Groups

(Last Updated: September 7, 2018)

These notes are derived primarily from *Algebra* by Michael Artin (2ed). Most of this material is drawn from Chapter 2.

2. Groups and subgroups

A law of composition (or binary operation) on a set S is a function of two variables S: $S \times S \to S$.

Definition. A group is a pair $(G, \cdot)$ with G a set and $\cdot$ a binary operation on G satisfying

(1) Associativity: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
(2) Identity: there exists $1 \in G$ such that $a \cdot 1 = 1 \cdot a$ for all $a \in G$.
(3) Inverses: for all $a \in G$ there exists an element $b \in G$ such that $a \cdot b = b \cdot a = 1$.

If in addition, $a \cdot b = b \cdot a$ for all $a, b \in G$ (commutativity) the group is said to be abelian.

Disclaimer/Warning: When the operation is understood we often will only write the set to denote the group. The most common operation symbols are $+$, $\cdot$, and $\circ$. When the operation is addition, the inverse of $a \in G$ is typically denoted $-a$. When the operation is multiplication or composition, it is denoted a^{-1}.

Example. The following are examples of groups.

- $\mathbb{Z}^+$, the integers under addition, is a group. One can also replace $\mathbb{Z}$ by $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$ (but not $\mathbb{N}$).
- $\mathbb{Q}^\times$, the rational numbers under multiplication is a group. One can also replace $\mathbb{Q}^\times$ by $\mathbb{R}^\times$ (but not $\mathbb{Z}^\times$).
- Let $M_n(\mathbb{R})$ denote the set of $n \times n$ matrices with entries in $\mathbb{R}$. Then $M_n(\mathbb{R})$ is a group under matrix addition. One can replace $\mathbb{R}$ with any field. Note that $M_n(\mathbb{R})$ is *not* a group under matrix multiplication.
- Let $\mathrm{GL}_n(\mathbb{R}) \subset M_n(\mathbb{R})$ denote the set of $n \times n$ *invertible* matrices. Then $\mathrm{GL}_n(\mathbb{R})$ is a group under matrix multiplication. (This follows from the fact that $\det(A)\det(B) = \det(AB)$ for all $n \times n$ matrices A, B.) Again, one can replace $\mathbb{R}$ with any field. Note that $\mathrm{GL}_n(\mathbb{R})$ is *not* a group under matrix addition.

Because for a group $(G, \cdot)$, G is a set and so it makes sense to write $|G| = n$ for the number of elements in the group/set G (so $|\mathbb{Z}_n| = n$). We call n the order of the group and say a group is finite if $n < \infty$. Otherwise the group is said to be infinite.

The next proposition establishes some basic properties of groups and inverses. The proofs are straightforward and are left to the reader.

Proposition 1. Let G be a group.

(1) The identity element of G is unique.
(2) For all $g \in G$, the inverse element $g^{-1} \in G$ is unique.
(3) For $g, h \in G$, $(gh)^{-1} = h^{-1}g^{-1}$.
(4) Left and right cancellation hold. That is, for all $a, b, c \in G$,

$$ba = ca \Rightarrow b = c$$

$$ab = ac \Rightarrow b = c.$$

Exercise. Let G be a group and $g \in G$. If $g' \in G$ satisfies $gg' = 1$ or $g'g = 1$, then $g' = g^{-1}$. (A left/right inverse element in a group is a two-sided inverse).

Definition. A subgroup H of a group G is a subset such that H is a group with respect to the operation associated to G.

Example. (1) Let G be a group. Then G is a subgroup of itself. If $e \in G$ is the identity element, then $\{e\}$ is a subgroup called the trivial subgroup. A subgroup of G that is not G and not the trivial subgroup is called proper.

(2) $\mathbb{Z}2$, the set of even numbers, is a subgroup of $\mathbb{Z}^+$ (under addition). The set of odd numbers is not a subgroup. In general, for $a \in \mathbb{Z}$, $a \neq 0$, the set

$$\mathbb{Z}a = \{n \in \mathbb{Z} : n = ka \text{ for some } k \text{ in } \mathbb{Z}\}$$

is a subgroup of $\mathbb{Z}^+$.

(3) $\mathbb{R}^\times$ is a group under multiplication and $\mathbb{Q}^\times$ is a subgroup.
(4) $\mathrm{SL}_n(\mathbb{R})$, the set of real $n \times n$ matrices with determinant 1, is a subgroup of $\mathrm{GL}_n(\mathbb{R})$ under matrix multiplication.
(5) $\mathrm{GL}_n(\mathbb{R})$ is a subset of $M_n(\mathbb{R})$ but not a subgroup because $M_n(\mathbb{R})$ is a group under matrix addition and $\mathrm{GL}_n(\mathbb{R})$ a group under matrix multiplication.

In general, to check that a subgroup is a group we need to verify first that it is a subset and then check that it is a group. The next proposition simplifies that process.

Proposition 2. A subset H of a group G is a subgroup if and only if

(1) the identity element $1 \in G$ is in H;
(2) $h_1, h_2 \in H \Rightarrow h_1h_2 \in H$;
(3) $h \in H \Rightarrow h^{-1} \in H$.

The next proposition is a shortcut to the shortcut, but it is only useful in certain circumstances.

Proposition 3. Let H be a subset of G. Then H is a subgroup of G if and only if $H \neq \emptyset$ and whenever $a, b \in H$, $ab^{-1} \in H$.

3. Subgroups of the additive group of integers

Here are some well-known theorems, though the first is more of an axiom. You should be familiar with these, though it is not difficult to find a proof. Note that we use use the notation $a \mid b$ in place of a divides b. A nonempty subset S of $\mathbb{Z}$ is well-ordered if S contains a least element.

Theorem 4. (1) (The Well-Ordering Principle) Every nonempty subset of $\mathbb{N}$ contains a least element.

(2) (The Division Algorithm) Let $a, b \in \mathbb{Z}$ with $a > 0$. Then there exist unique integers q and r such that $b = aq + r$ with $0 \leq r < a$.

(3) (The Euclidean Algorithm) Let $a, b \in \mathbb{Z}$. There exist integers r, s such that $\gcd(a, b) = ar + bs$. Furthermore, the gcd of a and b is unique.

(4) Let $a, b \in \mathbb{Z}$ and p a prime number. If $p \mid ab$, then $p \mid a$ or $p \mid b$.

(5) (The Fundamental Theorem of Arithmetic) Let $n \in NN$. Then $n = p_1 p_2 \cdots p_k$ where the p_i are prime. Furthermore, if $n = q_1 q_2 \cdots q_\ell$ where the q_i are prime, then $k = \ell$ and the q_i are a rearrangement of the p_i.

Theorem 5. Let S be a subgroup of $\mathbb{Z}^+$. Then either $S = \{0\}$, the trivial group, or else $S = \mathbb{Z}a$ for some nonzero $a \in \mathbb{Z}$.

Proof. Suppose S is not trivial and let $a \in S$ be a minimal positive integer. Such an integer exists because S must contain positive integers (because the inverse of a negative integer is positive and S is closed under inverses), and by the Well-Ordering Principle on $\mathbb{N}$. We claim that $S = \mathbb{Z}a$.

Note first that since $a \in S$, then $ka = a + a + \cdots + a \in S$ by closure. Thus, $\mathbb{Z}a \subset S$. We now must show that $S \subset \mathbb{Z}a$.

Let $b \in S$ be positive. The case of $b \in S$ negative is similar. By the division algorithm, there exist unique integers q and r such that $b = aq + r$ with $0 \leq r < b$. But $a, b \in S$ and so $r = b - aq \in S$. This contradicts our choice of a unless $r = 0$, whence $b \in \mathbb{Z}a$. □

Let $a, b \in \mathbb{Z}$ not both zero. Set $d = \gcd(a, b)$ and $m = \operatorname{lcm}(a, b)$. Define

$$\mathbb{Z}a + \mathbb{Z}b = \{n \in \mathbb{Z} : n = ra + sb \text{ for some } r, s \in \mathbb{Z}\}.$$

This is the subgroup of $\mathbb{Z}^+$ generated by a and b. One can check that $\mathbb{Z}a + \mathbb{Z}b = \mathbb{Z}d$ and $\mathbb{Z}a \cap \mathbb{Z}b = \mathbb{Z}m$.

A digression on the symmetric group

Definition. A permutation is a bijective function on the set X (from X to itself). The set of permutations on X is denoted $\mathcal{S}_X$. If X is finite we write $X = \{1, \ldots, n\}$ and denote $\mathcal{S}_X$ by $\mathcal{S}_n$ and call it the symmetric group on n letters.

Theorem 6. S_n is a group of order $n!$ under composition.

Proof. The composition of two bijective functions is again a bijective function (see Chapter 1). Moreover, the operation of composition is associative (check!). The identity function is given by $\mathrm{id}(x) = x$ for all $x \in X$ and a bijective function is invertible. The last part of the theorem follows because there are $n!$ permutations of a set with n elements. □

There are two standard types of notation to represent elements of $\mathcal{S}_n$: two-line and cycle. Two-line is in some ways easier to use at first but much clumsier. We will learn both but as we go on we will use cycle notation much more frequently.

In two-line notation we write the elements of $\mathcal{S}_n$ as $2 \times n$ matrices. For a given element $\pi \in \mathcal{S}_n$ we write in the first row $1, \ldots, n$ and in the second the image of each value under π:

$$\pi = \begin{pmatrix} 1 & 2 & \cdots & n \\ \pi(1) & \pi(2) & \ldots & \pi(n) \end{pmatrix}.$$

Disclaimer/Warning: The elements of S_n are functions and therefore we compose right-to-left. As one would expect with function the composition, the elements of $\mathcal{S}_n$ do not commute.

Example. Consider the following elements of S_3:

$$\pi = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

Then

$$\pi\tau = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad \text{and} \quad \tau\pi = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}.$$

A more compact way of representing elements of $\mathcal{S}_n$ is with *cycles.*

Definition. A permutation $\pi \in \mathcal{S}_n$ is a cycle of length k if there exists $a_1, \ldots, a_k \in \{1, \ldots, n\}$ such that

$$\pi(a_1) = a_2, \quad \pi(a_2) = a_3, \quad \ldots \quad \pi(a_k) = a_1$$

and $\pi(i) = i$ for $i \notin \{a_1, \ldots, a_k\}$. We denote the cycle by $(a_1 \ a_2 \ \cdots \ a_k)$.

To compose cycles, one *could* translate back to two-line notation but I strongly advise against that. Instead, we compose (from right-to-left) by tracking the image of each element through successive cycles, remembering to close cycles when we get back to where we started.

Example. In the previous example, the elements would be written in cycle notation by

$$\mathrm{id} = (1), \quad \pi = (1\ 4\ 3\ 2), \quad \tau = (1\ 3)(2\ 4), \quad \mu = (1\ 2\ 3\ 4).$$

Then

$$\pi\tau = (1\,4\,3\,2)(1\,3)(2\,4) = (2\,3\,4\,1) = (1\,2\,3\,4) = \mu.$$

Example. Consider $\mathcal{S}_3$. By the theorem, $\mathcal{S}_3$ has six elements. Let x be the cyclic permutation $(1\ 2\ 3)$ and y the transposition $(1\ 2)$. Of course, $x^3 = 1$ and $y^2 = 1$ where 1 is the trivial permutation. One can also check that $yx = x^2y$. (In both cases, it is the permutation $(2\,3)$). Thus, a complete list of elements of $\mathcal{S}_3$ is $\{1, x, x^2, y, xy, x^2y\}$.

Definition. Two cycles $\pi = (a_1\ a_2\ \cdots\ a_k)$ and $\tau = (b_1\ b_2\ \cdots\ b_\ell)$ are said to be disjoint if $a_i \neq b_j$ for all i, j.

Example. $(1\,3\,5)(2\,7)$ are disjoint but $(1\,3\,5)(3\,4\,7)$ are not. Note that $(1\,3\,5)(3\,4\,7) = (3\,4\,7\,5\,1)$.

Proposition 7. Disjoint cycles in $\mathcal{S}_n$ commute.

Proof. Let $\pi, \tau \in \mathcal{S}_n$ be disjoint. Write $\pi = (a_1\ a_2\ \cdots\ a_k)$ and $\tau = (b_1\ b_2\ \cdots\ b_\ell)$. We claim $(\pi\tau)(x) = (\tau\pi)(x)$ for all $x \in \{1, \ldots, n\}$.

Suppose $x \notin \{a_1, \ldots, a_k\}$ and $x \notin \{b_1 \ldots, b_\ell\}$. By definition of a cycle, $(\pi\tau)(x) = \pi(\tau(x)) = \pi(x) = x$ and similarly $(\tau\pi)(x) = \tau(\pi(x)) = \tau(x) = x$.

Now suppose $x \in \{a_1, \ldots, a_k\}$ (so $x \notin \{b_1, \ldots, b_\ell\}$). Then $\pi(x) \in \{a_1, \ldots, a_k\}$ and so $\pi(x) \notin \{b_1, \ldots, b_\ell\}$. Thus $(\pi\tau)(x) = \pi(\tau(x)) = \sigma(x)$ and $(\tau\pi)(x) = \tau(\pi(x)) = \sigma(x)$.

The proof for $x \in \{b_1, \ldots, b_\ell\}$ is similar. □

The next theorem says that we can decompose cycles in disjoint pieces. It also gives the algorithm for doing this.

Theorem 8. Every element in $\mathcal{S}_n$ can be written as the product of disjoint cycles.

Proof. Set $X = \{1, \ldots, n\}$. First we will decompose X into disjoint pieces and use these to define the cycles. Choose $\sigma \in \mathcal{S}_n$ and define $X_1 = \{1, \sigma(1), \sigma^2(1), \ldots\}$. Then X_1 is finite because σ has finite order. Choose $k \in X \backslash X_1$ and define $X_2 = \{k, \sigma(k), \sigma^2(k), \ldots\}$. Continue in this way. Note that the process *must* end because X is finite. Write $X = X_1 \cup X_2 \cup \cdots \cup X_r$.

Define a cycle σ_i by

$$\sigma_i(x) = \begin{cases} \sigma(x) & x \in X_i \\ x & x \notin X_i. \end{cases}$$

Then $\sigma = \sigma_1\sigma_2 \cdots \sigma_r$. Note that the σ_i are disjoint because the X_i are. □

Example. Write $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 1 & 5 & 6 & 4 \end{pmatrix}$ as a product of disjoint cycles.

Starting with 1 we have $\sigma_1 = (1\ 3)$. We choose a value, say 2, that is not yet accounted for and continue, $\sigma_2 = (2)$. Next we choose 4 and continue, $\sigma_3 = (4\ 5\ 6)$. This exhausts $\{1, \ldots, 6\}$ and so $\sigma = \sigma_1\sigma_2\sigma_3 = (1\ 3)(2)(4\ 5\ 6)$.

Note that (2) is equivalent to the identity so it is appropriate to omit it.

This decomposition is not unique. What is unique is the parity (even/odd) of the decomposition. This is a bit harder (but doable!) to prove and we don't do it here. We say a permutation is even if it can be written as an even number of permutations and odd otherwise. We define the sign of a permutation σ to be 1 if σ is even and -1 if it odd.

Here's another way to view the sign map. Let e_{ij} represent the $n \times n$ matrix units. That is, e_{ij} has a 1 in the i, j position and zeros elsewhere. Given a permutation $\sigma \in \mathcal{S}_n$, define a permutation matrix P by

$$P = \sum_i e_{\sigma(i),i}.$$

Now if $\mathbf{x}$ is an n-dimensional column vector, then

$$P\mathbf{x} = \begin{bmatrix} x_{\sigma(1)} & \ldots & x_{\sigma(n)} \end{bmatrix}.$$

The sign of the permutation σ is the determinant of the matrix P.

4. Cyclic groups

We now move on to study cyclic groups, which are the building blocks of all finite(ly generated) abelian groups.

Let G be a group and $a \in G$. The set

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

is a subgroup of G (this is not difficult to check).

Definition. Let G be a group. For $a \in G$, $\langle a \rangle$ is called the cyclic subgroup of G generated by a. If there exists $a \in G$ such that $\langle a \rangle = G$, then G is said to be a cyclic group and a a generator of G. The order of a is the smallest positive integer such that $a^n = e$ ($na = 0$ in additive notation) and we write $|a| = n$. If no such n exists then we say the order is infinite and write $|a| = \infty$.

It is also not hard to check that every cyclic group is abelian. The next proposition lays out how cyclic groups are related to subgroups of $\mathbb{Z}^+$.

Proposition 9. Let G be a group and $a \in G$. Set $H = \langle a \rangle$ and S to be the set of integers k such that $a^k = 1$.

(1) The set S is a subgroup of $\mathbb{Z}^+$.
(2) Two powers $a^r = a^s$, with $r \geq s$, if and only if $a^{r-s} = 1$ if and only if $r - s \in S$.
(3) Suppose that S is not the trivial group, so $S = \mathbb{Z}n$ for some positive integer n. The powers $1, a, a^2, \ldots, a^{n-1}$ are the distinct elements of $\langle a \rangle$ and the order of $\langle a \rangle$ is n.

Proof. We prove only the third item. By Theorem 5, $S = \mathbb{Z}n$ where n is the minimal positive integer such that $n \in \mathbb{Z}$. If $a^k \in S$, then by the Division algorithm there exists $q, r \in \mathbb{Z}$, $0 \leq r < k$, such that $k = qn + r$. Then $a^k = a^{qn+r} = (a^n)^q a^r = a^r$ so $a^k \in \{1, a, a^2, \ldots, a^{n-1}\}$. If $a^r = a^s$ for $0 \leq r, s < n$, then $r - s = 0$ by (2), a contradiction. Hence, these elements are distinct. □

We say the order of an element $a \in G$ is the order of the cyclic group $\langle a \rangle$.

Proposition 10. Let a be an element of finite order n in a group, and let k be an integer that is written as $k = nq + r$ for $q, r \in \mathbb{Z}$ with $0 \leq r < n$.

(1) $a^k = a^r$
(2) $a^k = 1$ if and only if $r = 0$.
(3) Let $d = \gcd(k, n)$. The order of a^k is n/d.

One can generalize the above and speak of groups generated by more elements. Let U be a subset of G. We define the subgroup generated by U to be the smallest subgroup of G containing U.

Example. The Klein four group V is the group consisting of the four matrices

$$\begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}.$$

This is the simplest group that is not cyclic. Note that the square of any element is the identity.

Example. The quaternion group H consists of eight matrices $H = \{\pm\mathbf{1}, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}\}$ where

$$\mathbf{1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{i} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \mathbf{k} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

Any two of $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ generate H. One can check that these elements satisfy the rules,

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -\mathbf{1}, \quad \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \quad \mathbf{jk} = -\mathbf{kj} = \mathbf{i}, \quad \mathbf{ki} = -\mathbf{ik} = \mathbf{j}.$$

More concisely, these last three rules can be summed up with the single rule $\mathbf{ijk} = -\mathbf{1}$.

5. Homomorphisms and Isomorphisms

One should think of a homomorphism as a *structure preserving map* between two groups G and H. In fact, whenever you see the word *morphism*, this is generally what is meant (though not always for groups).

Definition. A homomorphism is a function $\phi : (G, \cdot) \to (H, \circ)$ between groups such that

$$\phi(g_1 \cdot g_2) = \phi(g_1) \circ \phi(g_2) \quad \text{for all } g_1, g_2 \in G.$$

If ϕ is bijective, then we say it is an isomorphism and that G and H are isomorphic.

Example. The following maps are homomorphisms:

(1) the determinant function $\det : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^\times$.

