\ell}$. Since the F-automorphisms $\sigma\tau$ and $\tau\sigma$ have the same effect on the generator ζ_n, we must have $\sigma\tau = \tau\sigma$; that is, $\mathrm{Gal}(F(\zeta_n)/F)$ is abelian. The remaining assertion follows from the remarks above. □

We omit the proof of the following result, since it is not required in Section 24. Note that in the special case n is a prime power, the result follows from Exercise 17.9 together with Theorem 23.1. For a proof in the general case, see e.g. Garling [1].

23.2 Theorem. The polynomial $\Phi_n(X)$ is irreducible in $\mathbb{Q}[X]$. Therefore $[\mathbb{Q}(\zeta_n) : \mathbb{Q}] = \varphi(n)$, and the Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_n)/\mathbb{Q})$ is isomorphic to the group of units of the ring $\mathbb{Z}/n\mathbb{Z}$.

Exercises 23.

1. Compute $\Phi_n(X)$ explicitly for $n = 1, 2, 3, \dots, 10$.
2. Let n be a positive integer. Prove that $X^n - 1 = \prod_{d|n} \Phi_d(X)$, where the product is over all positive integers d dividing n.
3. Verify Theorem 23.2 in the special case n is a prime power.

24. Radical Extensions

In this final section, we briefly indicate the key ideas in the application of Galois theory to solvability of polynomials by radicals, as alluded to in the Prologue. An extension $E \supseteq F$ is an **elementary radical extension** if $E = F[\alpha]$ for some $\alpha \in E$ such that $\alpha^n \in F$ for some integer $n \geq 1$. This means that α is an n-th root of some element of F. More generally, $E \supseteq F$ is a **radical extension** (or **extension by radicals**) if there exist intermediate subfields

$$E = E_k \supseteq E_{k-1} \supseteq E_{k-2} \supseteq \cdots \supseteq E_1 \supseteq E_0 = F$$

such that each extension $E_i \supseteq E_{i-1}$ is an elementary radical extension. Clearly, to say that a number x is expressible in terms of elements of F using $+$, $-$, $\times$ and $/$, together with the extraction of roots, is equivalent to saying that x lies in some radical extension of F.

A polynomial $f(X) \in F[X]$ is **solvable by radicals** if there exists a radical extension $E \supseteq F$ which contains a splitting field for $f(X)$ over F. It is important to realize that we do not require E itself to be a splitting field for $f(X)$; very often E will be much larger. It is clear that what we mean by saying that the zeroes of $f(X)$ are expressible in terms of the coefficients of $f(X)$ using the standard four operations plus extraction of roots, is precisely the property of solvability of $f(X)$ by radicals.

In our example of $X^3 - 2 = (X - \alpha)(X - \omega\alpha)(X - \omega^2\alpha) \in \mathbb{Q}[X]$ where $\alpha = \sqrt[3]{2}$ and $\omega = e^{2\pi i/3}$, we see that each of the extensions $\mathbb{Q}[\alpha] \supset \mathbb{Q}$ and $\mathbb{Q}[\alpha, \omega] \supset \mathbb{Q}[\alpha]$ is an elementary radical extension, and so $\mathbb{Q}[\alpha, \omega] \supset \mathbb{Q}$ is an extension by radicals (but not an elementary radical extension). Since $\mathbb{Q}[\alpha] \supset \mathbb{Q}$ is itself a radical extension, we see that a radical extension is not necessarily normal.

24.1 Theorem. Suppose that F is a field of characteristic zero, and let $f(X) \in F[X]$. Then $f(X)$ is solvable by radicals over F, iff the Galois group of $f(X)$ over F is solvable.

We will not have time to fully prove this, but will give some indication of its proof. We observed that a radical extension is not necessarily Galois. However, consider an elementary radical extension $F[\alpha] \supseteq F$. This will be a Galois extension if F contains enough roots of unity, as we proceed to show. For example while $\mathbb{Q}[\alpha] \supset \mathbb{Q}$ is not Galois where $\alpha = \sqrt[3]{3}$, yet by adjoining a primitive cube root of unity ω, we obtain $F = \mathbb{Q}[\omega]$, and the extension $F[\alpha] \supset F$ is Galois.