(2) the sign homomorphism $\sigma : \mathcal{S}_n \to \{\pm 1\}$ that sends a permutation to its sign.

(3) the exponential map $\exp : \mathbb{R}^+ \to \mathbb{R}^\times$ defined by $x \mapsto e^x$. The exponential map is not an isomorphism unless we restrict the codomain to only positive real numbers.

(4) the map $\phi : \mathbb{Z}^+ \to G$ by $\phi(n) = g^n$ where g is a particular element in the group G. If g has infinite order, then ϕ is an isomorphism.

(5) the absolute value map $|\ | : \mathbb{C}^\times \to \mathbb{R}^\times$.

(6) the trivial homomorphism $\phi : G \to G$ given by $\phi(x) = x$ for all $x \in G$.

(7) the inclusion homomorphism $\phi : H \to G$ for a subgroup H of G given by $\phi(x) = x$ for all $x \in G$.

(8) the map $\phi : \mathcal{S}_n \to \mathcal{P}$ where $\mathcal{P}$ is the set of $n \times n$ permutation matrices (this is a subgroup of GL_n) that sends a permutation to its associated matrix is an isomorphism.

Proposition 11. Let $\phi : G \to H$ be a homomorphism of groups.

(1) If $a_1 \cdots a_k \in G$, then $\phi(a_1 \cdots a_k) = \phi(a_1) \cdots \phi(a_k)$.

(2) $\phi(1_G) = 1_H$.

(3) For any $g \in G$, $\phi(g^{-1}) = \phi(g)^{-1}$.

Definition. Let $\phi : G \to H$ be a homomorphism of group. The set $\operatorname{im}\phi = \{\phi(g) : g \in G\}$ is called the image of ϕ. The kernel of a homomorphism $\phi : G \to H$ is the set

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

Exercise. Verify that the image and kernel of a homomorphism are subgroups of H and G, respectively.

Example. (1) The kernel of the determinant homomorphism $\det : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^\times$ is $\mathrm{SL}_n(\mathbb{R})$.

(2) The kernel of the sign homomorphism $\sigma : \mathcal{S}_n \to \{\pm 1\}$ is the alternating group. That is, the alternating group A_n is the group of even permutations.

Definition. Let G be a group and H a subgroup of G. A left coset of H with representative in G is the set

$$gH = \{gh : h \in H\}.$$

Proposition 12. Let $\phi : G \to G'$ be a homomorphism of groups and let $a, b \in G$. Let K be the kernel of ϕ. The following are equivalent:

(1) $\phi(a) = \phi(b)$,
(2) $a^{-1}b \in K$,
(3) $b \in aK$,
(4) $bK = aK$.

Corollary 13. A homomorphism $\phi : G \to G'$ is injective if and only if $\ker \phi = \{1\}$.

Definition. A subgroup N of a group G is normal if $gng^{-1} \in N$ for all $g \in G$, $n \in N$.

Theorem 14. Let $\phi : G \to H$ be a group homomorphism. Then $\ker \phi$ is a normal subgroup of G.

Proof. In a previous exercise you verified that $\ker \phi$ was a subgroup. We will prove here that $\ker \phi$ is normal. Let $g \in G$ and $k \in \ker \phi$. Then

$$\phi(gkg^{-1}) = \phi(g)\phi(k)\phi(g)^{-1} = \phi(g)1_H\phi(g)^{-1} = 1_H.$$

Thus, $g \ker \phi g^{-1} \subset \ker \phi$ and so $\ker \phi$ is normal. □

Example. (1) The special linear group $\mathrm{SL}_n(\mathbb{R})$ is a normal subgroup of $\mathrm{GL}_n(\mathbb{R})$.
(2) The alternating group A_n is a normal subgroup of $\mathcal{S}_n$.
(3) Any subgroup of an abelian group is normal.
(4) The cyclic subgroup $\langle y \rangle$ in $\mathcal{S}_3$ is not a normal subgroup.
(5) The center of a group G, defined as

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

is always a normal subgroup.

Proposition 15. Let $\phi : G \to H$ be a homomorphism of groups.

(1) If G' is a subgroup of G, then $\phi(G')$ is a subgroup of H.
(2) If H' is a subgroup of H, then $\phi^{-1}(H') = \{g \in G : \phi(g) \in H'\}$ is a subgroup of G containing $\ker \phi$.
(3) H' is a normal subgroup of H if and only if $\phi^{-1}(H')$ is a normal normal subgroup in G.

Lemma 16. If $\phi : G \to H$ is an isomorphism, the inverse map $\phi^{-1} : H \to G$ is also an isomorphismm.

Definition. An automorphism is an isomorphism from a group G to itself.

Example. Let G be a group and $g \in G$. The map $\phi : G \to G$ given by $\phi(x) = gxg^{-1}$ (conjugation by g) is an automorphism of G.

7. Equivalence relations and partitions

Next we'll discuss equivalence relations on sets. This will tie quickly back into the idea of cosets and quotient groups.

Definition. An equivalence relation on a set S is a relation $\sim$ satisfying the following rules.

- For all $x \in X$, $x \sim x$ (reflexive).
- If $x \sim y$, then $y \sim x$ (symmetric).
- If $x \sim y$ and $y \sim z$, then $x \sim z$ (transitive).

The equivalence class of $x \in S$ is the set $C_x = \{y \in S : x \sim y\}$.

Example. Conjugacy is an example of an equivalence relation on a group. Let G be a group. We say $x \sim y$ if there exists $g \in G$ such that $y = gxg^{-1}$.

Definition. A partition P of a set S is a collection of nonempty sets $S_1, S_2, \ldots$ such that $S_i \cap S_j = \emptyset$ for all $i \neq j$ and $\bigcup_k S_k = S$.

Proposition 17. An equivalence relation on a set S determines a partition, and conversely.

Proof. Let $\sim$ be an equivalence relation on S. We claim that the equivalence classes partition S. If $a \in S$, then by the reflexive property $a \in C_a$. Hence, the union of all equivalence classes is S. Now we must show that the distinct equivalence classes are disjoint.

Suppose $C_a \cap C_b \neq \emptyset$. Choose an element d in the intersection, so $a \sim d$ and $b \sim d$. If $x \in C_b$, then $b \sim x$. Now by symmetry, $a \sim d$, $d \sim b$, and $b \sim x$. Thus, by transitivity, $a \sim x$, and so $x \in C_a$. Hence, $C_b \subset C_a$ and a similar proof shows that $C_a \subset C_b$. It now follows that $C_a = C_b$. This says that two equivalence relations are either equal or disjoint.

Conversely, let P be a partition of S. We define an equivalence relation by saying that $a \sim b$ if a and b lie in the same subset of P. We leave it to the reader to verify that this is a valid equivalence relation. □

Given a set S and an equivalence relation $\sim$, we denote by $\overline{S}$ the set of equivalence class of S under $\sim$. Instead of using the above notation for equivalence classes, it will be convenient to represent them like elements in S, but with a bar. There is a natural surjective map

$$\pi : S \to \overline{S}$$

that maps an element a in S to its equivalence class $\overline{a}$. Hence, if $a \sim b$, then in $\overline{S}$ we have $\overline{a} = \overline{b}$, and vice-versa.

Any map of sets $f : S \to T$ gives us an equivalence relation on S defined by $a \sim b$ if $f(a) = f(b)$. The inverse image of an element $t \in T$ is the set

$$f^{-1}(t) = \{s \in S : f(s) = t.\}$$

That is, $f^{-1}(t)$ is the subset of S consisting of preimages of t. These are called the fibres of the map f and the non-empty fibres are the equivalence classes for the relation above.

Let $\phi : G \to G'$ be a group homomorphism. Then ϕ defines an equivalence relation as in the previous example, though we usually write $\equiv$ instead of $\sim$ and read it as congruence. So $a \equiv b$ if $\phi(a) = \phi(b)$. As we saw before, this holds if and only if $b \in aK$ where K is the kernel of ϕ.

Proposition 18. Let K be the kernel of a homomorphism $\phi : G \to G'$. The fibre of ϕ that contains an element a of G is the coset aK of K. These cosets partition the group G, and they correspond to elements of the image of ϕ.

A digression on fibres

Let $f : S \to S'$ be a map of sets and let $\overline{S}$ denote the set of fibres of f. For an element $x \in S$, denote its fibre by $\overline{x}$. This is just the equivalence class to which it belongs according to the equivalence relation determined by $a \sim b$ if $f(a) = f(b)$.

Define a new map $\pi : S \to S'$ that sends an element $x \in S$ to its fibre in $\overline{S}$, so $\pi(x) = \overline{x}$. Then there is a *unique* map $\overline{f}$ that makes the following diagram commute:

$$\begin{tikzcd}
S \arrow[rr, "f"] \arrow[dr, "\pi"'] & & S' \\
& \overline{S} \arrow[ur, dashed, "\overline{f}"'] &
\end{tikzcd}$$

That is, $f = \overline{f} \circ \pi$.

Choosing $\overline{f}$ is trivial. We just define $\overline{f}(\overline{x}) = f(x)$. This works, of course, in the composition, but it is not immediate that this map is even well-defined since x is but one representative for the equivalence class $\overline{x}$. Well-defined means that if $\overline{x} = \overline{y}$, then $\overline{f}(\overline{x}) = \overline{f}(\overline{y})$.

Suppose $\overline{y} = \overline{x}$. This means that $y \in \overline{x}$ (and also that $x \in \overline{y}$) and further that $f(x) = f(y)$. But then $\overline{f}(\overline{x}) = f(x) = f(y) = \overline{f}(\overline{y})$. Hence, $\overline{f}$ is well-defined.

The only thing left to check is uniqueness. Suppose $g : \overline{S} \to S'$ is another map such that $f = g \circ \pi$. Then for any $x \in S$, $g(\pi(x)) = f(x)$, so $g(\overline{x}) = f(x)$. Thus, $\overline{f}$ and g agree on their common domain, $\overline{S}$, and so $\overline{f} = g$. Thus, $\overline{f}$ is unique.

8. Cosets

Let H be a subgroup of a group G and $a \in G$. Recall that a left coset of H in G is

$$aH = \{ah : h \in H\}.$$

The cosets of H in G are the equivalence classes for the congruence relation

$$a \equiv b \text{ if } b = ah \text{ for some } h \in H.$$

Exercise. Verify that congruence is an equivalence relation.

Corollary 19. The left cosets of a subgroup H in a group G partition the group..

Example. Let $H = \langle y \rangle$ in $\mathcal{S}_3$. The cosets of H in $\mathcal{S}_3$ are

$$H = \{1, y\} = yH, \quad xH = \{x, xy\} = xyH, \quad x^2H = \{x^2, x^2y\} = x^2yH.$$

The next proposition is essentially a restatement of an earlier result.

Proposition 20. Let H be a subgroup of a group G and let $a, b \in G$. The following are equivalent.

(1) $b = ah$ for some $h \in H$,
(2) $a^{-1}b \in H$,
(3) $b \in aH$,
(4) $aH = bH$.

The number of left cosets of a subgroup H in a group G is called the index of H in G and is denoted $[G : H]$. In the previous example, $[\mathcal{S}_3 : H] = 3$.

Lemma 21. All left cosets aH of a subgroup H of a group G have the same order.

Proof. Left multiplication by a defines a map $H \to aH$ defined by $h \mapsto ah$. This map is bijective because its inverse is multiplication by a^{-1}. Hence, $|aH| = |H|$. □

The cosets all have the same order and they partition the group. From this we obtain the Counting Formula

$$|G| = |H|[G : H]$$
$$\text{(Order of G)} = \text{(Order of H)(number of cosets)}.$$

We now obtain the next two statements as easy corollaries of this formula.

Theorem 22 (Lagrange's Theorem)**.** Let H be a subgroup of a finite group G. The order of H divides the order of G.

Corollary 23. The order of an element of a finite group divides the order of the group.

Proof. Let $a \in G$, then $|a| = \langle a \rangle$. Apply Lagrange's Theorem with $H = \langle a \rangle$. □

Corollary 24. Let G be a group with $|G| = p$, p prime. Then G is cyclic and any $a \in G$, $a \neq e$, is a generator.

Proof. Let $a \in G$, $a \neq e$. Then $|a| \mid |G|$ by the previous corollary, so $|a| = 1$ or p. But $g \neq e$ so $|\langle a \rangle| > 1$. Hence $|\langle a \rangle| = p$ and so $\langle a \rangle = G$. □

From the above results and Proposition 18 we obtain

$$[G : \ker \phi] = |\operatorname{im} \phi|.$$

Proposition 25 (Multiplicative Property of Indices)**.** Let H, K be subgroups of a finite group G such that $K \subset H \subset G$. Then $[G : K] = [G : H][H : K]$.

Proof. If G is finite, then by Lagrange's Theorem,

$$[G : K] = \frac{|G|}{|K|} = \frac{|G|}{|H|} \cdot \frac{|H|}{|K|} = [G : H][H : K].$$

Assume now that $[G : H] = m < \infty$ and $[H : K] = n < \infty$. List the m distinct cosets (by choosing representatives) of H in G: $g_1H, \dots, g_mH$. Note that these cosets partition G. Similarly, the n distinct cosets of K in H partition H: $h_1K, \dots, h_nK$. The operation of left multiplication by $g_i \in G$ is an invertible operation (its inverse is multiplication by g_i^{-1}) and thus $g_iH = g_ih_1K \cup \cdots \cup g_ih_nK$. It follows that these cosets partition g_iH. Hence, G is partitioned into the mn cosets g_ih_jK.

A similar argument applies when one of $[G : H]$ or $[H : K]$ is infinite. □

We have done everything here in terms of left cosets, but of course all can be done using right cosets Ha in an analogous way. The structure of left and right cosets are related.

Proposition 26. Let H be a subgroup of a group G. The following conditions are equivalent:

(1) H is a normal subgroup: For all $h \in H$ and all $g \in G$, $ghg^{-1} \in H$.
(2) For all $g \in G$, $gHg^{-1} = H$.
(3) For all $g \in G$, $gH = Hg$.
(4) Every left coset of H in G is a right coset.

Proof. (1) $\Rightarrow$ (2). By hypothesis, $gHg^{-1} \subset H$ for all $g \in G$. Substituting g^{-1} for g gives $g^{-1}Hg \subset H$. Left and right multiplication give $H \subset gHg^{-1}$ and so the two sets are equal.

(2) $\Rightarrow$ (1). This is immediate.

(2) $\Leftrightarrow$ (3). This is clear as we can right multiply $gHg^{-1} = H$ to get $gH = Hg$ and similarly for the converse.

(3) $\Rightarrow$ (4). This is immediate.

(4) ⇒ (3). By hypothesis, every left coset is a right coset. Hence, the partitioning of G into cosets is the same. Let $g \in G$, then gH and Hg have an element in common, namely g. As two sets in a partition are either equal or disjoint, we have $gH = Hg$.

□

Proposition 27. (1) If H is a subgroup of a group G and g is an element of G, the set gHg^{-1} is also a subgroup.
(2) If a group G has just one subgroup H of order r, then that subgroup is normal.

Proof. The first statement follows because conjugation by $g \in G$ is an automorphism of H with image gHg^{-1}. Now gHg^{-1} is a subgroup with the same order as H. Thus, $gHg^{-1} = H$ so H is normal by the previous proposition. □

9. Modular Arithmetic

Fix a positive integer n. Two integers a and b are said to be congruent mod n if n divdies $b - a$ (equivalently, $b = a + nk$ for some $k \in \mathbb{Z}$). We write

$$a \equiv b \mod n.$$

The equivalence class of $a \in \mathbb{Z}$ is

$$\overline{a} = \{\dots, a - 2n, a - n, a, a + n, a + 2n, \dots\}.$$

In particular, the equivalence class of 0 is

$$\overline{0} = \{\dots, -2n, -n, 0, n, 2n, \dots\} = \mathbb{Z}n\}$$

and this is a subgroup of $\mathbb{Z}^+$. Let $H = \mathbb{Z}n$. In general, the equivalence class $\overline{a}$ is a right coset of $\mathbb{Z}^+$ written, additively,

$$\overline{a} = a + H = \{a + kn : k \in \mathbb{Z}\}.$$

The next result is now implied by earlier work.

Proposition 28. There are n congruence classes modulo n, namely $\overline{0}, \overline{1}, \dots, \overline{n-1}$. The index $[\mathbb{Z} : \mathbb{Z}n]$ of the subgroup $\mathbb{Z}n$ in $\mathbb{Z}$ is n.

Let $\overline{a}$ and $\overline{b}$ be the equivalence classes of $a, b \in \mathbb{Z}$ under congruence mod n. Then we have the arithmetic rules

$$\overline{a} + \overline{b} = \overline{a + b} \quad \text{and} \quad \overline{a}\overline{b} = \overline{ab}.$$

This follows from the next proposition.

Proposition 29. If $a' \equiv a$ and $b' \equiv b \mod n$, then $a' + b' \equiv a + b \mod n$ and $a'b' \equiv ab \mod n$.

We will typically denote the set of congruence classes mod n as $\mathbb{Z}/\mathbb{Z}n$. So, to say $a = b$ in $\mathbb{Z}/\mathbb{Z}n$ means that $a \equiv b \mod n$. Under addition, $\mathbb{Z}/\mathbb{Z}n$ is a group. Because it contains 0, it is not a group under multiplication. However, one can form a group when n is prime by removing 0.

10. The Correspondence Theorem

The theorem we study in this section has significant implications in the study of *quotient groups*. We will not define that term here, but a good example is $\mathbb{Z}/\mathbb{Z}n$ from the previous section.

Let $\phi : G \to \mathcal{G}$ be a group homomorphism, and let H be a subgroup of G. The restriction of ϕ to H is the homomorphism

$$\phi\big|_H : H \to \mathcal{G}$$

defined by restricting the domain to H. It is not difficult to check that this is a homomorphism (since H is closed under multiplication). Moreover,

$$\ker(\phi\big|_H) = (\ker\phi) \cap H \quad \text{and} \quad \operatorname{im}\phi\big|_H = \phi(H).$$

By the Counting Formula, the order of $|\phi(H)|$ divides $|H|$ and $|\mathcal{G}|$. Hence, if $|H|$ and $|\mathcal{G}|$ have no common factors, then $\phi(H) = \{1\}$ and so $H \subset \ker\phi$.

Example. Let H be a subgroup if $\mathcal{S}_n$ of odd order. The image of the sign homomorphism $\sigma : \mathcal{S}_n \to \{\pm 1\}$ has order 2, and so $H \subset \ker\sigma = A_n$. This says that a cycle of odd length is an even permutation.

The next proposition is just a restatement of Proposition 15

Proposition 30. Let $\phi : G \to \mathcal{G}$ be a homomorphism with kernel K, let $\mathcal{H}$ be a subgroup of G, and let $H = \phi^{-1}(\mathcal{H})$.

(1) H is a subgroup of G that contains K.
(2) If $\mathcal{H}$ is a normal subgroup of $\mathcal{G}$, then H is a normal subgroup of G.
(3) If ϕ is surjective and if H is a normal subgroup of G, then $\mathcal{H}$ is a normal subgroup of G.

Theorem 31 (Correspondence Theorem)**.** Let $\phi : G \to \mathcal{G}$ be a surjective group homomorphism with kernel K. There is a bijective correspondence between subgroups of $\mathcal{G}$ and subgroups of G that contain K.

The correspondence itself is fairly direct. If H is a subgroup of G that contains K, then it is matched to its image $\phi(H)$ in $\mathcal{G}$. Conversely, if $\mathcal{H}$ is a subgroup of $\mathcal{G}$, then it is matched to its inverse image $\phi^{-1}(\mathcal{H})$ in G.

Here are two consequences of this theorem. Let H and $\mathcal{H}$ are corresponding subgroups. Then

- H is normal in G if and only if $\mathcal{H}$ is normal in $\mathcal{G}$.
- $|H| = |\mathcal{H}||K|$.

11. Product Group

Given two groups, G and G', one can form the cartesian product

$$G \times G' = \{(a, a') : a \in G, a' \in G'\}.$$

This is a group under the operation,

$$(a, a') \cdot (b, b') = (ab, a'b')$$

where the product in the first coordinate is computed in G and the product in the second coordinate is computed in G'. The pair $(1, 1)$ is the identity and (a^{-1}, a'^{-1}) is the inverse of (a, a').

These groups are all related by a set of homomorphisms diagrammed below:

$$\begin{tikzcd}
G \arrow[dr, "i"] & & G \\
& G \times G' \arrow[ur, "p"] \arrow[dr, "p'"] & \\
G' \arrow[ur, "i'"] & & G'
\end{tikzcd}$$

These maps are defined by the rules

$$i(x) = (x, 1), \quad i'(x) = (1, x), \quad p(x, y) = x, \quad p'(x, y) = y.$$

The surjective maps p, p' are called projections with kernels $1 \times G'$ and $G \times 1$, respectively. The next question will be when we can decompose a group into a direct product of two of its subgroups.

Proposition 32. Let r and s be relatively prime integers A cyclic group of order rs is isomorphic to the product of a cyclic group of order r and a cyclic group of order s.

The next proposition does this more generally.

Proposition 33. Let H and K be subgroups of a group G, and let $f : H \times K \to G$ be the multiplication map, defined by $f(h, k) = hk$. Its image is the set $HK = \{hk : h \in H, k \in K\}$.

(1) f is injective if and only if $H \cap K = \{1\}$.
(2) f is a homomorphism from the product group $H \times K$ to G if and only if elements of K commute with element of H: $hk = kh$.
(3) If H is a normal subgroup of G, then HK is a subgroup of G.
(4) f is an isomorphism from the product group $H \times K$ to G if and only if $H \cap K = \{1\}$, $HK = G$, and also H and K are normal subgroups of G.

One can use the above to show that every group of order 4 is isomorphic to either C_4 or $C_2 \times C_2$.

12. Quotient Groups

Subgroups only tell part of the story of the structure of a given group. The other part is determined by its quotients. We have already seen this with congruence mod n and the additive structure of cosets.

Let G be a group and N a *normal* subgroup. We denote the set of cosets of N in G by G/N and simplify this to $\overline{G}$ when the group N is clear.

There is a law of composition on $\overline{G}$ defined by

$$(aN)(bN) = (ab)N,$$

for all $aN, bN \in \overline{G}$. It is left as an exercise to verify that this operation is well-defined (independent of coset representative) and associative. The coset with representative 1 is denoted just N and under this operation, N is the identity in $\overline{G}$. Moreover, the coset $a^{-1}N$ is the inverse of aN in $\overline{G}$. Hence, $\overline{G}$ is a group under the given operation.

Theorem 34. Let N be a normal subgroup of a group G, and let $\overline{G}$ denote the set of cosets of N in G. There is a law of composition on $\overline{G}$ that makes this set into a group, such that the map $\pi : G \to \overline{G}$ defined by $\pi(a) = \overline{a}$ is a surjective homomorphim whose kernel is N.

It is not difficult to verify the last claim in the theorem. We call the map π the canonical map from G to $\overline{G}$. This now leads to one of the most important theorems of elementary abstract algebra.

Theorem 35 (First Isomorphism Theorem)**.** Let $\phi : G \to G'$ be a surjective group homomorphism with kernel N, and let $\overline{G} = G/N$. Let $\pi : G \to \overline{G}$ be the canoncial map. There is a unique isomorphism $\overline{\phi} : G \to G'$ such that $\phi = \overline{\phi} \circ \pi$.

$$\begin{tikzcd}
G \arrow[rr, "\phi"] \arrow[dr, "\pi"'] & & G' \\
& \overline{G} \arrow[ur, dashed, "\overline{\phi}"'] &
\end{tikzcd}$$

Another, less precise way to say the above is that the quotient group $\overline{G}$ is isomorphic to G'.

Proof. The elements of $\overline{G}$ are the cosets of N, and they are also the fibres of the map ϕ (Proposition 18). The map $\overline{\phi}$ sends a nonempty fibre to its image: $\overline{\phi}(x) = \phi(x)$. For any surjective map of sets $\phi : G \to G'$, one can form the $\overline{G}$ of fibres, and then one obtains the a diagram as above, in which $\overline{\phi}$ is the bijective map that sends a fibre to its image (see the A Digression on Fibres above). When ϕ is a group homomorphism, we have for all $a, b \in \overline{G}$,

$$\overline{\phi}(ab) = \phi(ab) = \phi(a)\phi(b) = \overline{\phi}(a)\overline{\phi}(b),$$

whence $\overline{\phi}$ is an isomorphism. □

More Group Theory

(Last Updated: September 28, 2018)

These notes are derived primarily from *Algebra* by Michael Artin (2ed). Most of this material is drawn from Chapter 7.

0. Group operations

Disclaimer/Warning: This material in this section is primarily drawn from Chapter 6.

The set of automorphisms of an algebraic object X forms a group (under composition of maps). Thus, the automorphisms can be though of as a symmetry of X. That is, the automorphism permutes the objects of X whilst maintaining the algebraic structure of X. The most familiar example to most students will be the action of the symmetric group $\mathcal{S}_n$ on the set of n letters. This is an example of the concept of a group operation.

Definition. An operation (or action) of a group G on a set S is a map

$$\begin{aligned} G \times S &\to S \\ (g, s) &\mapsto g * s \end{aligned}$$

satisfying

(1) $1 * s = s$ for all $s \in S$, and
(2) $(gg') * s = g * (g' * s)$ for all $g, g' \in G$ and $s \in S$ (associative law).

We typically omit the $*$ (other authors use .). When G acts on S, then each $g \in G$ acts as a map $S \to S$. We denote this map by m_g. That is, $m_g : S \to S$ defined by the rule $m_g(s) = gs$. Each m_g is a permutation of S because it has inverse $m_{g^{-1}}$.

Example. Choose coordinates so as to identify the plane P with the space $\mathbb{R}^2$. An isometry of $\mathbb{R}^2$ is a map $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that, for all $u, v \in \mathbb{R}^2$,

$$|f(u) - f(v)| = |u - v|.$$

That is, f is a distance preserving map.

The isometries of $\mathbb{R}^2$ form a group that acts on $\mathbb{R}^2$. We list three families of isometries below.

(1) translation t_a by a vector a: $t_a(x) = x + a = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$.

(2) rotation ρ_θ by an angle θ about the origin: $\rho_\theta(x) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

(3) reflection r about the e_1-axis: $r(x) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

A remarkable theorem (6.3.4 in the text) shows that every isometry of the plane is either a translation, rotation, reflection (about some line ℓ), or a glide reflection (reflection about a line ℓ followed by translation by a nonzero vector parallel to ℓ).

Given an operation of a group G on a set S, one can form an equivalence relation on S by declaring

$$s \sim s' \text{ if } s' = gs \text{ for some } g \in G.$$

An equivalence class of this relation is called an orbit. That is, the orbit of $s \in S$ is defined as

$$O_s = \{s' \in S : s' = gs \text{ for some } g \in G\}.$$

When S consists of just one orbit, the operation is said to be transitive.

The stabilizer of an element $s \in S$ is the set of elements that leave s fixed, so

$$G_s = \{g \in G : gs = s\}.$$

For each $s \in S$, G_s is a subgroup of G.

Proposition 1. Let S be a set on which a group G operates, let $s \in S$, and let $H = G_s$.

(1) If $a, b \in G$, then $as = bs$ if and only if $a^{-1}b \in H$ if and only if $b \in aH$.
(2) Suppose that $as = s'$. Then $H' = G_{s'}$ is a conjugate subgroup:

$$H' = aHa^{-1} = \{g \in G : g = aha^{-1} \text{ for some } h \in H\}.$$

Proof. (1) We need only observe that $as = bs$ if and only if $s = a^{-1}bs$, and so $a^{-1}b \in G_s$.

(2) Let $g \in aHa^{-1}$, so $g = aha^{-1}$ for some $h \in H$. Then $gs' = (aha^{-1})(as) = ahs = as = s'$. Thus, $aHa^{-1} \subset H'$. Since $s = a^{-1}s'$, then the above argument shows that $a^{-1}H'a \subset H$ so $H' \subset aHa^{-1}$, as desired. □

Let H be a subgroup of G and G/H the set of left cosets of H in G. Of course, G/H is only a group if H is normal, but regardless, G acts on H in the following way. We denote by $[C]$ the coset C as an element of G/H. Then the action of $g \in G$ on $[C]$ is given by $g[C] = [gC]$. Thus, if $[C] = [aH]$, then $[gC] = [gaH]$. It is left to the reader to check that this is a well-defined operation.

Example. Let $G = \mathcal{S}_3$ and let $H = \{1, y\}$. The left cosets of H in G are

$$C_1 = H = \{1, y\}, \qquad C_2 = xH = \{x, xy\}, \qquad C_3 = x^2H = \{x^2, x^2y\}.$$

Thus, $G/H = \{[C_1], [C_2], [C_3]\}$. Now x and y act on the set G/H as elements of $\mathcal{S}_3$ on $\{1, 2, 3\}$ via the association

$$m_x \leftrightarrow (1\ 2\ 3) \qquad m_y \leftrightarrow (2\ 3).$$

For example, $yC_2 = C_3$ (check this!).

Proposition 2. Let H be a subgroup of a group G.

(1) The operation of G on the set G/H of cosets is transitive.
(2) The stabilizer of the coset H is the subgroup H.

Proposition 3. Let S be a set on which a group G operates, and let $s \in S$. Let H and O_s be the stabilizer and orbit of s, respectively. There is a bijective map $\epsilon : G/H \to O_s$ defined by $[aH] \mapsto as$. This map is compatible with the operations of the group: $\epsilon(g[C]) = g\epsilon([C])$ for every coset C and every element $g \in G$.

Proof. First we claim that the map defined in the proposition is actually well-defined. Suppose $a, b \in G$ such that $aH = bH$. We claim that $as = bs$. Since $a^{-1}b \in H$ by basic properties of cosets and H is the stabilizer of s, then $a^{-1}bs = s$, so $as = bs$ as claimed.

Now that ϵ is well-defined, we claim that it is bijective. Essentially using the same argument as above (but in reverse), one sees that ϵ is injective. Now let $gs \in O_s$, then $\epsilon([gH]) = gs$, so ϵ is surjective.

The last claim (compatibility) is clear. □

The next result is sometimes called the Orbit-Stablizer Theorem (though apparently not by Artin). It is essentially a restatement of the Counting Formula.

Proposition 4. Let S be a finite set on which a group G operates, and let G_s and O_s be the stabilizer and orbit, respectively, of an element $s \in S$. Then

$$|G| = |G_s||O_s|$$
$$\text{(order of G)} = \text{(order of stablizer)} \cdot \text{(order of orbit)}.$$

A permutation representation of a group G is a homomorphism from the group to a symmetric group:

$$\phi : G \to \mathcal{S}_n.$$

Proposition 5. Let G be a group. There is a bijective correspondence between operations of G on the set $S = \{1, \ldots, n\}$ and permutation representations $G \to \mathcal{S}_n$.

Proof. The map $\phi : G \to \mathcal{S}_n$ was essentially given above. Set $\phi(g) = m_g$ and recall that each m_g defines a permutation of S. To show that ϕ is a homomorphism, let $g, h \in G$, then for some $i \in S$ we have

$$m_g(m_h i) = g(hi) = (gh)i = m_{gh} i.$$

Hence, $\phi(g)\phi(h) = \phi(gh)$. □

There's nothing special about the set S above. For an arbitrary set, let $\mathrm{Perm}(S)$ denote the group of permutations of S. We also call a homomorphism $\phi : G \to \mathrm{Perm}(S)$ a permutation representation of G.

Corollary 6. Let $\mathrm{Perm}(S)$ denote the group of permutations of S. There is a bijective correspondence between operations of G on S and permutation representations $G \to \mathrm{Perm}(S)$.

If a permutation representation $\phi : G \to \mathrm{Perm}(S)$ is injective, then the corresponding operation of G on S is said to be faithful. This is equivalent to saying that the only element $g \in G$ such that $gs = s$ for all $s \in S$ is the identity.

A digression on the dihedral group

Recall our list of isometries of the plane.

(1) translation t_a by a vector a: $t_a(x) = x + a = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$.

(2) rotation ρ_θ by an angle θ about the origin: $\rho_\theta(x) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

(3) reflection r about the e_1-axis: $r(x) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

The isometries that fix the origin (in P) are the orthogonal linear operators, so when coordinates are chosen, the orthogonal group O_2 becomes a subgroup of isometries M. The next theorem describes such subgroups.

Theorem 7. Let G be a finite subgroup of the orthogonal group O_2. There is an integer n such that G is one of the following groups:

(1) C_n: the *cyclic group* of order n generated by the rotation ρ_θ, where $\theta = 2\pi/n$.

(2) D_n: the *dihedral group* of order $2n$ generated by two elements: the rotation ρ_θ, where $\theta = 2\pi/n$, and a reflection r' about a line ℓ through the origin.

Proposition 8. The dihedral group D_n has order $2n$. It is generated by two elements x and h that satisfy the relations

$$x^n = 1, \qquad y^2 = 1, \qquad yx = x^{-1}y.$$

From this presentation it is not difficult to see that $D_3 \cong \mathcal{S}_3$.

There is an operation of D_n on the vertices $\{v_1, \ldots, v_n\}$ of a regular n-gon. This operation defines a permutation representation $\phi : D_n \to \mathcal{S}_n$.

1. Cayley's Theorem

Some of the most important group operations are those operations on the group itself. In this chapter we present several of these and this leads to the Sylow Theorems, which are fundamental tools in the classification of finite groups.

A group acts on itself by left multiplication:

$$\begin{aligned} G \times G &\to G \\ (g, x) &\mapsto gx. \end{aligned}$$

As there is only one orbit, this operation is transitive.

Theorem 9 (Cayley's Theorem)**.** Every finite group is isomorphic to a subgroup of a permutation group. A group of order N is isomorphic to a subgroup of the symmetric group $\mathcal{S}_n$.

Proof. Since the operation of left multiplication is faithful (the stabilizer of any element is the trivial subgroup $\langle 1 \rangle$), the permutation representation

$$\begin{aligned} G &\to \mathrm{Perm}(G) \\ g &\mapsto m_g \end{aligned}$$

is injective. Thus, G is isomorphic to its image in $\mathrm{Perm}(G)$. If $|G| = n$, then $\mathrm{Perm}(G)$ is isomorphic to $\mathcal{S}_n$. □

2. The Class Equation

Another way G operates on itself is by conjugation:

$$\begin{aligned} G \times G &\to G \\ (g, x) &\mapsto gxg^{-1}. \end{aligned}$$

As opposed to left multiplication, it is less obvious that this is actually an operation. Clearly, $1 * x = 1x1^{-1} = x$. Now let $g, h \in G$, then

$$g * (h * x) = g * (hxh^{-1}) = g(hxh^{-1})g^{-1} = (gh)x(gh)^{-1} = (gh) * x.$$

Thus, the operation is associative.

The stabilizer of an element $x \in G$ for the operation of conjugation is called the centralizer of x, denoted $Z(x)$:

$$Z(x) = \{g \in G : gxg^{-1} = x\} = \{g \in G : gx = xg\}.$$

Said another way, the centralizer is the set of elements that commute with x.

The orbit of x for conjugation is called the conjugacy class of x, denoted $C(x)$:

$$C(x) = \{y \in G : y = gxg^{-1} \text{ for some } g \in G\}.$$

By the Counting Formula,

$$|G| = |Z(x)| \cdot |C(x)|.$$

Proposition 10. (1) The centralizer $Z(x)$ of an element $x \in G$ contains x, and it contains the center Z of G.

(2) An element $x \in G$ is in the center if and only if its centralizer $Z(x)$ is the whole group G, and this happens if and only if the conjugacy class $C(X)$ consists of the element x alone.

The conjugacy classes are orbits for a group operation and, since the operation is on the group itself, they partition the group. Thus, if G is finite and $C_1, \ldots, C_k$ are the distinct conjugacy classes of G, then

$$|G| = |C_1| + \cdots + |C_k|.$$

Note that $C(1) = \{1\}$, and so we denote this conjugacy class by C_1 (thus, $|C_1| = 1$). Also note that any conjugacy class of a central element (element of Z) is a singleton. That is, $|C(x)| = 1$ if and only if $x \in Z$. Also, because each conjugacy class is an orbit, it divides the order of G (by the Counting Formula). To summarize: the number on the right side of the class equation divide the order of the group, and at least one of them is equal to 1.

Example. Let $G = \mathcal{S}_3$, then $|G| = 6$. The centralizer of x $(= (1\ 2\ 3))$ contains x (in fact, it contains the subgroup $\langle x \rangle$). Hence, $|Z(x)| = 3$ or 6 (since its order must divide the order of the group - Counting Formula). But $y \notin Z(x)$, so $|Z(x)| = 3$. Hence, $C(x) = 2$ (Counting Formula, again). Similar logic shows that $|C(y)| = 3$. Thus, the class equation for $\mathcal{S}_3$ is

$$6 = 1 + 2 + 3.$$

3. p-Groups

A group whose order is a positive power of a prime p is said to be a p-group.

Proposition 11. The center of a p-group is not the trivial group.

Proof. Say that $|G| = p^e$ for some $e \geq 1$. By the class equation,

$$|G| = |C_1| + \cdots + |C_k| \Rightarrow p^e = 1 + |C_2| + \cdots + |C_k|.$$

If $|Z| = 1$, then for all $i > 1$, $|C_i| = p^k$ for some $k \geq 1$. But then p divides $|C_2| + \cdots + |C_k|$ but not $p^e - 1$, a contradiction. □

The proof of the next theorem is similar.

Theorem 12 (Fixed Point Theorem)**.** Let G be a p-group, and let S be a finite set on which G operates. If the order of S is not divisible by p, there is a fixed point for the operation of G on S - an element s whose stabilizer is the whole group.

Proposition 13. Every group of order p^2 is abelian.

Proof. By the Proposition 11 and Lagrange's Theorem, $|Z| = p$ or p^2. If $|Z| = p^2$, then $Z = G$ and we are done.

Suppose $|Z| = p$ and let $x \in G \backslash Z$. However, $Z(x)$ contains Z and contains x, so $|Z(x)| > Z$. But $|Z(x)|$ divides $|G| = p^2$, so $|Z(x)| = p^2$, implying $x \in Z$, a contradiction. □

Corollary 14. A group of order p^2 is either cyclic, or the product of two cyclic groups of order p.

Proof. Let G be a group of order p^2. If G contains an element of p^2, then it is cyclic. Assume otherwise. That is, every element of G that is not the identity has order p. Let $x \in G$ be a non-identity element and choose an element y not in $\langle x \rangle$. Then $\langle y \rangle \cap \langle x \rangle = \{1\}$ and so $G = \langle x \rangle \langle y \rangle \cong \langle x \rangle \times \langle y \rangle$. □

6. Normalizers

Just as G acts on itself by conjugation, so too does it act on the set of subgroups of G. Let H be a subgroup of G. The orbit $[H]$ of H under conjugation is the set of subgroups gHg^{-1} for $g \in G$ (recall that gHg^{-1} is always a subgroup of G). The stablizer of $[H]$ for this operation is called the normalizer of H, denoted $N(H)$:

$$N(H) = \{g \in G : gHg^{-1} = H\}.$$

By the Counting Formula:

$$|G| = |N(H)| \cdot (\text{number of conjugate subgroups}).$$

Hence, since $N(H)$ is a subgroup of G, then the number of conjugate subgroups is equal to $[G : N(H)]$.