24.2 Lemma. Let $E = F[\alpha] \supseteq F$ be an elementary radical extension, say $\alpha^n = \gamma \in F$. If F contains a primitive n-th root of unity, then the extension $E \supseteq F$ is Galois, and the Galois group $\mathrm{Gal}(E/F)$ is cyclic of order dividing n.

Proof. Since there exists a primitive n-th root of unity $\zeta = \zeta_n \in F$, we have $E = F[\alpha] = F[\alpha, \zeta\alpha, \zeta^2\alpha, \ldots, \zeta^{n-1}\alpha]$, which is the splitting field of $X^n - \gamma = \prod_{k=0}^{n-1}(X - \zeta^k\alpha)$ over F. So the extension $E \supseteq F$ is normal, and hence Galois.

Let $\sigma \in G = \mathrm{Gal}(E/F)$. Then σ must permute the zeroes of $X^n - \gamma$, so $\alpha^\sigma = \zeta^k\alpha$ for some integer k. Since α generates E over F, σ is uniquely determined by k. So it is reasonable to rename this σ as σ_k. Now we have a one-to-one map $G \to \langle\zeta\rangle$ given by $\sigma_k \mapsto \zeta^k$. This map is a group homomorphism since $(\alpha^{\sigma_k})^{\sigma_\ell} = (\alpha\zeta^k)\zeta^\ell = \alpha\zeta^{k+\ell} = \alpha^{\sigma_{k+\ell}}$. Thus G is isomorphic to a (possibly proper) subgroup of $\langle\zeta\rangle$, and so G is cyclic of order dividing n. □

Consider an arbitrary radical extension

$$E = E_k \supseteq E_{k-1} \supseteq E_{k-2} \supseteq \cdots \supseteq E_1 \supseteq E_0 = F$$

where $E_i = E_{i-1}[\alpha_i]$ and α_i is an n_i-th root of $\gamma_i \in E_{i-1}$. Let $n = \mathrm{lcm}(n_1, n_2, \ldots, n_k)$. We adjoin a primitive n-th root of unity to obtain new fields $E_i' = E_i(\zeta_n)$. This gives a new tower of extensions

$$E' = E_k' \supseteq E_{k-1}' \supseteq E_{k-2}' \supseteq \cdots \supseteq E_1' \supseteq E_0' = F(\zeta_n) \supseteq F.$$

Each of the extensions $E_i' \supseteq E_{i-1}'$ is Galois with cyclic Galois group, by Lemma 24.2. Also, the last extension $F(\zeta_n) \supseteq F$ is Galois, with abelian Galois group, by Theorem 23.1. We are not quite home free, since we still don't know that E' (or E) is a Galois extension of F. Fortunately it is possible to reduce to this case; for details, see Garling [1]. Assuming however that both E and E' are Galois extensions of F, then by the Fundamental Theorem 22.3, we have a normal series for $\mathrm{Gal}(E'/F)$ given by

$$\begin{aligned} 1 = \mathrm{Gal}(E'/E') \trianglelefteq \mathrm{Gal}(E'/E_{k-1}') \trianglelefteq \mathrm{Gal}(E'/E_{k-2}') \trianglelefteq \cdots \\ \cdots \trianglelefteq \mathrm{Gal}(E'/E_1') \trianglelefteq \mathrm{Gal}(E'/E_0') \trianglelefteq \mathrm{Gal}(E'/F) \end{aligned}$$

in which the factors are $\mathrm{Gal}(E'/E_{k-1}')$, $\mathrm{Gal}(E_{k-1}'/E_{k-2}')$, $\ldots$, $\mathrm{Gal}(E_1'/E_0')$, and $\mathrm{Gal}(E_0'/F)$, all of which are abelian. Therefore $\mathrm{Gal}(E'/F)$ is solvable.