Proposition 15. Let H be a subgroup of a group G, and let N be the normalizer of H.

(1) H is a normal subgroup of N.
(2) H is a normal subgroup of G if and only if $N = G$.
(3) $|H|$ divides $|N|$ and $|N|$ divides $|G|$.

7. The Sylow Theorems

Let G be a group of order n, an let p be a prime integer that divies n. Let p^e denote the largest power of p that divides n, so that $n = p^e m$ where m is an integer not divisible by p Subgroups H of G of order p^e are called Sylow p-subgroups of G. Said another way, a Sylow p-subgroup is a p-group whose index is the group is not divisible by p.

We will begin by stating the three Sylow theorems, along with some of their consequences, and some applications. The proofs of the Sylow theorems is deferred to the end.

Theorem 16 (First Sylow Theorem)**.** A finite group whose order is divisible by a prime p contains a Sylow p-subgroup.

Corollary 17. A finite group whose order is divisible by a prime p contains an element of order p.

Proof. Let G be such a group, and let H be a Sylow p-subgroup of G. Then H contains a non-identity element x. By Lagrange's Theorem, $|x|$ divides $|H|$, so $|x| = p^k$ for some $k \geq 1$. Then $x^{p^{k-1}}$ has order p. □

Theorem 18 (Second Sylow Theorem)**.** Let G be a finite group whose order is divisible by a prime p.

(1) The Sylow p-subgroups of G are conjugate subgroups.
(2) Every subgroup of G that is a p-group is contained in a Sylow p-subgroup.

Corollary 19. A group G has just one Sylow p-subgroup H if and only if that subgroup is normal.

Proof. Suppose H is the only Sylow p-subgroup. Then for any $g \in G$, gHg^{-1} is another Sylow p-subgroup by the Second Sylow Theorem. This implies $gHg^{-1} = H$, so H is normal.

Conversely, suppose H is a normal Sylow p-subgroup. Then it is conjugate to all other Sylow p-subgroups. But by normality, $gHg^{-1} = H$ for all $g \in G$, so H is the only Sylow p-subgroup. □

Theorem 20 (Third Sylow Theorem)**.** Let G be a finite group whose order n is divisible by a prime p. Say $n = p^e m$, where p does not divide m, and let s denote the number of Sylow p-subgroups. Then s divides m and s is congruent to 1 modulo p: $s = kp + 1$ for some integer $k \geq 0$.

We'll now apply the Sylow theorems to classify groups of low order.

Example. Let G be a group of order 15. By the Third Sylow Theorem, the number of its Sylow 3-subgroups divides 5 and is congruent to 1 mod 3. Hence, this number is 1 and so there is a unique Sylow 3-subgroup, H, and it is normal. Similar logic shows that there is a unique Sylow 5-subgroup, K. The subgroups H and K are cyclic of orders 3 and 5, respectively, and $H \cap K = \{1\}$. Thus, $G \cong H \times K \cong C_3 \times C_5 \cong C_{15}$. Hence, all groups of order 15 are cyclic.

Example. Let G be a group of order 6. The logic as above shows that there is a unique Sylow 3-subgroup, H. On the other hand, the number of Sylow 2-subgroups is either 1 or 3. If there is a unique Sylow 2-subgroup, K, then the same argument as in the previous example shows that $G \cong C_6$.

Suppose there are 3 distinct Sylow 2-subgroups: K_1, K_2, K_3. The group G acts by conjugation on the set $S = \{[K_1], [K_2], [K_3]\}$ of order 3, and this gives a homomorphism $\phi : G \to \mathcal{S}_3$ (permutation representation). By the Second Sylow Theorem, this action is transitive, so the stabilizer in G of the element $[K_i]$, which is the normalizer $N(K_i)$, has order 2. Hence, $N(K_i) = K_i$. Since $K_1 \cap K_2 = \{1\}$, then 1 is the only element of G that fixes all elements of S. Thus, the operation is faithful and ϕ is injective. Since $|G| = |\mathcal{S}_3|$, then ϕ is an isomorphism.

Now we proceed to proving the Sylow Theorems themselves.

Lemma 21. Let U be a subset of a group G. The order of the stabilizer $\mathrm{Stab}([U])$ of $[U]$ for the operation of left multiplication by G on the set of its subsets divides both of the orders $|U|$ and $|G|$.

Proof. If H is a subgroup of G, the H-orbit of an element u of G for left multiplication by H is the right coset Hu. Let H be the stabilizer of $[U]$. Then multiplication by H permutes the elements of U, so U is partitione into H-orbits, which are right cosests. Each coset has order H, so $|H|$ divides $|U|$. Because H is a subgroup, $|H|$ divides $|G|$. □

Lemma 22. Let $n = p^e m$ be an integer with p prime, $e > 0$, and m an integer not divisible by p. The number N of subsets of order p^e in a set of order n is not divisible by p.

Proof. The number N is the binomial coefficient

$$\binom{n}{p^e} = \frac{n \cdot (n-1) \cdots (n-k) \cdots (n-p^e+1)}{p^e(p^e-1) \cdots (p^e-k) \cdots 1}.$$

If p divides $(n-k)$, then p divides $(p^e - k)$ the same number of times. It follows that $N \not\equiv 0 \mod p$. □

Proof of the First Sylow Theorem. Let G be a group of order $p^e m$, where p is a prime, $e > 0$, and p does not divide m. Let $\mathcal{S}$ be the set of all subsets of G of order p^e. We claim that one of the subsets in $\mathcal{S}$ is a subgroup of G.

Note that G acts on $\mathcal{S}$ by left multiplication. We decompose $\mathcal{S}$ into orbits for this operation, so

$$N = |\mathcal{S}| = \sum_{\text{orbits } O} |O|.$$

By Lemma 22, $p \nmid N$, so there is some subset $[U]$ whose orbit, $O_{[U]}$, has an order not divisible by p. Let $H = \mathrm{Stab}([U])$. By Lemma 21, $|H|$ divides $|U| = p^e$ (by definition of $\mathcal{S}$). Hence, $|H|$ is a power of p. By the Counting Formula, $|H| \cdot |O_{[U]}| = |G| = p^e m$. But, $|O_{[U]}|$ is not divisible by p, so $|H| = p^e$, whence H is a Sylow p-subgroup. □

Proof of the Second Sylow Theorem. Let G be a finite group whose order is divisible by a prime p, let H be a Sylow p-subgroup of G (note that H exists by the First Sylow Theorem), and let K be a p-subgroup of G. We will show that some conjugate H' of H contains K. Assume this is true. Then if K is a Sylow p-subgroup, then $|K| = |H'|$, proving (a).

Let $\mathcal{C}$ be a set on which G operates satisfying

- $p \nmid |\mathcal{C}|$,
- the operation is transitive,
- $\mathcal{C}$ contains an element c such that $\text{Stab}(c) = H$.

Since we could take $\mathcal{C}$ to be the set of left cosets of H in G, such a set exists.

We restrict the operation of G on $\mathcal{C}$ to the p-group K. Since $p \nmid |\mathcal{C}|$, there is a fixed point c' for the operation on K by the Fixed Point Theorem. The operation of G is transitive, and so $c' = gc$ for some $g \in G$. Hence, the stabilizer of c' is $H' = gHg^{-1}$. Since K fixes c', then $K \subset H'$. This proves (b). □

Proof of the Third Sylow Theorem. Let G be a group of order $p^e m$, where p is a prime, $e > 0$, and p does not divide m. Let s denote the number of Sylow p-subgroups. By the Second Sylow Theorem, the operation of G on the set S of Sylow p-subgroups (conjugation) is transitive. Let H be a Sylow p-subgroup. Then $\text{Stab}([H]) = N(H) = N$. By the Counting Formula, $s = |S| = [G : N]$. Since $H \subset N$, and since $[G : H] = m$, then s divides m.

Now we decompose S into orbits for the operation of conjugation by H. The H-orbit of $[H]$ has order 1. Since H is a p-group, the order of any H-orbit is a power of p. To prove $s \equiv 1 \mod p$, it suffices that to show that $[H]$ is the only element of S fixed by H.

Let H' be a p-Sylow subgroup such that H fixes $[H']$ under conjugation. Then $H \subset N' = N(H')$, so both H and H' are Sylow p-subgroups of N'. By the Second Sylow Theorem, the Sylow p-subgroups of N' are conjugates of N'. But H' is a normal subgroup of H', so $H' = H$. □

8. Groups of order 12

In this section we fully exploit the power of the Sylow Theorems to classify groups of order 12. However, this also shows the limit of the theorems as we have many cases to consider.

Theorem 23. There are five isomorphism classes of groups of order 12. They are represented by:

- the product of cyclic groups $C_4 \times C_3$,
- the product of cyclic groups $C_2 \times C_2 \times C_3$,

- the alternating group A_4,
- the dihedral group D_6,
- the group generated by elements x and y, with relations $x^4 = 1$, $y^3 = 1$, $xy = y^2x$.

Proof. Let G be a group of order 12. By the First Sylow Theorem, G has a Sylow 2-subgroup, H, which has order 4. Thus, H is isomorphic to either the cyclic group C_4 or the Klein-4 group $C_2 \times C_2$. By the Third Sylow Theorem, the number of Sylow 2-subgroups is 1 or 3. Similarly, G has a Sylow 3-subgroup $K \cong C_3$, and the number of Sylow 3-subgroups is 1 or 4.

The next part is not necessary for the proof but can be a useful tool in using the Sylow Theorems. Suppose K is not normal, so there are four distinct Sylow 3-subgroups: $K_1, \ldots, K_4$, with $K_1 = K$. As each K_i is cyclic of order 3, any non-identity element is a generator. It follows that $K_i \cap K_j = \{1\}$ when $i \neq j$. Hence, between the four K_i's there are 9 elements of G. Also note that $K_i \cap H = \{1\}$ by Lagrange's Theorem. Hence, there are only three non-identity elements that may belong to the collection of Sylow 2-subgroups. Since $|H| = 4$, this shows that H is the *only* Sylow 2-subgroup, whence H is normal. We have now shown that at least one of H or K is normal.

Case 1: H and K are both normal.

Since $H \cap K = \{1\}$, then the product map $H \times K \to G$ is bijective. Since H and K are both normal, then $G \cong H \times K$ so either

$$G \cong C_4 \times C_3 \quad \text{or} \quad G \cong C_2 \times C_2 \times C_3.$$

Case 2: K is not normal.

(It is irrelevant in this case whether H is normal and this is why the discussion above was unnecessary. However, some of the discussion will be useful here.)

As discussed above, there are four conjugate Sylow 3-subgroups $(K_1, \ldots, K_4)$ and G acts on the set of Sylow 3-subgroups by conjugation. This defines a permutation representation $\phi : G \to \mathcal{S}_4$. We will show that ϕ is injective and $\operatorname{im}\phi \cong A_4$.

The normalizer $N_i = N(K_i)$ contains K_i, and $|N_i| = 3$ by the Counting Formula. Thus, $N_i = K_i$. Because $K_i \cap K_j = \{1\}$ when $i \neq j$, only the identity stabilizes all of the subgroups. Thus, the operation of G is faithful, and so ϕ is injective and $G \cong \operatorname{im}\phi$.

Since G has four subgroups of order 3, it contains eight elements of order 3. Their images are the 3-cycles in S_4, which generates A_4 (exercise!). Hence, the image of G contains A_4. But $|A_4| = 12$, so $\operatorname{im}\phi \cong A_4$.

Case 3: K is normal, but H is not.

In this case, H operates by conjugation on $K = \{1, y, y^2\}$. Since H is not normal, it contains an element x that doesn't commute with y, whence $xyx^{-1} = y^2$. There are now two subcases depending on the isomorphism class of H.

Case 3a: H is a cyclic group.

The element x generates H, so G is generated by x and y with the relations

$$x^4 = 1, \quad y^3 = 1, \quad xy = y^2x.$$

These relations completely determine the multiplication table of G (check this!). Hence, there is at most one isomorphism class of such groups. One needs to show that the group does not collapse further. Said another way, that there are not additional relations amongst the generators. To do this, consider the matrices,

$$x = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad y = \begin{bmatrix} \omega & 0 \\ 0 & \omega^2 \end{bmatrix}, \omega = e^{2\pi i/3}.$$

These matrices satisfy the relations above and one can check that they generate a group of order 12.

Case 3b: $H \cong C_2 \times C_2$.

The stabilizer of y for the operation of H by conjugation on the set $\{y, y^2\}$ has order 2. Hence, H contains an element $z \neq 1$ such that $zy = yz$ and an element x such that $xy = y^2x$. Since H is abelian, $xz = zx$. Then G is generated by three elements x, y, z, with relations

$$x^2 = 1, y^3 = 1, z^2 = 1, yz = zy, xz = zx, xy = y^2x.$$

Again, these relations completely determine the multiplication table and so there is at most one isomorphism class of such groups. Since we are in the last case, this must be D_6. □

9. The free group

Let $S = \{a, b, \cdots\}$ be a set. We call the elements of S symbols or letters. A finite string of elements of S is a word and we let W denote the set of words in S. Concatenation defines an (associative) law of composition on W in which the *empty word*, denoted by 1, is the identity element. Thus, W is a monoid[1], called the free monoid on S. It is not a group and our goal will be to define a group structure derived from S.

Let $S' = \{a, a^{-1}, b, b^{-1}, \cdots\}$ and let W' be a free monoid on S'. If a word contains as a subword xx^{-1} or $x^{-1}x$ then we cancel the subword and reduce the length of the word. A word is reduced if there are no such cancellations that can be made. Given a word w, we denote by w_0 a reduced form of the word.

[1] Artin calls this a semigroup for reasons I don't understand. Generally a semigroup just has an associative law of composition and a monoid is a semigroup with an identity element.

Proposition 24. There is only one reduced form of a given word.

We call two words w and w' equivalent, and write $w \sim w'$, if they have the same reduced form. This defines an equivalence relation on W'. Let $\mathcal{F}$ denote the set of equivalence classes of words in W'. We call $\mathcal{F}$ the free group on the set S.

Proposition 25. The set $\mathcal{F}$ along with the law of composition induced from W' defines a group.

Proof. We need to show that the binary operation is well-defined. That is, suppose $w \sim w'$ and $v \sim v'$. We must show that $wv \sim w'v'$. Let w_0 and v_0 be the reduced forms of w and v, respectively. Then the reduced form of wv is the same as the reduced form of w_0v_0 (note this word may not be reduced) and the same holds for w' and v'.

That the operation is associative and (the equivalence class of) the empty word 1 being the identity are inherited from W'. Let $w = xy \cdots z$ be a word in W'. Then the inverse of w is $z^{-1} \cdots y^{-1}x^{-1}$ (here, by convention, we set $(a^{-1})^{-1} = a$). □

Example. The free group on a set $S = \{a\}$ is an infinite cyclic group.

10. Generators and relations

In this section we make more precise the terms generators and relations that we have used previously.

Definition. A relation R among elements $x_1, \ldots, x_n$ of a group G is a word r in the free group on the set $\{x_1, \ldots, x_n\}$ that evaluates to 1 in G.

Example. Let D_n be the dihedral group of symmetries of a regular n-gon. Let x be a rotation by $2\pi/n$ and y a reflection. These generators satisfy the relations

$$x^n = 1, \quad y^2 = 1, \quad xyxy = 1.$$

From these rules one can work out the entire multiplication table for D_n. We call these rules the defining relations of the group.

Lemma 26. Let R be a subset of a group G. There exists a unique smallest normal subgroup N of G that contains R, called the normal subgroup generated by R. If a normal subgroup of G contains R, it contains N. The elements of N can be described in either of the following ways:

(1) An element of G is in N if it can be obtained from the elements of R using a finite sequence of the operations of multiplication, inversion, and conjugation.

(2) Let R' be the set consisting of elements r and r^{-1} with $r \in R$. An element of G is in N if it can be written as a product $y_1, \ldots, y_r$ of some arbitrary length, where each y_v is a conjugate of an element of R'.

Proof. Let N be obtained in any one of the given ways. It is clear that N is normal since a normal subgroup is closed under those operations. Now we claim that N is the unique smallest subgroup containing R.

Let K be any normal subgroup of G containing R. As the intersection of two normal subgroups is again normal, then $K \cap N$ is normal and contains R. If $K \cap N$ is strictly contained in N, then we should be able to generate it by the operations given from R. But then $K \cap N = N$. □

By convention, we say that the empty set (word) generates the trivial group.

Definition. Let $\mathcal{F}$ be the free group on the set $S = \{x_1, \ldots, x_n\}$, and let $R = \{r_1, \ldots, r_k\}$ be a set of elements of $\mathcal{F}$. The group generated by S, with relations $r_1 = 1, \ldots, r_k = 1$, is the quotient group $\mathcal{G} = \mathcal{F}/\mathcal{R}$, where $\mathcal{R}$ is the normal subgroup of $\mathcal{F}$ generated by R. We denote $\mathcal{G}$ by $\langle x_1, \ldots, x_n : r_1, \ldots, r_k \rangle$.

We have the canonical homomorphism,

$$\pi : \mathcal{F} \to \mathcal{F}/\mathcal{R} = \mathcal{G}$$

where a word w is sent to the coset $\overline{w} = [wR]$, and the kernel of π is $\mathcal{R}$. We generally ignore the bar notation and just remember that $w_1, w_2 \in \mathcal{F}$ are equal in $\mathcal{G}$ if $w_1\mathcal{R} = w_2\mathcal{R}$.

Proposition 27 (Mapping property of the free group)**.** Let $\mathcal{F}$ be the free group on a set $S = \{a, b, \ldots\}$, and let G be a group. Any map of sets $f : S \to G$ extends in a unique way to a group homomorphism $\phi : \mathcal{F} \to G$. If we denote the image of $f(x)$ of $x \in S$ by $\underline{x}$, then ϕ sends a word $S' = \{a, a^{-1}, b, b^{-1}, \ldots\}$ to the corresponding product of the elements $\{\underline{a}, \underline{a}^{-1}, \underline{b}, \underline{b}^{-1}, \ldots\}$ in G.

Proof. We must show that ϕ, as defined, is a homomorphism. Note that, by definition, $\phi(a) = \underline{a}$ and $\phi(a^{-1}) = \underline{a}^{-1}$ for all $a \in S$.

Let $w \sim v$ in W', so $w = v$ in $\mathcal{F}$. By definition, w and v have the same reduced form and therefore after cancelling, $w_0 = v_0$. Note that the same reductions happen in G (cancelling inverses) and so ϕ is well-defined.

Let $w \in \mathcal{F}$, then $w = a_1 a_2 \cdots a_n$ for some $a_i \in S'$. We may assume that w is reduced. By the rule in the statement of the proof,

$$\phi(w) = \phi(a_1 a_2 \cdots a_n) = \underline{a_1 a_2 \cdots a_n} = \underline{a_1} \cdot \underline{a_2} \cdots \underline{a_n} = \phi(a_1)\phi(a_2) \cdots \phi(a_n).$$

Hence, ϕ is a homomorphism.