By definition, E contains a splitting field K for $f(X)$. Now $E' = E(\zeta_n) \supseteq E \supseteq K \supseteq F$, in which each containment is normal, and so $\mathrm{Gal}(K/F) \cong \mathrm{Gal}(E'/F)/\mathrm{Gal}(E'/K)$ is solvable. This gives a proof of one direction of Theorem 24.1, except with gaps as mentioned above. That's all we'll say about the proof of Theorem 24.1.

If one assumes Theorem 24.1, then every polynomial $f(X) \in F[X]$ of degree at most four is solvable by radicals. This is because, by Theorem 22.4, the Galois group of $f(X)$ over F is isomorphic to a subgroup of S_n where $n = \deg f(X) \leq 4$, and this group is solvable (see Section 13).

This is not the case for polynomials of degree greater than or equal to 5. Suppose that $f(X) \in F[X]$ is irreducible of degree 5, so that the Galois group of $f(X)$ over F is a transitive subgroup of S_5. Up to equivalence, there are five such permutation groups (Exercise 10.2), and each of these five groups occurs as the Galois group of some irreducible polynomial of degree 5 over $\mathbb{Q}$. We list representatives of each of these five types, as given by Garling [1]:

(i) $X^5 + X^4 - 4X^3 - 3X^2 + 3X + 1$, with Galois group $\langle(12345)\rangle \cong C_5$;

(ii) $X^5 - 5X + 12$, with Galois group $\langle(12345), (25)(34)\rangle \cong D_5$;

(iii) $X^5 - 2$, with Galois group $\langle(12345), (1243)\rangle$ of order 20;

(iv) $X^5 + 20X + 16$, with Galois group A_5, and

(v) $X^5 - 4X + 2$, with Galois group S_5.

Polynomials (i)–(iii) above are solvable by radicals, since they have solvable Galois groups over $\mathbb{Q}$; polynomials (iv) and (v) are not solvable by radicals, since their Galois groups are nonsolvable.

Exercises 24.

1. This exercise shows that the polynomial (i) above has solvable Galois group $\cong C_5$, and that it is solvable by radicals (in fact, using 11th roots of unity).

   Let $\zeta = e^{2\pi i/11} \in \mathbb{C}$ and let $\alpha = \zeta + \zeta^{-1}$. Since $\zeta^{11} = 1 \in \mathbb{Q}$, the cyclotomic extension $E = \mathbb{Q}(\zeta) \supset \mathbb{Q}$ is an elementary radical extension. Recall (Theorem 23.2) that the Galois group of this extension is $\mathrm{Gal}(E/\mathbb{Q}) \cong C_{10}$.

   (a) Prove that $\alpha \in \mathbb{R}$. Conclude that the field $F = \mathbb{Q}(\alpha)$ is a *proper* subfield of E.

   (b) Show that ζ is a zero of a quadratic polynomial in $F[X]$. Conclude that $[E : F] = 2$ and that $\mathrm{Gal}(E/F) \cong C_2$.

   (c) Using the Fundamental Theorem of Galois Theory, show that the extension $F \supset \mathbb{Q}$ is Galois and that $\mathrm{Gal}(F/\mathbb{Q}) \cong C_5$.

   (d) Show that α is a zero of the polynomial $f(X) = X^5 + X^4 - 4X^3 - 3X^2 + 3X + 1 \in \mathbb{Q}[X]$ given in (i) above.

   (e) Show that $f(X)$ is irreducible and that $f(X) = \mathrm{Irr}_{\alpha,\mathbb{Q}}(X)$.

   (f) Conclude that F is the splitting field of $f(X)$ and so the Galois group of $f(X)$ is isomorphic to C_5.

Appendix: Zorn's Lemma

At a couple points during the course we have benefited from Zorn's Lemma. Here we outline the statement of Zorn's Lemma and give an example of its use.

Let S be a set. A **partial order** on S is a binary relation $\leq$ such that for all $x, y, z \in S$,

(i) $x \leq x$;
(ii) if $x \leq y$ and $y \leq x$, then $x = y$; and
(iii) if $x \leq y$ and $y \leq z$, then $x \leq z$.

Note that there can be many pairs of elements $\{x, y\}$ in X which are incomparable, i.e. $x \not\leq y$ and $y \not\leq x$. A **chain** is a subset $C \subseteq x$ such that for all $x, y \in S$, either $x \leq y$ or $y \leq x$. We write $x < y$ as an abbreviation for the statement that '$x \leq y$ and $x \neq y$'. If $S \subseteq X$, an *upper bound* for S is an element $b \in X$ such that $s \leq b$ for all $s \in S$. We say that S is **bounded above** if such an upper bound for S exists. Note that b is *not* required to belong to the subset S in this case. A **maximal element** in X is an element $m \in X$ such that no element of X is larger than m; that is, there does not exist $x \in X$ such that $m < x$.

Example: $\mathbb{Z}$ with Divisibility. An example is the relation of divisibility on the set of integers, in which the pair $\{4, 15\}$ is incomparable since $4 \not| \, 15$ and $15 \not| \, 4$. In this setting, $\{1, 2, 4, 8, 16, \ldots\}$ is a chain with no upper bound. The chain $\{3, 12, 36, 1440\}$ has many choices of upper bound: 1440 is an upper bound (the *least* upper bound), and 2880 is also an upper bound. There is no maximal element in $\mathbb{Z}$ for the divisibility relation.

Example: $X \subset \mathbb{Z}$ with Divisibility. Now consider the set X consisting of integers expressible as a product of at most 5 prime factors. For example, X contains $2^3 3^1 = 24$ and $2^3 3^1 7^1 = 168$ but *not* $2^3 3^1 5^1 7^1 = 840$. We use divisibility as our relation on X. Every chain in X has at most six elements. Moreover every chain $C \subset X$ has an upper bound: either $C = \emptyset$, in which case 1 (or any element of X) is an upper bound for C, or the largest element of C is an upper bound for C. The element $32 \in X$ (or, for that matter, *any* element with exactly 5 prime factors, not necessarily distinct) is a maximal element of X. Note, however, that 32 is *not* an upper bound for X.

Zorn's Lemma. Let X be a nonempty partially ordered set, and suppose every chain in X is bounded above. Then X has a maximal element.

Like most authors, we assume this result rather than proving it. The reason for this is that one cannot prove this result without assuming the Axiom of Choice (or something at least as strong). This is because Zorn's Lemma is equivalent to the Axiom of Choice, given the Zermelo-Fraenkel axioms of set theory. It is typically used as a convenient crutch, where no maximal element is explicitly constructible. This should not be of great concern, however, since in practical situations where a maximal element is desired, we can typically get by without one. We will try to make this point clear in the context of an example.

Corollary. Every vector space has a basis.

Proof. Let V be a vector space over a field F. We assume $V \neq 0$; otherwise $\emptyset$ is a basis for V.

Let $\mathcal{I}$ be the collection of all linearly independent subsets of V. Recall that a subset $S \subseteq V$ is *linearly dependent* if there exist distinct vectors $v_1, v_2, \ldots, v_k \in S$ and scalars $a_1, a_2, \ldots, a_k \in F$, not all zero, such that $a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$. Thus $S \in \mathcal{I}$ iff $S \subset V$ and S is *not* linearly dependent. Clearly $\mathcal{I}$ is nonempty, since every nonzero vector $v \in V$ gives rise to a linearly independent subset $\{v\} \in \mathcal{I}$.