For uniqueness, suppose $\phi' : \mathcal{F} \to G$ is another homomorphism such that $\phi'(a) = f(a) = \underline{a}$ for every $a \in S$. If $w = a_1a_2\cdots a_n \in \mathcal{F}$ with $a_i \in S'$, then

$$\begin{aligned}\phi(w) = \phi(a_1a_2\cdots a_n) = \phi(a_1)\phi(a_2)\cdots\phi(a_n). &= \underline{a_1}\cdot\underline{a_2}\cdots\underline{a_n} = f(a_1)\cdot f(a_2)\cdots f(a_n)\\ &= \phi'(a_1)\cdot\phi'(a_2)\cdots\phi'(a_n) = \phi'(a_1a_2\cdots a_n) = \phi'(w).\end{aligned}$$

Hence, $\phi = \phi'$. □

The previous proposition encapsulates what we mean for G to be generated by S. That is, $G = \operatorname{im}\phi$.

Proposition 28 (Mapping property of the quotient groups)**.** Let $\phi : G' \to G$ be a group homomorphism with kernel K, and let N be a normal subgroup of G' that is contained in K. Let $\overline{G}' = G'/N$, and let $\pi : G' \to \overline{G}'$ be the canonical map $a \mapsto \overline{a}$. The rule $\overline{\phi}(\overline{a}) = \phi(a)$ defines a homomorphism $\overline{\phi} : \overline{G}' \to G$, and $\overline{\phi} \circ \pi = \phi$.

This generalizes the First Isomorphism Theorem, and we can represent it diagrammatically by

$$\begin{tikzcd}
G' \arrow[rr, "\phi"] \arrow[dr, "\pi"'] & & G \\
& \overline{G}' \arrow[ur, dashed, "\overline{\phi}"'] &
\end{tikzcd}$$

Corollary 29. Let $S = \{x_1, \ldots, x_n\}$ be a subset of a group G, $R = \{r_1, \ldots, r_k\}$ a set of relations among the elements of S, $\mathcal{F}$ the free group on S, $\mathcal{R}$ the normal subgroup of $\mathcal{F}$ generated by R, and $\mathcal{G} = \langle x_1, \ldots, x_n : r_1, \ldots, r_k\rangle$.

(1) There is a canonical homomorphism $\psi : \mathcal{G} \to G$ that sends $x_i \mapsto x_i$.

(2) ψ is surjective if and only if the set S generates G

(3) ψ is injective if and only if every relation among the elements of S is in $\mathcal{R}$.

Proof. (1) By the mapping property of the free group, there is a homomorphism $\phi : \mathcal{F} \to G$ with $\phi(x_i) = x_i$. The relations $r_i \in R$ evaluate to 1 in G and so $R \subset \ker\phi$. As $\ker\phi$ is a normal subgroup, $\mathcal{R} \subset \ker\phi$ as well. Now by the mapping property of quotient groups we have a map $\overline{\phi} : \mathcal{G} \to G$ and this is the map ψ.

(2) This is clear by the discussion above. That is, because $\psi \circ \pi = \phi$, then $G = \operatorname{im}\phi$ if and only if ψ is surjective.

(3) The map ψ is injective if and only if $\ker\psi = \{1\}$. By the mapping property of quotient groups, this is equivalent to $\ker\phi = \mathcal{R}$. □

Rings

(Last Updated: October 30, 2018)

These notes are derived primarily from *Algebra* by Michael Artin (2ed), though some material is drawn from *Abstract Algebra, Theory and Applications* by Thomas Judson (16ed).

1. Rings

Calling $\mathbb{Z}$ a group (under addition) obscures the fact that there are actually two well-defined (binary) operations on $\mathbb{Z}$: addition and multiplication. Moreover, these two operations play nicely together (via the distributive law).

Definition. A ring is a set R along with two laws of composition (typically $+$ and $\cdot$) satisfying the following

(1) $R^+ = (R, +)$ is an additive abelian group whose identity denoted 0.
(2) $(R, \cdot)$ is an abelian monoid whose identity is denoted 1.
(3) the distributive property holds: for all $a, b, c \in R$,

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c).$$

A subring of a ring is a subset that is closed under the operations of addition, subtraction, and multiplication and that contains the element 1.

Disclaimer/Warning: My biggest beef with Artin is that he assumes throughout that the multiplicative operation on a ring R is commutative. Most authors will *not* make this assumption and define a ring to be a *commutative ring* when this holds. Additionally, some authors will not require that R have a multiplicative identity (so $(R, \cdot)$ is only required to be a semigroup).

Example. The following are rings:

(1) $\mathbb{Z}, \mathbb{R}, \mathbb{Q}, \mathbb{C}$ under addition and multiplication.
(2) $\mathbb{Z}/(n) = \mathbb{Z}_n$ under addition and multiplication mod n.

Example. The Gaussian integers $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}$ are a subring of $\mathbb{C}$. In general, for any complex number $\alpha \in \mathbb{C}$, $\mathbb{Z}[\alpha]$ denotes the smallest subring of $\mathbb{C}$ that contains α.

A complex number $\alpha \in \mathbb{C}$ is algebraic if it is a root of a nonzero polynomial with integer coefficients. If no such polynomial exists, α is said to be transcendental.

Disclaimer/Warning: Because R^+ is assumed to be an (additive) abelian group, we denote the additive inverse of an element $a \in R$ by $(-a) = (-1)a$.

Proposition 1. Let R be a ring with $a, b \in R$. Then

(1) $a0 = 0a = 0$.
(2) $a(-b) = (-a)b = -(ab)$.
(3) $(-a)(-b) = ab$.

Proof. (1) We have $a0 = a(0+0) = a0 + a0$, so $a0 = 0$.

(2) By the (left) distributive property and (1), $ab + a(-b) = a(b-b) = a0 = 0$. Thus, $a(-b) = -(ab)$. Similarly $(-a)b = -(ab)$.

(3) By (2), $(-a)(-b) = -(a(-b)) = -(-(ab)) = ab$. □

Let R be a ring containing just one element. By definition, that element must be the additive identity 0. We call such a ring the zero ring.

Proposition 2. A ring R in which the elements 1 and 0 are equal is the zero ring.

Proof. Let $a \in R$. By Proposition 1, $0a = 0$ for all $a \in R$. If $1 = 0$, then $a = 1a = 0a = 0$. Hence, R is the zero ring. □

A unit of a ring is an element that has a multiplicative inverse. The units of $\mathbb{Z}$ are ± 1 and the units of $\mathbb{Z}[i]$ are $\pm 1, \pm i$. If every nonzero element in a (commutative) ring R has a multiplicative inverse, then R is said to be a field. Equivalently, $(R \backslash \{0\}, \cdot)$ is a group.

Example. Let p be a prime number. Then $\mathbb{Z}_p$ is a field, denoted $\mathbb{F}_p$.

Disclaimer/Warning: In general, if the multiplication operation is not assumed to be commutative, then we say that R is a division ring and a field is just a commutative division ring.

Proposition 3. Let R be a ring with multiplicative identity 1.

(1) The multiplicative identity is unique.
(2) If $a \in R$ is a unit, then a is not a zero divisor.
(3) If $a \in R$ is a unit, then its multiplicative inverse is unique.

Proof. (1) Exercise.

(2) Let a^{-1} be an inverse of a and suppose $ba = 0$. Then $0 = (ba)a^{-1}b(aa^{-1}) = b$. Similarly, $ab = 0$ implies $b = 0$.

(3) Suppose b, c are multiplicative inverses of a. Then $ba = 1 = ca$, so $(b-c)a = 0$. Thus, by (2), $b - c = 0$, or $b = c$. □

Example. Recall the quaternions are the group $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ under multiplication with identity element 1 satisfying $(-1)^2 = 1$, $i^2 = j^2 = k^2 = -1$, and

$$ij = k, \quad ji = -k, \quad jk = i, \quad kj = -i, \quad ki = j, \quad ik = -j.$$

Let $\mathbb{H} = \{a+bi+cj+dk : a,b,c,d \in \mathbb{R}\}$. That is, as a set, it is a real vector space with basis $\{1, i, j, k\}$ Define addition and multiplication on $\mathbb{H}$ as follows. Let $a_1+b_1i+c_1j+d_1k, a_2+b_2i+c_2j+d_2k \in \mathbb{H}$, then

$$\begin{aligned}(a_1 + b_1 i + c_1 j + d_1 k) + (a_2 + b_2 i + c_2 j + d_2 k) &= (a_1 + a_2) + (b_1 + b_2)i + (c_1 + c_2)j + (d_1 + d_2)k \\ (a_1 + b_1 i + c_1 j + d_1 k)(a_2 + b_2 i + c_2 j + d_2 k) &= \alpha + \beta i + \gamma j + \delta k\end{aligned}$$

where

$$\begin{aligned}\alpha &= a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2 \\ \beta &= a_1b_2 + a_2b_1 + c_1d_2 - d_1c_2 \\ \gamma &= a_1c_2 - b_1d_2 + c_1a_2 - d_1b_2 \\ \delta &= a_1d_2 + b_1c_2 - c_1b_2 - d_1a_2.\end{aligned}$$

Thus, $\mathbb{H}$ is a ring with identity. The multiplication operation is noncommutative (since $ij \neq ji$). One can now check that for $a + bi + cj + dk \neq 0$,

$$(a + bi + cj + dk)\left(\frac{a - bi - cj - dk}{a^2 + b^2 + c^2 + d^2}\right) = 1.$$

Thus, $\mathbb{H}$ is a (noncommutative) division ring.

2. Polynomials

Throughout this section, R is a (commutative) ring with identity

An expression of the form

$$f(x) = \sum_{i=0}^{n} a_i x^i$$

where $a_i \in R$ is called a polynomial over R in indeterminate x. The set of such polynomials is denoted $R[x]$.

The elements a_i are the coefficients of f. The degree of f is the largest m such that $0 \neq a_m$ if such an m exists. A polynomial of degree zero is called constant We write $\deg(f) = m$ and say a_m is the leading coefficient. Otherwise $f = 0$ and we set $\deg(f) = -\infty$. A nonzero polynomial with leading coefficient 1 is called monic.

Let $p(x), q(x) \in R[x]$ be nonzero polynomials with degrees n and m, respectively. Write

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$
$$q(x) = b_0 + b_1 x + \cdots + b_m x^m.$$

The x^i are considered independent and so the polynomials $p(x)$ and $q(x)$ are equal ($p(x) = q(x)$) if and only if $n = m$ and $a_i = b_i$ for all i. We can define two binary operations, addition and multiplication, on $R[x]$. Suppose $n \geq m$ and set $b_i = 0$ for $i > m$. Then

$$p(x) + q(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n.$$

This is similar if $m > n$. Now

$$p(x)q(x) = c_0 + c_1 x + \cdots + c_{m+n} x^{m+n},$$

where

$$c_i = \sum_{k=0}^{i} a_0 b_i + a_1 b_{i-1} + \cdots + a_{i-1} b_1 + a_i b_0.$$

Theorem 4. Let R be a ring. The set of polynomials $R[x]$ along with the operations defined above is a ring.

Proof. Above we showed that addition and multiplication are binary operations. It is easy to check that $(R[x], +)$ is an abelian group where the zero polynomial is the (additive) identity. The multiplicative identity is the constant polynomial 1. Associativity of multiplication and the distributive property are easy (albeit annoying) proofs. The details are left as an exercise. □

If R is a ring (commutative ring with identity) such that $ab = 0$ implies $a = 0$ or $b = 0$ (no zero divisors), then R is said to be a *domain*. Then next lemma justifies why we set $\deg(0) = -\infty$.

Lemma 5. Let R be an integral domain and let $p(x), q(x) \in R[x]$ be polynomials with degrees n and m. Then $\deg(p(x)q(x)) = \deg(p(x)) + \deg(q(x))$.

The next proposition is essentially the division algorithm for polynomials that you learn in high school.

Proposition 6. Let R be a ring. Let $f, g \in R[x]$ with f monic. Then there exist unique polynomials $q, r \in R[x]$ such that

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

where $\deg r < \deg f$.

If $r = 0$ in the above proposition, then we say that f divides g in $R[x]$. It is not strictly necessary that f be monic in the previous proposition, but we do need that the leading coefficient of f is a unit and in this case we can just factor that unit out.

If $g(x) \in R[x]$ and $\alpha \in R$, then the remainder of $g(x)$ when divided by $x - \alpha$ is just $g(\alpha)$. We will formally define this in the next section but for now it suffices to say that it is $g(x)$ but where x has been *replaced* by the constant α.

So far we've discussed polynomials in one variable, but it is relatively straightforward, albeit very tedious, to define polynomials in two or more variables. By the above theorem, if R is a commutative ring with identity then so is $R[x]$. If y is another indeterminate, then it makes sense to define $(R[x])[y]$. One could then show that this ring is isomorphic to $(R[y])[x]$. Both of these rings will be identified with the ring $R[x, y]$ and call this the ring of polynomials in two indeterminates x and y with coefficients in R. Similarly (or inductively), one can then define the ring of polynomials in n indeterminates with coefficients in R, denoted $R[x_1, \dots, x_n]$.

Disclaimer/Warning: If it is understood that there are a set of variables $x_1, \dots, x_n$, then $R[x_1, \dots, x_n]$ will just be denoted $R[x]$. Otherwise $R[x]$ will mean a polynomial in several variables.

A monomial is a formal product of some variables $x_1, \dots, x_n$ of the form

$$x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$$

with each i_k a nonnegative integer. The (total) degree of the monomial is $i_1 + \cdots + i_n$. For simplicity, we may refer to $i = (i_1, \dots, i_n)$ as a multi-index and simplify the above monomial to x^i. The monomial x^0 where $0 = (0, \dots, 0)$ is denoted by 1. Now a polynomial $f \in R[x_1, \dots, x_n] = R[x]$ is just

$$f(x) = \sum_i a_i x^i$$

where i runs through all multi-indices.

A polynomial in which all monomial with nonzero coefficients have the same (total) degree is called homogeneous.

3. Homomorphisms and ideals

We now extend notions of homomorphisms, cosets, and factor groups to rings.

Definition. A map $\phi : R \to S$ of rings is a (ring) homomorphism if

$$\phi(a + b) = \phi(a) + \phi(b), \qquad \phi(ab) = \phi(a)\phi(b), \text{ and } \qquad \phi(1) = 1.$$

The image of ϕ is $\operatorname{im} \phi = \{\phi(a) : a \in R\}$ and the kernel of ϕ is $\ker \phi = \{x \in R : \phi(x) = 0_s\}$. An isomorphism (of rings) is a bijective homomorphism.

Disclaimer/Warning: Not all authors require that $\phi(1) = 1$ and some allow the *zero map* to be a ring homomorphism.

Example. The map $\phi : \mathbb{Z} \to \mathbb{Z}_n$ given by $\phi(a) = a \mod n$ is a surjective homomorphism with $\ker \phi = n\mathbb{Z}$.

Proposition 7 (Substitution Principle)**.** Let $\phi : R \to R'$ be a ring homomorphism, and let α be an element of R'. There is a unique homomorphism $\Phi : R[x] \to R'$ that agrees with the map ϕ on constant polynomials, and that sends $x \mapsto \alpha$.

There is also a version of the substitution principle for multi-variable polynomial rings but it is not significantly different from the above.

Proof. Denote the image $\phi(a) \in R'$ of an element $a \in R$ by a'. By definition, $\Phi(1) = \phi(1) = 1$. Let $f = \sum a_i x^i$ and $g = \sum b_i x^i$ be polynomials in $R[x]$. Then

$$\begin{aligned}\Phi(f+g) = \Phi(\sum(a_i+b_i)x^i) &= \sum \Phi(a_i+b_i)\Phi(x)^i = \sum \phi(a_i+b_i)\alpha^i \\ &= \sum \phi(a_i)\alpha^i + \sum \phi(b_i)\alpha^i = \Phi(f) + \Phi(g).\end{aligned}$$

The proof that $\Phi(fg) = \Phi(f)\Phi(g)$ is similar.

Now let Ψ be any homomorphism that agrees with the map ϕ on constant polynomials and sends $x \mapsto \alpha$. Then

$$\Psi(\sum a_i x^i) = \sum \Psi(a_i)\Psi(x)^i = \sum a_i'\alpha^i = \Phi(\sum a_i x^i).$$

Hence, $\Psi = \Phi$. □

An easy consequence of the Substitution Principle is that any ring homomorphism $R \to S$ extends to a ring homomorphism $R[x] \to S[x]$.

Example. Let p be a prime and let $\phi : \mathbb{Z} \to \mathbb{F}_p$ be the homomorphism defined by $\phi(a) = \overline{a}$ where $\overline{a}$ denotes the residue of a modulo p. Then ϕ extends to a homomorphism $\Phi : \mathbb{Z}[x] \to \mathbb{F}_p[x]$ that sends $x \mapsto x$.

Ideals take the place of normal subgroups in ring theory in the sense that they are the right structure to allow us to define quotient rings.

Definition. An ideal of a ring R is a nonempty subset of R such that I is closed under addition, and if $s \in I$ and $r \in R$, then $rs \in I$.

Disclaimer/Warning: In a noncommutative ring, this is the definition of a *left ideal.* A *right ideal* is similar but we require that $sr \in I$. An *ideal*, or two-sided ideal, is both a left and right ideal.

Every ring R has two ideals: $\{0\}$ (called the trivial ideal) and R itself. An ideal I of R that is not one of these is called a proper ideal. As another example, $2\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

Proposition 8. Let $\phi : R \to S$ be a homomorphism of rings. Then $\ker \phi$ is an ideal of R.

In a ring R, the set $\langle a \rangle = \{ra : r \in R\}$ is an ideal in R, called the principal ideal generated by a. The elements of $\langle a \rangle$ are the R-multiples of a. This ideal is also denoted aR or Ra. Note that $R = (1)$ and the zero ideal is (0).

Example. Let $\phi : \mathbb{R}[x] \to \mathbb{R}$ be the substitution map that sends $x \mapsto 2$. Then $\ker \phi$ consists of those polynomials that have 2 as a root. Thus, $\ker \phi$ is the principal ideal $(x - 2)$.

Example. Let $\Phi : \mathbb{R}[x, y] \to \mathbb{R}[t]$ be the homomorphism that is the identity on $\mathbb{R}$ and sends $x \mapsto t^2$, $y \mapsto t^3$. It is clear that the polynomial $f(x, y) = y^2 - x^3 \in \ker \Phi$. We claim that $\ker \Phi = (f)$.

Suppose $g(x, y) \in \ker \Phi$. That is, $g(t^2, t^3) = 0$. We regard f as a (monic) polynomial in y with coefficients in $\mathbb{R}[x]$. Hence, by the division algorithm, $g = fq + r$ where q, r are polynomials and $\deg_y(r) < \deg_y(f) = 2$. Thus, $r(x, y) = r_1(x)y + r_0(x)$. Since $g, fq \in \ker \Phi$, then so is $r = g - fq$. Thus, $0 = r(t^2, t^3) = r_1(t^2)t^3 + r_0(t^2)$. The polynomials in $r_0(t^2)$ have even degree while those in $r_1(t^2)t^3$ have odd degree. It follows that $r_1 = r_0 = 0$ and so $r = 0$. That is, f divides g and so $g \in (f)$.

More generally, we can think of an ideal is like a spanning set. The ideal I of a ring R generated by the set $\{a_1, \ldots, a_n\} \subset R$ is the set of linear combinations

$$r_1 a_1 + \cdots + r_k a_k$$

with $r_i \in R$. Equivalently, I is the smallest ideal of R containing the a_i. We denote this ideal by $(a_1, \ldots, a_n)$.

Proposition 9. A ring R is a field if and only if it has exactly two ideals.

Proof. Suppose R is a field and I a nonzero ideal. Let $a \in I$. Then $1 = a^{-1}a \in I$, so $I = (1) = R$. Thus, R has exactly two ideals: (0) and (1).

Conversely, suppose R has exactly two ideals. Since the zero ring has only one ideal, we may assume that R is not the zero ring. Hence, the two ideals are (0) and (1), and they are not equal. Let $a \in R$ be nonzero and consider the principal ideal (a). This is not the zero ideal, and so $(a) = (1)$. Thus, $ra = 1$ for some $r \in R$, so $r = a^{-1}$. Hence, R is a field. □

Proposition 10. Every homomorphism $\phi : F \to R$ from a field F to a nonzero ring R is injective.

Proof. Recall that $K = \ker \phi$ is an ideal of F. By the previous proposition, $K = (0)$ or $K = (1)$. If $K = (1)$, then $\phi(1) = 0$, contradicting the definition of a ring homomorphism. Hence, $K = (0)$ and ϕ is injective. □

Proposition 11. Every ideal in $\mathbb{Z}$ is principal.

Proof. An ideal is necessarily a subgroup of the additive group $\mathbb{Z}^+$. We proved previously that every subgroup of $\mathbb{Z}^+$ is of the form $\mathbb{Z}n$. □

Proposition 12. Let F be a field. Every ideal in $F[x]$ is principal.

Proof. Let I be a nonzero ideal in $F[x]$. Set $S = \{\deg p(x) : p(x) \in I, p(x) \geq 0\}$. Since $I \neq 0$, then $S \neq \emptyset$ and $S \subset \mathbb{N}$. Thus, by the Well-Ordering Principal, S has a least element, d. Let $g(x) \in I$ be a polynomial of degree d. We claim $I = \langle g(x) \rangle$. It is clear that $\langle g(x) \rangle \subset I$ and so it is left only to prove the reverse inclusion.

Let $f(x) \in I$. If $\deg f(x) = d$, then either $f(x) = ag(x)$ for some $a \in F$ or else $\deg(f(x) - g(x)) < d$, a contradiction since $f(x) - g(x) \in I$. Now assume $\deg f(x) > d$. By the division algorithm, $f(x) = g(x)q(x) + r(x)$ for some $q(x), r(x) \in I$ with $\deg r(x) < \deg g(x)$. But then $r(x) = f(x) - g(x)q(x) \in I$. If $\deg r(x) \geq 0$, then this contradicts the minimality of d. Thus, $r(x) = 0$ and $f(x) \in \langle g(x) \rangle$ □

Disclaimer/Warning: The above result *does not* hold for $F[x, y]$. In particular, the ideal (x, y) is not principal.

Example. Let $\gamma = \sqrt[3]{2}$ be the real cube root of 2 and let $\phi : \mathbb{Q}[x] \to \mathbb{C}$ be the substitution map that sends $x \mapsto \gamma$. The kernel of ϕ is a principal ideal, generated by the lowest degree monic polynomial in $\mathbb{Q}[x]$ that has γ as a root. Since $x^3 - 2 \in \ker \phi$ and $\gamma \notin \mathbb{Q}$, then $x^3 - 2$ is not the product of two nonconstant polynomials with rational coefficients. Hence, $\ker \phi = (x^3 - 2)$. Note that this result still holds if we replace $\mathbb{Q}$ by $\mathbb{Z}$.

Lemma 13. Let f be a monic integer polynomial, and let g be another integer polynomial. If f divides g in $\mathbb{Q}[x]$, then f divides g in $\mathbb{Z}[x]$.

Proof. Since f is monic, we divide in $\mathbb{Z}[x]$, so $g = fq + r$ where r has integer coefficients. This remains true in $\mathbb{Q}[x]$. Hence, $g = fq + r$ and $r = 0$ because f divides g. □

The following corollary is proved in a manner similar to the corresponding results for integers.

Corollary 14. Let R denote the polynomial ring $F[x]$ in one variable over a field F and let $f, g \in R$, not both zero. Their greatest common divisor $d(x)$ is the unique monic polynomial that generates the ideal (f, g). It has these properties:

(1) $Rd = Rf + Rg$.
(2) d divides f and g.
(3) If a polynomial $e = e(x)$ divides both f and g, it also divides d.
(4) There are polynomial p and q such that $d = pf + qg$.

4. Quotient rings

Let I be an ideal of a ring R and let $\overline{R} = R^+/I^+$ denote the set of additive cosets $[a+I]$. Then $\overline{R}$ is a group under addition. We will next show that there is a multiplication operation such that $\overline{R}$ is a ring, called the quotient ring of R.

Theorem 15. Let I be an ideal of a ring R. There is a unique ring structure on the set $\overline{R}$ of additive cosets of I such that the map $\pi : R \to \overline{R}$ that sends $a \mapsto \overline{a} = [a+I]$ is a ring homomorphism. The kernel of π is the ideal I.

Proof. We know that $\overline{R}$ is an abelian group under addition. It is only left to show that the multiplication operation is well-defined, that it is associative, and that the distributive properties hold. Let $r+I, s+I \in R/I$. Suppose $r+I = r'+I$ and $s+I = s'+I$ for some $r', s' \in R$. We will show that $r's'+I = rs+I$. Checking associativity and the distributive properties are left as an exercise.

Our hypothesis implies that $r' \in r+I$ so $r' = r+a$ for some $a \in I$. Similarly, $s' = s+b$ for some $b \in I$. Then

$$r's' = (r+a)(s+b) = rs + rb + as + ab.$$

Since I is an ideal, $rb + as + ab \in I$, whence $r's' \in rs + I$. Thus, because cosets are either equal or disjoint, $r's' + I = rs + I$.

It is clear that ϕ is a surjective group homomorphism with kernel I. We need only show that it respects multiplication. Let $r, s \in R$. Then

$$\phi(r)\phi(s) = (r+I)(s+I) = rs + I = \phi(rs). \qquad \square$$

Theorem 16 (Mapping property of quotient rings)**.** Let $f : R \to R'$ be a ring homomorphism with kernel K and let I be another ideal. Let $\pi : R \to \overline{R}$ be the canonical map from R to $\overline{R} = R/I$.

(1) If $I \subset K$, there is a unique homomorphism $\overline{f} : \overline{R} \to R'$ such that $\overline{f}\pi = f$:

$$\begin{tikzcd}
R' \arrow[rr, "f"] \arrow[dr, "\pi"'] & & R' \\
& \overline{R} = R/I \arrow[ur, dashed, "\overline{f}"'] &
\end{tikzcd}$$

(2) If f is surjective and $I = K$, $\overline{f}$ is an isomorphism.

Theorem 17 (Correspondence Theorem)**.** Let $\phi : R \to \mathcal{R}$ be a surjective ring homomorphism with kernel K. There is a bijective correspoindence between the set of all ideals of $\mathcal{R}$ and the set of ideals of R that contain K:

$$\{\text{ideals of } R \text{ that contain } K\} \leftrightarrow \{\text{ideals of } \mathcal{R}\}.$$

The correspondence is defined as follows:

- If I is an ideal of R and if $K \subset I$, the corresponding ideal of $\mathcal{R}$ is $\phi(I)$.
- If $\mathcal{I}$ is an ideal of $\mathcal{R}$, the corresponding ideal of R is $\phi^{-1}(I)$.

If the ideal I of R corresponds to the ideal $\mathcal{I}$ of $\mathcal{R}$, the quotient rings R/I and $\mathcal{R}/\mathcal{I}$ are naturally isomorphic.

Proof. Let $\mathcal{I}$ be an ideal of $\mathcal{R}$ and I an ideal of R that contains K. We must prove the following:

- $\phi(I)$ is an ideal of $\mathcal{R}$.
- $\phi^{-1}(\mathcal{I})$ is an ideal of R, and it contains K.
- $\phi(\phi^{-1}(\mathcal{I})) = \mathcal{I}$, and $\phi^{-1}\phi(I)) = I$.
- If $\phi(I) = \mathcal{I}$, then $R/I \cong \mathcal{R}/\mathcal{I}$.

Note that these are the same items we proved for the Correspondence Theorem for groups. Hence, in each case, the corresponding results hold for the additive subgroup structure.

To show that $\phi(I)$ is an ideal of $\mathcal{R}$, we need only show that it is closed under multiplication by elements of $\mathcal{R}$. Let $\tilde{r} \in \mathcal{R}$ and $\tilde{x} \in \phi(I)$. By surjectivity, there exists $r \in R$ and $x \in I$ such that $\phi(r) = \tilde{r}$ and $\phi(x) = \tilde{x}$. Then $rx \in I$ because I is an ideal. Hence, $\tilde{r}\tilde{x} = \phi(r)\phi(x) = \phi(rx) \in \phi(I)$.

Similarly, for the second point we need only show that $\phi^{-1}(\mathcal{I})$ is closed under multiplication by elements of R. Let $a \in \phi^{-1}(\mathcal{I})$ and $r \in R$. Then there exists $\tilde{a} \in \mathcal{I}$ such that $\phi(a) = \tilde{a}$. Now $\phi(ra) = \phi(r)\phi(a) = \phi(r)\tilde{a} \in \mathcal{I}$ because $\mathcal{I}$ is an ideal. It follows that $ra \in \phi^{-1}(cI)$. Note that we never used surjectivity in this part.

The third point follows because ϕ is a group homomorphism.

For the last point, suppose I is an ideal of R that contains K. Set $\mathcal{I} = \phi(I)$ and assume that $I = \phi^{-1}(\mathcal{I})$. Let $\tilde{\pi} : R \to \mathcal{R}/\mathcal{I}$ be the canonical map and let $f = \tilde{\pi}\phi : R \to \mathcal{R} \to \mathcal{R}/\mathcal{I}$. Now $\ker f$ is the set of elements $x \in R$ such that $\tilde{\pi}\phi(x) = 0$. This holds if and only if $\phi(x) \in \mathcal{I}$ if and only if $x \in \phi^{-1}(\mathcal{I}) = I$. Thus, $\ker f = I$ and now the result follows by the First Isomorphism Theorem. □

Example. Let $\phi : \mathbb{C}[x, y] \to \mathbb{C}[t]$ be the homomorphism that sends $x \mapsto t$ and $y \mapsto t^2$. This is a surjective map and $K = \ker \phi = (y - x^2)$ (one can prove this in a manner similar to a previous example). Let I be an ideal of $\mathbb{C}[x, y]$ that contains K. Let $J = \phi^{-1}(I)$, so by the Correspondence Theorem, $I = \phi(J)$. By the division algorithm, $J = (p(x))$ for some monic polynomial $p(x)$. Now let I_1 be the ideal of $\mathbb{C}[x, y]$ generated by $y - x^2$ and $p(x)$. It is clear that $\phi^{-1}(I_1) = J$ and so $I_1 = I$. Hence, every ideal of $\mathbb{C}[x, y]$ that contains $y - x^2$ is of the form $(y - x^2, p(x))$ for some polynomial $p(x)$.

Let R be a ring and $I = (a)$ a principal ideal of R. One way to view $\overline{R}$ is as R but with the relation that $a = 0$. (E.g., $\mathbb{F}_7$ is $\mathbb{Z}$ where $7 = 0$).

Let $\pi : R \to \overline{R}$ be the canonical map. Then $\pi(a) = 0$. The elements of $\overline{R}$ are the cosets $b + I$. In R, these are the elements $b + ra$ for some $r \in R$ and so, using the idea above, we can view $\overline{R}$ as R but with the relation $b = b + ra$. This story also plays out in a similar way when I is generated by finitely many elements $a_1, \ldots, a_n$. Here, we will have $b = b + r_1a_1 + \cdots + r_na_n$ for some $r_i \in R$.

The Correspondence Theorem asserts that we can achieve this result either by factoring out individual elements one by one or all at the same time. That is, $R/(a, b) \cong (R/(a))/(b)$.

Example. Consider the ring $\overline{R} = \mathbb{Z}[i]/(i-2)$. We will show, somewhat unintuitively, that $\overline{R} \cong \mathbb{F}_5$.

Consider the map $\mathbb{Z}[x] \to \mathbb{Z}[i]$ defined by $x \mapsto i$. The kernel of this map is $f = x^2 + 1$ and so $\mathbb{Z}[x]/(f) \cong \mathbb{Z}[i]$ by the First Isomorphism Theorem. The image of $g = x - 2$ is $i - 2$, and so $\overline{R} \cong \mathbb{Z}[x]/I$ where $I = (f, g)$.

Alternatively, we could obtain $\overline{R}$ by factoring out the elements f and g in the opposite order. The principal ideal (g) is the kernel of the homomorphism $\mathbb{Z}[x] \to \mathbb{Z}$ that sends $x \mapsto 2$. The image of f under this map is 5 and $\mathbb{Z}/(5) \cong \mathbb{F}_5$. This is summarized in the diagram below.

$$\begin{tikzcd}
\mathbb{Z}[x] \arrow[r, "x-2"] \arrow[d, "x^2+1"'] \arrow[dr] & \mathbb{Z} \arrow[d, "5"] \\
\mathbb{Z}[i] \arrow[r, "i-2"'] & \mathbb{F}_5
\end{tikzcd}$$

5. Adjoining Elements

Let R be a ring that is a proper subring of R'. Let $\alpha \in R' \backslash R$, then as we have seen before, $R[\alpha]$ is defined as the smallest subring of R' containing R and α. The question in this section is how do we (formally) adjoin an element to a ring if the larger ring is not available?

Example. Let $P = \mathbb{R}[x]$. We form the quotient ring $\overline{P}/(x^2 + 1)$. Then we denote by i the residue of x. The relation $\overline{x}^2 + 1 = 0$ holds in $\overline{P}$ because the map $\pi : P \to \overline{P}$ is a homomorphism with kernel $(x^2 + 1)$. Hence, it follows that $\overline{P} \cong \mathbb{C}$.

In general, we wish to *adjoin* an element α that satisfies a polynomial relation $f(x) = 0$, where

$$f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0, \quad a_i \in R.$$

Let $R' = R[x]/(f)$, where (f) is the principal ideal of $R[x]$ generated by f. Let α denote the reside $\overline{x}$ of $x \in R'$. Then the map $\pi : R[x] \to R[x]/(f)$ is a homomorphism

$$\pi(f(x)) = \overline{f(x)} = \overline{a}_n\alpha^n + \cdots + \overline{a}_0 = 0$$

where $\overline{a}_i$ denotes the image in R' of the constant $a_i \in R$. Ignoring the bar notation, we see that α satisfies the relation $f(\alpha) = 0$. We denote this ring by $R[\alpha]$.

Disclaimer/Warning: It turns out the two rings we have denoted by $R[\alpha]$ agree (up to isomorphism) so long as α satisfied a polynomial relation. This second method will not apply when we try to adjoin a transcendental number. Say, for example, the ring $\mathbb{Q}[\pi]$.

Example. Let a be an element of a ring R. An inverse of a is an element α that satisfies the relation $a\alpha - 1 = 0$. We can adjoin an inverse by forming the quotient ring $R' = R[x]/(ax-1)$. What happens if $a^{-1} \in R$?

Proposition 18. Let R be a ring, and let $f(x)$ be a monic polynomial of positive degree n with coefficients in R. Let $R[\alpha]$ denote the ring $R[x]/(f)$ obtained by adjoining an element satisfying the relation $f(\alpha) = 0$.

(1) The set $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is a basis of $R[\alpha]$ over R.
(2) Addition of two linear combinations is vector addition.
(3) Multiplication of linear combinations is as follows: Let $\beta_1, \beta_2 \in R[\alpha]$ and let $g_1(x), g_2(x)$ be polynomials such that $\beta_1 = g_1(\alpha)$, $\beta_2 = g_2(\alpha)$. One divides the product $g_1 g_2$ by f, say $g_1 g_2 = fq + r$, where $\deg(r) < n$. Then $\beta_1\beta_2 = r(\alpha)$.

Proof. First we note that, because f is a monic polynomial of degree n in a polynomial ring $R[x]$, then every nonzero element of (f) has degree at least n.

(1) Since $R[\alpha]$ is a quotient of $R[x]$, every element β of $R[\alpha]$ is the residue of a polynomial $g(x)$, i.e., $\beta = g(\alpha)$. Since f is monic, we can perform division with remainder: $g(x) = f(x)q(x) + r(x)$ where $\deg(r(x)) < n$. Then since $f(\alpha) = 0$, $\beta = g(\alpha) = r(\alpha)$. Hence, β is written as a combination of the basis. The expression for β is unique because the principal ideal (f) contains no element of degree $< n$.

(2) and (3) follow because the operation of addition in $R[x]$ is vector addition. □

It's sufficient in the previous proposition to assume that f has a leading coefficient u that is a unit in R because in this case, $(f) = (u^{-1}f)$ and $u^{-1}f$ is monic. It is possible (but much harder) to adjoin an element when the polynomial relation isn't monic.

Example. Let $\phi : \mathbb{Z}[x] \mapsto \mathbb{C}$ be the substitution map sending $x \mapsto \gamma = \sqrt[3]{2}$. Then $\ker\phi = (x^3 - 2) = f(x)$ so $\mathbb{Z}[\gamma] \cong \mathbb{Z}[x]/(x^3-2)$. Then $\{1, \gamma, \gamma^2\}$ is a $\mathbb{Z}$-basis for $\mathbb{Z}[\gamma]$ and its elements are linear combinations $a_0 + a_1\gamma + a_2\gamma^2$, $a_i \in \mathbb{Z}$. If $\beta_1 = \gamma^2 - \gamma$ and $\beta_2 = \gamma^2 + 1$, then

$$\beta_1\beta_2 = \gamma^4 - \gamma^3 + \gamma^2 - \gamma = f(\gamma)(\gamma - 1) + (\gamma^2 + \gamma - 2) = \gamma^2 + \gamma - 2.$$

Example. Note that 3 does not have a square root in $\mathbb{F}_5$. Let $R' = \mathbb{F}_5[\delta]$ where δ satisfies the polynomial equation $f(x) = (x^2 - 3) = 0$. Thus, δ becomes an abstract square root of 3.

The elements of R' are the 25 linear combinations $a+b\delta$ with $a, b \in \mathbb{F}_5$. We claim that R' is a field. Set $c = (a+b\delta)(a-b\delta) = (a^2 - 3b^2) \in \mathbb{F}_5$. Since 3 isn't a square in $\mathbb{F}_5$ then $c = 0$ if and only if $a = b = 0$. Hence, if $a + b\delta \neq 0$, then its inverse is $(a - b\delta)c^{-1}$.

6. Product rings

Let $R \times R'$ be the Cartesian product of the rings R and R'. As we have seen previously, there is a group structure on $R \times R'$ defined by

$$(x, x') + (y, y') = (x + y, x' + y') \quad \text{for all } x, y \in R, x', y' \in R'.$$

This operation is associative, the (additive) identity is $(0_R, 0_{R'})$, and the inverse of (x, y) is $(-x, -y)$. There is also a multiplication operation defined by

$$(x, x')(y, y') = (xy, x'y') \quad \text{for all } x, y \in R, x', y' \in R'.$$

This operation is also associative and the element $(1_R, 1_{R'})$ is a multiplicative identity. It is easy to check that the distributive property holds for these two operations. Hence, $R \times R'$ is a ring under these two operations called the product ring of R and R'. The projection maps $\pi : R \times R' \to R$ and $\pi' : R \times R' \to R'$ given by $\pi(x, x') = x$ and $\pi'(x, x') = x'$ are ring homomorphisms with kernels $\{0\} \times R'$ and $R \times \{0\}$, respectively. A reasonable question now is whether a given ring can be *decomposed* into the product of two subrings.