Let $\mathcal{C} \subset I$ be any chain. We claim that $\mathcal{C}$ is bounded above by $\bigcup \mathcal{C}$. (Recall that $\bigcup \mathcal{C}$ is the union of all members of $\mathcal{C}$; that is, $\bigcup \mathcal{C} = \bigcup_{S \in \mathcal{C}} S$.) We must first show that $\bigcup \mathcal{C} \in \mathcal{I}$. Consider any distinct vectors $v_1, v_2, \ldots, v_k \in \bigcup \mathcal{C}$ and let $a_1, a_2, \ldots, a_k \in F$. For every $i = 1, 2, \ldots, k$, the fact that $v_i \in \bigcup \mathcal{C}$ means that $v_i \in S_i$ for some linearly independent subset $S_i \in \mathcal{C}$. Since $\mathcal{C}$ is a chain, the S_i's are totally ordered by inclusion. This means we may assume that $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_k$; at least this will be the case if $v_1, v_2, \ldots, v_k$ were listed in a suitable order. But now $v_1, v_2, \ldots, v_k$ all belong to the linearly independent set S_k, and so the scalars $a_1, a_2, \ldots, a_k$ must all be zero. This shows that $\mathcal{C}$ is linearly independent, so $\bigcup \mathcal{C} \in \mathcal{I}$. We still need to show that $\bigcup \mathcal{C}$ is an upper bound for the chain $\mathcal{C}$. But this is obvious since for every linearly independent subset $S \in \mathcal{C}$, we have $S \subseteq \bigcup \mathcal{C}$ by definition.

Let B be a maximal element for $\mathcal{I}$, which exists by Zorn's Lemma. So B is linearly independent. It remains to be shown that B spans V. Let $v \in V$. We must show that v is in the span of B. If $v \in B$ then this is clear; so we may assume that $v \notin B$, so that B is a proper subset of $B \cup \{v\}$. Since B is a maximal element of $\mathcal{I}$, it must be the case that $B \cup \{v\}$ is linearly dependent. Thus there exist distinct vectors $v_1, v_2, \ldots, v_k \in B \cup \{v\}$ and scalars $a_1, a_2, \ldots, a_k \in F$, not all zero, such that

$$a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0.$$

Clearly $v \in \{v_1, v_2, \ldots, v_k\}$ since B itself is linearly independent; we may assume that $v_1 = v$. Moreover $a_1 \neq 0$, for otherwise we have found a nontrivial linear relation between $v_2, v_3, \ldots, v_k \in B$, which cannot occur since B is linearly dependent. Thus

$$v = -a_1^{-1}\big(a_2v_2 + a_3v_3 + \cdots + a_kv_k\big)$$

lies in the span of B, as required. Thus B spans V. Since B is also linearly independent, B is a basis for V. □

For finite dimensional vector spaces, it is very easy to produce bases explicitly, and so Zorn's Lemma is not needed in such cases. For many infinite-dimensional vector spaces, this is not an option. For example, the vector space $C([0,1])$ consisting of continuous functions $[0,1] \to \mathbb{R}$, has a basis, by Zorn's Lemma. But you will never see an explicit basis for this vector space! since none can be written down. But in any practical situation in which $C([0,1])$ arises, this is not an issue since we typically deal with only certain well-known proper subspaces of $C([0,1])$ for which explicit bases are known.

Bibliography

[1] D. J. H. Garling, *A Course in Galois Theory,* Cambridge Univ. Press, 1986. Quite readable. Appropriate for the later course material.

[2] D. Gorenstein, 'The enormous theorem', *Scientific American* **253** (1985), pp.104–115. A layman's introduction to the classification of finite simple groups.

[3] I. N. Herstein, *Topics in Algebra,* Wiley, New York, 1975. A standard general algebra text.

[4] T. W. Hungerford, *Algebra,* Springer-Verlag, 1974b. Probably more useful as a general algebra reference than as a textbook.

[5] L. Infeld, *Whom the Gods Love,* Whittlesey House, New York, 1948. A well-recommended biography of the life of Evariste Galois.

[6] C. C. Pinter, *A Book of Abstract Algebra,* 2nd ed., McGraw-Hill, 1990.

[7] P. Samuel, *Algebraic Theory of Numbers,* Kershaw, London, 1972. Very helpful for ring theory, field theory and Galois theory.

Index