An idempotent element e of a ring S is an element of S such that $e^2 = e$.

Example. The standard matrix units e_{ii} of $M_n(\mathbb{R})$ are idempotent.

Given an idempotent e is a ring S, the element $e' = 1 - e$ is also an idempotent, called the complement of e.

Proposition 19. Let e be an idempotent element of a ring S.

(1) The element $e' = 1 - e$ is also idempotent, $e + e' = 1$, and $ee' = 0$.
(2) The principal ideal eS of S is a ring (with inherited operations from S) with identity element e, and multiplication by e defines a ring homomorphism $S \to eS$.
(3) The ideal eS is not a subring of S unless e is the unit element 1 of S and $e' = 0$.
(4) The ring S is isomorphic to the product ring $eS \times e'S$.

Proof. (1) Since e is idempotent, $(e')^2 = (1-e)^2 = 1 - 2e + e^2 = 1 - 2e + e = 1 - e$. Clearly, $e + e' = 1$. Then $ee' = e(1-e) = e - e^2 = e - e = 0$.

(2) Since eS is an ideal of S, it has all of the properties of a ring except a multiplicative identity. Let $x \in eS$, then $x = es$ for some $s \in S$. Then $ex = e(es) = e^2 s = es = x$. Hence, e is a multiplicative identity for eS.

Define a map $\phi : S \to eS$ by $\phi(s) = es$. Let $a, b \in S$, then $\phi(a+b) = e(a+b) = ea+eb = \phi(a)+\phi(b)$ and $\phi(ab) = e(ab) = e^2(ab) = (ea)(eb) = \phi(a)\phi(b)$.

(3) A subring of S must contain the(multiplicative) identity element of S. But e is already a identity of eS and so if eS is a ring then $e = 1$ by uniqueness of the identity.

(4) Define a map $\psi : S \to eS \times e'S$ by $\psi(x) = (ex, e'x)$. Since the map ψ is a homomorphism in each coordinate by (2), then ψ is a homomorphism. (We also use the fact that, in the product ring, the operations are componentwise.)

Let $(a, b) \in eS \times e'S$. Then $a = ex$ and $b = e'y$ for some $x, y \in S$. Now by part (1),

$$\psi(ex + e'y) = (e(ex + e'y), e'(ex + e'y)) = (ex, e'y) = (a, b).$$

Hence, the map is surjective. It is left to show that ψ is injective. Assume $z \in \ker \psi$. Then $\psi(z) = 0$, so $ez = e'z = 0$. But then $0 = ez + e'z = (e + e')z = z$. Hence, $\ker \psi = \{0\}$. □

Disclaimer/Warning: Note that in two we made use of the fact that the multiplication operation in S is commutative.

Example. Let δ be an abstract square root of 3 in $\mathbb{F}_{11}$. The elements of $R' = \mathbb{F}_{11}[\delta]$ are the 11^2 linear combinations $a + \delta b$. Note that R' is *not* a field because $\mathbb{F}_{11}$ already contains two square roots of 3, namely ±5.

The elements $e = \delta - 5$ and $e' = -\delta - 5$ are complementary idempotents in R'. Thus, R' is isomorphic to $eR' \times e'R'$. As $|eR| = |e'R'| = 11$, then both eR and $e'R'$ are isomorphic to $\mathbb{F}_{11}$.

7. Fractions

The intent of this section is to make formal the process of forming the rationals $\mathbb{Q}$ from the integers $\mathbb{Z}$. This process, known as localization, is applicable to a large class of rings. We will define $\mathbb{Q}(x)$, the ring of rational functions in one variable (over $\mathbb{Q}$), using this process. Another goal here is to extend notions of divisibility and factorization in $\mathbb{Z}$ to integral domains.

Though we normally write each rational number in the form $p/q \in \mathbb{Q}$, we could just as easily think of them as ordered pairs $(p, q) \in \mathbb{Z} \times \mathbb{Z}$ subject to the condition that $q \neq 0$. Of course, the representation of an element in $\mathbb{Q}$ is not unique. Thus, if we are to represent the rationals as ordered pairs, we need a way to detect if they actually represent the same number in $\mathbb{Q}$. Recall that $a/b = c/d$ if and only if $ad = bc$. This then defines an equivalence relation on the ordered pairs.

We will work in our usual setup with one additional condition. Let R be a (commutative) ring (with identity). Also, we will assume that R is an integral domain (or just a domain). It is possible to do localization outside of this context, but the details are much more difficult.

Suppose R is an integral domain such that $ab = ac$. Then $a(b - c) = 0$ so $a = 0$ or $b = c$. Hence, an integral domain satisfies the *cancellation law*:

$$\text{If } ab = ac \text{ and } a \neq 0, \text{ then } b = c.$$

Lemma 20. Let R be any integral domain and define the set

$$S = \{(a, b) : a, b \in R \text{ and } b \neq 0\}.$$

The relation $\sim$ on S given by $(a, b) \sim (c, d)$ if $ad = bc$ is an equivalence relation.

Proof. Let $(a, b) \in S$. Then $ab = ba$ because R is an integral domain (and hence commutative). Thus, $(a, b) \sim (a, b)$ and so $\sim$ is reflexive.

Let $(a, b), (c, d) \in S$ such that $(a, b) \sim (c, d)$. Then $ad = bc$ and so $cb = da$ because R is an integral domain. Thus, $(c, d) \sim (a, b)$ and so $\sim$ is reflexive.

Suppose $(a, b) \sim (c, d)$ and $(c, d) \sim (e, f)$ in S. Then $ad = bc$ and $cf = de$. Since R is an integral domain without zero divisors, then $ade = bce$ so $acf = bce$. Thus, $af = be$ so $(a, b) \sim (e, f)$ and $\sim$ is transitive. □

Let $[a, b]$ denote the equivalence class of $(a, b) \in S$ under $\sim$. The set of such equivalence classes is denoted F_R. We will now prove that F_R is in fact a field, justifying the name *fraction field* of R.

The operations on F_R are defined to mimic the operations of adding and multiplying fractions. Let $[a, b], [c, d] \in F_R$ and define operations of addition and multiplication by

$$[a, b] + [c, d] = [ad + bc, bd]$$
$$[a, b] \cdot [c, d] = [ac, bd].$$

Lemma 21. The operations of addition and multiplication above are well-defined binary operations on F_R.

Proof. That F_R is closed under these operations is clear. We need only show that they are well-defined (independent of equivalence class).

Let $[a_1, b_1], [a_2, b_2], [c_1, d_1], [c_2, d_2] \in F_R$ with $[a_1, b_1] = [a_2, b_2]$ and $[c_1, d_1] = [c_2, d_2]$. Thus, $a_1b_2 = b_1a_2$ and $c_1d_2 = d_1c_2$. We claim $[a_1, b_1] + [c_1, d_1] = [a_2, b_2] + [c_2, d_2]$. Equivalently,

$$[a_1d_1 + b_1c_1, b_1d_1] = [a_2d_2 + b_2c_2, b_2d_2]$$

or

$$(a_1d_1 + b_1c_1)(b_2d_2) = (b_1d_1)(a_2d_2 + b_2c_2).$$

We have

$$\begin{aligned}(a_1d_1 + b_1c_1)(b_2d_2) &= (a_1d_1)(b_2d_2) + (b_1c_1)(b_2d_2)\\ &= (a_1b_2)(d_1d_2) + (b_1b_2)(c_1d_2)\\ &= (b_1a_2)(d_1d_2) + (b_1b_2)(d_1c_2)\\ &= (b_1d_1)(a_2d_2 + b_2c_2).\end{aligned}$$

Checking the operation of multiplication is left as an (easier) exercise. □

Lemma 22. The set F_R with the above binary operations is a field.

Proof. We claim the additive identity is $[0, 1]$. Let $[a, b] \in F_R$, then

$$[a, b] + [0, 1] = [a \cdot 1 + b \cdot 0, b \cdot 1] = [a, b].$$

Associativity is left as an exercise. We claim the additive inverse of $[a, b]$ is $[-a, b]$ (note $-a \in R$ because R is a ring). One checks that $[a, b] + [-a, b] = [ab + b(-a), b^2] = [0, b^2] = [0, 1]$. (Note that this last equality follows because $0 \cdot 1 = b^2 \cdot 0$.) Thus, $(F_R, +)$ is an abelian group.

Associativity and commutivity of multiplication are left as an exercise. We check the left distributive property. Let $[a, b], [c, d], [e, f] \in F_R$. Then

$$\begin{aligned}[a, b][c, d] + [a, b][e, f] &= [ac, bd] + [ae, bf]\\ &= [acbf + bdae, bdbf]\\ &= [acf + dae, bdf]\\ &= [a, b][cf + de, df]\\ &= [a, b]([c, d] + [e, f]).\end{aligned}$$

By commutivity, the right distributive property also holds. Hence, F_D is a commutative ring.

The multiplicative identity in F_R is $[1, 1]$. This is an easy check: $[a, b][1, 1] = [a, b]$. Moreover, the multiplicative inverse of $[a, b]$ is $[b, a]$ as

$$[a, b][b, a] = [ab, ba] = [1, 1].$$

Thus, F_R a field. □

Definition. The field F_R defined above is the fraction field of R.

The next theorem is a sort of uniqueness result for F_R. It shows that F_R is in some sense the *smallest* field containing R.

Theorem 23. Let R be an integral domain. Then R can be embedded in a field of fractions F_R, where any element in F_R can be expressed as a quotient of two elements in R.

Furthermore, if E is any field containing R, then there exists a map $\psi : F_R \to E$ given an isomorphism with a subfield of E such that $\psi(a) = a$ for all $a \in R$.

Proof. Define a map

$$\begin{aligned}\phi : R &\to F_R \\ a &\mapsto [a, 1].\end{aligned}$$

We claim ϕ is an injective homomorphism. Let $a, b \in R$, then

$$\begin{aligned}\phi(a) + \phi(b) &= [a, 1] + [b, 1] = [a + b, 1] = \phi(a + b) \\ \phi(a)\phi(b) &= [a, 1][b, 1] = [ab, 1] = \phi(ab).\end{aligned}$$

Thus, ϕ is a ring homomorphism. Now suppose $\phi(a) = 0$, then $[a, 1] = [0, 1]$, so $a = 0$. Thus, $\ker \phi = \{0\}$ and ϕ is injective.

Now if $[a, b] \in F_R$, then

$$[a, b] = [a, 1][1, b] = [a, 1][b, 1]^{-1} = \phi(a)[\phi(b)]^{-1},$$

so every element in F_R can be expressed as a quotient of two elements in R.

Let E be any field containing R. Define a map

$$\begin{aligned}\psi : F_R &\to E \\ [a, b] &\mapsto ab^{-1}.\end{aligned}$$

We must show that ψ is well-defined. Suppose $[a_1, b_1], [a_2, b_2] \in F_R$ with $[a_1, b_1] = [a_2, b_2]$. Then $a_1b_2 = b_1a_2$. Thus, in E, $a_1b_1^{-1} = a_2b_2^{-1}$ and so $\psi(a_1b_1^{-1}) = \psi(a_2b_2^{-1})$.

Now we claim ψ is an injective homomorphism. Let $[a, b], [c, d] \in F_R$. Then

$$\begin{aligned}\psi([a, b] + [c, d]) &= \psi([ad + bc, bd]) = (ad + bc)(bd)^{-1} = ab^{-1} + cd^{-1} = \psi([a, b]) + \psi([c, d]) \\ \psi([a, b][c, d]) &= \psi([ac, bd]) = (ac)(bd)^{-1} = (ab^{-1})(cd^{-1}) = \psi([a, b])\psi([c, d]).\end{aligned}$$

For injectivity, suppose $\psi([a, b]) = 0$. Then $ab^{-1} = 0$. Multiplying both sides by b gives $a = 0$. Thus, $[a, b] = [0, 1]$, the additive identity in F_R, so ψ is injective.

Identifying a with $[a, 1]$ through ϕ, we see that $\psi(a) = \psi([a, 1]) = a1^{-1} = a$. □

Such a result as above is called a *universal property*. We can visualize it using the following *commutative diagram.*

$$\begin{tikzcd} R \arrow[r, "\iota"] \arrow[d, "\phi"'] & E \\ F_R \arrow[ur, dashed, "\psi"'] & \end{tikzcd}$$

Here, ϕ and ψ are as in the theorem and ι is the inclusion map.

Example. The polynomial ring $\mathbb{Q}[x]$ is an integral domain. The fraction field of $\mathbb{Q}[x]$ is the set of rational expressions $p(x)/q(x)$ for polynomials $p(x), q(x) \in \mathbb{Q}[x]$ with $q(x) \neq 0$. We denote this field by $\mathbb{Q}(x)$.

8. Maximal ideals

A maximal ideal M of a ring R is a *proper* ideal such that if I is any ideal such that $M \subset I \subset R$, then $I = M$ or $I = R$.

Proposition 24. (1) Let $\phi : R \to R'$ be a surjective ring homomorphism, with kernel K. The image R' is a field if and only if K is a maximal ideal.
(2) An ideal I of a ring is maximal if and only if $\overline{R} = R/I$ is a field.
(3) The zero ideal of a ring R is maximal if and only if R is a field.

Proof. (1) Suppose R' is a field, then R' has only two ideals ((0) and $(1) = F$) and so by the Correspondence Theorem the inverse image of (1) is $K = \ker \phi$. Thus, the only ideals that contain K are K and R. Hence, K is a maximal ideal.

Conversely, if K is a maximal ideal, then under the Correspondence Theorem K corresponds to the zero ideal. But then R' has exactly two ideals and so R' is a field.

(2) Apply part (1) to the canonical map $R \to R/I$.

(3) This statement is equivalent to saying that R has exactly two ideals. □

Proposition 25. The maximal ideals of the ring $\mathbb{Z}$ of integers are the principal ideals generated by prime integers.

Proof. Every ideal of $\mathbb{Z}$ is principal. Consider the principal ideal (n), with $n > 0$. (Note that the zero ideal is not maximal because $\mathbb{Z}$ is not a field). If $n = p$ with p prime, then $\mathbb{Z}/(n) \cong \mathbb{F}_p$, a field. Hence (n) is maximal. Conversely, suppose $n = pm$ for some prime p, $m \neq 1$. Then $(n) \subset (p)$ and (n) is not maximal. □

Next we consider a story analogous to the one for the integers, but with polynomials over a field. A polynomial with coefficients in a field is called irreducible if it is not constant and if it is not the product of two polynomials, neither of which is a constant.

Proposition 26. (1) Let F be a field. The maximal ideals of $F[x]$ are the principal ideals generated by the monic irreducible polynomials.

(2) Let $\phi : F[x] \to R'$ be a homomorphism to an integral domain R', and let K be the kernel of ϕ. Either K is a maximal ideal, or $K = (0)$.

Proof. The proof of (1) is similar to the previous proposition. For (2), suppose that $K \neq (0)$. Then $K = (p(x))$ for some nonzero polynomial $p(x) \in F[x]$. Note that $p(x)$ has minimal possible degree.

We claim that $p(x)$ is irreducible. Suppose otherwise. That is, $p(x) = p_1(x)p_2(x)$ for some nonconstant polynomials $p_1(x), p_2(x) \in F[x]$. Then

$$0 = \phi(p(x)) = \phi(p_1(x)p_2(x)) = \phi(p_1(x))\phi(p_2(x)).$$

Because R' is an integral domain, this implies that $p_1(x) \in K$ or $p_2(x) \in K$, contradicting the choice of $p(x)$ (they must have lower degree than $p(x)$). Consequently, $p(x)$ is irreducible and so, by (1), K is a maximal ideal. □

Corollary 27. There is a bijective correspondence between maximal ideals of the polynomial ring $\mathbb{C}[x]$ in one variable and points in the complex plane. The maximal ideal M_a that corresponds to a point a of $\mathbb{C}$ is the kernel of the substitution homomorphism $s_a : \mathbb{C}[x] \to \mathbb{C}$ that sends $x \mapsto a$. It is the principal ideal generated by the linear polynomial $x - a$.

Proof. The kernel M_a of the substitution homomorphism consists of those polynomials that have a as a root, which are divisible by $x - a$. Hence, $M_a = (x - a)$. Conversely, if M is a maximal ideal of $\mathbb{C}[x]$ then M is generated by a monic irreducible polynomial, so $M = (x - a)$. □

The word *Nullstellensatz* is a German word that is the combinations of three words whose translations are zero, places, theorem.

Theorem 28 (Hilbert's Nullstellensatz)**.** The maximal ideals of the polynomial ring $\mathbb{C}[x_1, \dots, x_n]$ are in bijective correspondence with points of complex n-dimensional space. A point $a = (a_1, \dots, a_n) \in \mathbb{C}^n$ corresponds to the kernel M_a of the substitution map $s_a : \mathbb{C}[x_1, \dots, x_n] \to \mathbb{C}$ that sends $x_i \mapsto a_i$. The kernel M_a is generated by the n linear polynomials $x_i - a_i$.

9. Algebraic Geometry

A point $a = (a_1, \dots, a_n) \in \mathbb{C}^n$ is called a zero of a polynomial $f(x_1, \dots, x_n) \in \mathbb{C}[x_1, \dots, x_n]$ if $f(a_1, \dots, a_n) = 0$. We say f vanishes at a. The common zeros of a set $\{f_1, \dots, f_r\} \subset \mathbb{C}[x_1, \dots, x_n]$ are the set of points in $\mathbb{C}^n$ that vanish on each f_i.

Definition. A subset V of a complex n-space $\mathbb{C}^n$ that is the set of common zeros of a finite number of polynomials in n variables is called an (algebraic) variety.

Example. The following are examples of varieties.

(1) A point $(a, b) \in \mathbb{C}^2$ is a variety because it is the set of solutions of $\{x - a, y - b\}$.
(2) A complex line in the (x, y)-plane $\mathbb{C}^2$ is the set of solutions of a linear equation $ax + by + c = 0$.
(3) The group $\mathrm{SL}_2(\mathbb{C}) \subset \mathbb{C}^{2\times 2}$ is the set of common zeros of the polynomial $x_{11}x_{22} - x_{12}x_{21} - 1$.

Theorem 29. Let I be an ideal of $\mathbb{C}[x_1, \ldots, x_n]$ generated by some polynomials $f_1, \ldots, f_r$, and let $R = \mathbb{C}[x_1, \ldots, x_n]/I$. Let V be the variety of (common) zeros of the polynomials $f_1, \ldots, f_r$ in $\mathbb{C}^n$. The maximal ideals of R are in bijective correspondence with the points of V.

Proof. By the Correspondence Theorem, the maximal ideals of R correspond to maximal ideals of $\mathbb{C}[x_1, \ldots, x_n]$ that contain I. An ideal of $\mathbb{C}[x_1, \ldots, x_n]$ will contain I if and only if it contains the generators $f_1, \ldots, f_r$ of I. Every maximal ideal of the ring $\mathbb{C}[x_1, \ldots, x_n]$ is the kernel M_a of the substitution map that sends $x_i \mapsto a_i$ for some point $a = (a_1, \ldots, a_n) \in \mathbb{C}^n$, and the polynomials $f_1, \ldots, f_r$ are in M_a if and only if $f_1(a) = \cdots = f_r(a) = 0$ if and only if $a \in V$. □

Algebraic geometry is (roughly) the study of the relationships between the ring $R = \mathbb{C}[x_1, \ldots, x_n]/I$ and the geometric properties of V.

Theorem 30. Let R be a ring. Every proper ideal I of R is contained in a maximal ideal.

Proof. Let I be an ideal, if I is not maximal, then there exists a proper ideal I_1 properly containing I. Continue this process inductively. It follows that the set of proper ideals forms a partially ordered set. Thus, by Zorn's Lemma (every ascending chain in a partially ordered set has a least upper bound), I is contained in some proper maximal ideal. □

Corollary 31. The only ring R having no maximal ideals is the zero ring.

Corollary 32. If a system of polynomial equations $f_1 = \cdots = f_r = 0$ in n variables has no solution in $\mathbb{C}^n$, then 1 is a linear combination $1 = \sum g_i f_i$ with polynomial coefficients g_i.

Proof. If the system has no solution, there is no maximal ideal that contains the ideal $I = (f_1, \ldots, f_r)$. Hence, I is the unit ideal and $1 \in I$. □

Generically, in a polynomial ring with two variables, one would expect that a system of three polynomials $f_1 = f_2 = f_3 = 0$ would have no solution and thus $I = (f_1, f_2, f_3) = (1)$.

Lemma 33. Let $f(t, x)$ be a polynomial and let $\alpha \in \mathbb{C}$. The following are equivalent:

(1) $f(t, x)$ vanishes at every point of the locus $\{t = \alpha\}$ in $\mathbb{C}^2$,
(2) The one-variable polynomial $f(\alpha, x)$ is the zero polynomial,

(3) $t - \alpha$ divides f in $\mathbb{C}[t, x]$.

Proof. (1) $\Rightarrow$ (2) If f vanishes at every point of the locus $t = \alpha$, then $f(\alpha, x) = 0$ for every x. A nonzero polynomial in one variable has finitely many roots, so $f(\alpha, x)$ is the zero polynomial.

(2) $\Rightarrow$ (3) We make a change of variable $t = t' + \alpha$. If $f(0, x)$ is the zero polynomial, then t divides every monomial that occurs in f, so t divides f.

(3) $\Rightarrow$ (1) Clear. □

Let $\mathcal{F} = \mathbb{C}(t)$. The ring $\mathbb{C}[t, x]$ is a subring of $\mathcal{F}[x]$, whose elements are of the form

$$f(t, x) = a_n(t)x^n + \cdots + a_1(t)x + a_0(t), \tag{1}$$

where $a_i(t)$ are rational functions in t.

Proposition 34. Let $h(t, x)$ and $f(t, x)$ be nonzero elements of $\mathbb{C}[t, x]$. Suppose that h is not divisible by any polynomial of the form $t - \alpha$. If h divides f in $\mathcal{F}[x]$, then h divides f in $\mathbb{C}[t, x]$.

Proof. In $\mathcal{F}$, we divde f by h to get $f = hq$. We claim that $q \in \mathbb{C}[t, x]$. Since $q \in \mathcal{F}[x]$, then q has the form (1). We multiply both sides of $f = hq$ by t to clear denominators in these coefficients to arrive that

$$u(t)f(t, x) = h(t, x)q_1(t, x),$$

where $u(t)$ is a monic polynomial in t and $q_1 \in \mathbb{C}[t, x]$. We proceed by induction on $\deg(u)$.

If u has positive degree, it has a complex root α. Then $t - \alpha$ divides the left-hand side of the equation, so it also divides the right-hand side. Thus, $h(\alpha, x)q_1(\alpha, x)$ is the zero polynomial in x. By hypothesis, $t - \alpha$ does not divide h, so $h(\alpha, x)$ is not zero. But $\mathbb{C}[x]$ is a domain so $q_1(\alpha, x) = 0$, and the lemma shows that $t - \alpha$ divides $q_(t, x)$. We cancel $t - \alpha$ from u and q_1. The proof is now complete by induction. □

Let $f, g \in \mathbb{C}[x_1, \ldots, x_n]$ have degrees m and n, respectively. The *Bezout bound* says that the number of common zeros of f and g is at most mn. We won't prove this statement but will prove that the bound is finite in the $n = 2$ case.

Theorem 35. Two nonzero polynomials $f(t, x)$ and $g(t, x)$ in two variables have only finitely many common zeros in $\mathbb{C}^2$, unless they have a common nonconstant factor in $\mathbb{C}[t, x]$.

Proof. Assume f and g have no common factors and let $I = (f, g)$ in $\mathcal{F}[x]$ where $\mathcal{F} = \mathbb{C}(t)$. Then $I = (h)$ where h is the gcd of f and g in $\mathcal{F}[x]$.

Suppose $h \neq 1$. Then h is a polynomial whose coefficients may have denominators that are polynomials in t. We multiply through to clear denominators to obtain $h_1 \in \mathbb{C}[t, x]$ and we may assume that h_1 is not divisible by $t - \alpha$ (else this would be a common factor of f and g). Note we are multiplying by units and so $(h) = (h_1)$ and so h_1 divides f and g in $\mathcal{F}[x]$. Now by the previous proposition, h_1 divides f and g in $\mathbb{C}[t, x]$, contradicting our hypothesis.

Hence, the gcd of f and g in $\mathcal{F}[x]$ is 1 and so $1 = rf+sg$ for some $r, s \in \mathcal{F}[x]$. Clearing denominators from r and s by multiplying by a suitable polynomial $u(t)$ gives

$$u(t) = r_1(t, x)f(t, x) + s_1(t, x)g(t, x),$$

where all polynomials on the right are in $\mathbb{C}[t, x]$. Thus, if (t_0, x_0) is a common zero of f and g, then it must be a root of u. But u is a polynomial in one variable and hence has only finitely many roots. Thus, amongst the roots of f and g, t takes on only finitely many values. A similar argument shows the same for x. □

The locus X of zeros in $\mathbb{C}^2$ of a polynomial $f(t, x)$ is called the Riemann surface of f. (It is also called a plane algebraic curve because as a topological space X has dimension 2.)

Assume $f = f(t, x)$ is irreducible and has positive degree in the variable x. Let $X = \{(t, x) \in \mathbb{C}^2 : f(t, x) = 0\}$ be its Riemann surface, and let T denote the complex t-plane. Sending $(t, x) \mapsto t$ defines a continuous projection map $\pi : X \to T$.

Definition. Let X and T be Hausdorff spaces[1]. A continuous map $\pi : X \to T$ is an n-sheeted covering space if every fibre consists of n points, and if it has the property: Let x_0 be a point of X and let $\pi(x_0) = t_0$. Then π maps an open neighborhood U of x_0 in X homeomorphically to an open neighborhood V of t_0 in T.

A map π from X to the complex plane T is an n-sheeted branched covering if X contains no isolated points, the fibres of π are finite, and if there is a finite set Δ of points of T called brach points, such that the map $(X - \pi^{-1}\Delta) \to (T - \Delta)$ is an n-sheeted covering space. We refer to the points in Δ as branch points

Theorem 36. Let $f(t, x)$ be an irreducible polynomial in $\mathbb{C}[t, x]$ that has positive degree n in the variable x. The Riemann surface of f is an n-sheeted branched covering of the complex plane T.

Proof. The points of the fibre $\pi^{-1}(t_0)$ are the points (t_0, x_0) such that x_0 is a root of the one-variable polynomial $f(t_0, x)$. We must show that, except for a finite set of values $t = t_0$ (our branch points), this polynomial has n distinct roots. We write

$$f(t, x) = f(x) = a_n(t)x^n + \cdots + a_1(t)x + a_0(t)$$

with $a_i(t) \in \mathbb{C}[t]$. The polynomial $f(t_0, x)$ has x-degree at most n and so it has at most n roots. Therefore the fibre $\pi^{-1}(t_0)$ contains at most n points. It will have fewer than n points if either

(1) the degree of $f(t_0, x)$ is less than n, or

[1] [Munkres] A topological space X is Hausdorff if for each pair of distinct points x_1 and x_2 in X there exists open neighborhoods U_1 and U_2 of x_1 and x_2, respectively, that are disjoint.

(2) $f(t_0, x)$ has a multiple root.

The first case occurs when t_0 is a root of $a_n(t)$. Since $a_n(t)$ is a polynomial, there are finitely many such values (this is our Δ). For the second case we note that a complex number x_0 is a multiple root of a polynomial $h(x)$ if $(x - x_0)^2$ divides $h(x)$ if and only if $(x - x_0)$ divides $h(x)$ and $h'(x)$. Take $h(x) = f(t_0, x)$ (so taking a derivative is equivalent to taking the partial derivative $\frac{\partial f}{\partial x}$). Thus, case 2 occurs at the points (t_0, x_0) at points (t_0, x_0) that are common zeros of f and $\frac{\partial f}{\partial x}$. Since f cannot divide its partial derivative (lower degree in x) and f is assumed to be irreducible, then f and $\frac{\partial f}{\partial x}$ have no common nonconstant factors, whence there are finitely many common zeros by the previous theorem.

It is left only to check the second condition in the definition. Let t_0 be a point of T such that the fibre $\pi^{-1}(t_0)$ consists of n points, and let (t_0, x_0) be a point of X in the fibre. Then x_0 is a simple root of $f(t_0, x)$, and therefore $\frac{\partial f}{\partial x}$ is not zero at this point. The Implicit Function Theorem[2] implies that one can solve for x as a function $x(t)$ of t in a neighborhood of t_0, such that $x(t_0) = x_0$. The neighborhood U referred to in the definition of covering space is the graph of this function. □

[2] Let $f(x, y)$ be a complex polynomial. Suppose that for some $(a, b) \in \mathbb{C}^2$, $f(a, b) = 0$, and $\frac{\partial f}{\partial x}(a, b) \neq 0$. There is a neighborhood U of x in $\mathbb{C}$ on which a unique continuous function $Y(x)$ exists having the properties $f(x, Y(x)) = 0$ and $Y(a) = b$.

Factoring

(Last Updated: October 20, 2018)

These notes are derived primarily from *Algebra* by Michael Artin (2ed), though some material is drawn from *Abstract Algebra, Theory and Applications* by Thomas Judson (16ed).

1. Factoring Integers

Before studying division in rings, we recall basic facts about division in the integers. All of these facts will be proved in more generality in this chapter. Most of these facts follow from Division Algorithm.

Division Algorithm (for $\mathbb{Z}$): If a and b are integers and a is positive, there exist (unique) integers q and r so that $b = aq + r$ with $0 \leq r < a$.

Theorem 1. (1) Every ideal of the ring $\mathbb{Z}$ of integers is principal.

(2) A pair $a, b \in \mathbb{Z}$, not both zero, has a greatest common divisor, a positive integer with these properties:
   (a) $\mathbb{Z}d = \mathbb{Z}a + \mathbb{Z}b$,
   (b) d divides both a and b,
   (c) if an integer e divides a and b, then e divides d.
   (d) There are integers r and s such that $d = ra + sb$.

(3) If a prime integer p divides a product ab of integers, then p divides a or p divides b.

(4) *Fundamental Theorem of Arithmetic:* Every positive integer $a \neq 1$ can be written as a product $a = p_1 \cdots p_k$ of positive prime integers. This expression is unique up to reordering.

2. Unique Factorization Domains

Throughout this section, R is an integral domain. Our interest will be in discovering what rings satisfy a version of Theorem 1. We note immediately that this is nontrivial. Consider the ring $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$. The element 6 has two factorizations:

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

Thus, the FTA does not hold in $\mathbb{Z}[\sqrt{-5}]$.

Before proceeding, let us review some terminology from $\mathbb{Z}$ and then reframe it in the context of ideals.

term	definition in $\mathbb{Z}$	ideal-theoretic definition
u is a unit	u has a multiplicative inverse in R	$(u) = (1)$
a divides b	$b = aq$ for some $q \in R$	$(b) \subset (a)$
a is a proper divisor of b	$b = aq$ and neither a nor q is a unit	$(b) \subsetneq (a) \subsetneq (1)$
a and b are associates	$b = ua$ where u is a unit	$(b) = (a)$
a is irreducible	if $b \mid a$, then b is a unit or an associate	$(a) \subsetneq (1)$ and if $(a) \subsetneq (c)$, then $(c) = 1$
p is a prime element	p is not a unit and if $p \mid ab$, then $p \mid a$ or $p \mid b$	$ab \in (p)$ implies $a \in (p)$ or $b \in (p)$.

The next lemma shows how, in integral domains, some of these concepts are related.

Lemma 2. In an integral domain R, a prime element is irreducible.

Proof. Let p be a prime element and suppose that $p = ab$. This means that both a and b divide p. But as p is prime, it divides one of the factors, say a. Hence, a and p are associates and b is a unit. Thus, the factorization is not proper. □

Roughly, a Euclidean domain is an integral domain in which division "works". More precisely, we define a size function on a ring R is any function $\sigma : R\backslash\{0\} \to \mathbb{Z}_{\geq 0}$. An integral domain R is a Euclidean domain if there is a size function σ on R such that for any $a, b \in R$ with $a \neq 0$, there are elements $q, r \in R$ such that $b = aq + r$ and either $r = 0$ or $\sigma(r) < \sigma(a)$. (As an alternate convention, can also take $\sigma : R\backslash\{0\} \to \mathbb{Z}_{\geq 0} \cup \{-\infty\}$ and assign $\sigma(0) = -\infty$.)

Disclaimer/Warning: We do not assume that the division is unique in general.

Proposition 3. (1) The ring $\mathbb{Z}$ of integers is a Euclidean domain, with size function $\sigma(a) = |a|$.
(2) A polynomial ring $F[x]$ in one variable over a field F is a Euclidean domain, with $\sigma(f) = \deg(f)$.
(3) The ring $\mathbb{Z}[i]$ of Gauss integers is a Euclidean domain, with $\sigma(a) = |a|^2$.

Proof. The first two points were proved previously. We prove the third as in Artin, but with a few more details[1]. The Gaussian integers may be viewed as integer lattice of points in the plane. Hence, any nonzero $\alpha \in \mathbb{Z}[i]$ can be represented as $\alpha = re^{i\theta}$, where θ is the angle of rotation and r is the distance from the origin to α.

Let $\beta \in \mathbb{Z}[i]$ and set $\gamma = \beta/\alpha$. Note that $\gamma \in \mathbb{C}$ but γ need not be a Gaussian integer. Choose $q \in \mathbb{Z}[i]$ such that $\sigma(\gamma - q) < \frac{1}{2}$. Then

$$\sigma(\beta - q\alpha) = \sigma((\gamma - q)\alpha) = \sigma(\gamma - q)r^2 \leq r^2/2.$$

Thus, $\beta = q\alpha + r$ with $\sigma(r) < \sigma(\alpha)$. □

[1] Courtesy of James M$^{\text{c}}$Kernan's lecture notes: `http://math.mit.edu/~mckernan/Teaching/12-13/Spring/18.703/l_20.pdf`.

A principal ideal domain (PID) is an integral domain in which every ideal is principal. The next proof closely follows the proof for $\mathbb{Z}$ and $F[x]$.

Proposition 4. A Euclidean domain is a PID.

Proof. Let R be an Euclidean domain with size function σ and let A be a nonzero ideal (because the zero ideal is automatically principal). Choose $a \in A$ such that $\sigma(a)$ is as small as possible (Well-Ordering Principle). Clearly $(a) \subset A$ so we must show the reverse inclusion. Let $b \in A$, then we divide a into b to get $b = aq + r$ for some $q, r \in R$. If $r \neq 0$, then $\sigma(r) < \sigma(a)$. But then $r = b - aq \in A$, contradicting our choice for a. Hence, $r = 0$ so $b \in (a)$. Hence, $A \subset (a)$ and we have equality. □

Let R be an integral domain and $a, b \in R$ not both zero. A greatest common divisor (gcd) of a and b is an element $d \in R$ such that d divides a and b, and if e divides a and b then e divides d. Greatest common divisors need not be unique, or exist at all. However, the second condition tells us that if d and d' are gcds of a and b, then they divide each other and therefore are associates. We say that a and b are relatively prime if a gcd d exists and $d = 1$. The proof of the next proposition is not much different than that for integers.

Proposition 5. Let R be a PID, and let $a, b \in R$ not both zero. An element $d \in R$ that generates the ideal (a, b) is a gcd of a and b. There are elements $r, s \in R$ such that $d = ra + sb$.

Proof. Since $d \mid a$ and $d \mid b$, then d divides any R-linear combination of a and b. Hence, $(a, b) \subset (d)$. As R is a PID, then $(a, b) \subset (c)$ for some $c \in R$. But then $c \mid a$ and $c \mid b$, so $c \mid d$ because d is the gcd of a and b. But then $(d) \subset (c)$. It follows that there are no principal ideals contained properly contained between (a, b) and (d), so $(d) = (a, b)$. The second claim follows directly. □

Corollary 6. Let R be a PID.

(1) If elements a and b are relatively prime, then 1 is a linear combination $ra + sb$.
(2) An element of R is irreducible if and only if it is a prime element.
(3) The maximal ideals of R are the principal ideals generated by the irreducible elements.

Proof. (1) This follows from the proposition.

(2) As R is a PID, it is an integral domain and hence any prime element is irreducible (Lemma 2). Now let $q \in R$ be irreducible and suppose $q \mid ab$. Assume $q \nmid a$. We must show $q \mid b$. Let $d = \gcd(q, a)$. As q is irreducible, its divisors are units and its associates. But $q \nmid a$, so d is not an associate of q, whence d is a unit. It follows that d is a unit so a and q are relatively prime, and $1 = ra + sb$ with $r, s \in R$. Thus, $b = rab + sqb$ and q divides both terms on the right side. Thus, $q \mid b$.

(3) Let q be an irreducible element, so its divisors are units and its associates. Therefore, the only principal ideals that contain (q) are (q) itself and (1). Thus, (q) is maximal. Conversely, if an element a has a proper divisor a, then $(b) \subsetneq (a) \subsetneq (1)$, so (b) is not maximal. □

The existence of units other than ±1 complicates stating a more general FTA. However, we need only rephrase in terms of associates. We say a factoring of an element a in an integral domain R is an expression $a = p_1 \cdots p_m$ with each p_i irreducible. This factoring is unique if whenever a is factored in two ways into irreducibles

$$p_1 \cdots p_m = a = q_1 \cdots q_n$$

then $m = n$ and after reordering each q_i is an associate of p_i for each i.

Example. In $\mathbb{Z}[i]$, $(2+i)(2-i) = 5 = (1+2i)(1-2i)$ but this is unique in the sense above because the terms on the left are associates of the terms on the right: $-i(2+i) = 1-2i$ and $i(2-i) = 1+2i$.

Assume an element a in an integral domain R is not irreducible. Then we can write $a = a_1b_1$ where neither a_1 nor b_1 is a unit. This process continues by factoring a_1 and b_1 (if they are not irreducible). We say factoring terminates if this process ends with all irreducibles for all elements of R. An integral domain is a unique factorization domain (UFD) if factoring terminates and the factorization of a into irreducibles is unique in the sense above. By a previous example, $\mathbb{Z}[\sqrt{-5}]$ is an a UFD.

Proposition 7. Let R be an integral domain. Then factoring terminates if and only if R does not contain an infinite strictly increasing chain of principal ideals: $(a_1) \subsetneq (a_2) \subsetneq (a_3) \subsetneq \cdots$.

Lemma 8. Let $I_1 \subset I_2 \subset I_3 \subset \cdots$ be an increasing chain of ideals in a ring R. The union $J = \cup I_n$ is an ideal.

Proof. Exercise. □

Proposition 9. (1) Let R be an integral domain. Suppose that factoring terminates in R. Then R is a UFD if and only if every irreducible element is a prime element.
(2) A PID is UFD.
(3) The rings $\mathbb{Z}$, $\mathbb{Z}[i]$, and $F[x]$ (F a field) are UFDs.

Proof. (1) Let R be a ring in which every irreducible element is prime, and say that a factors in two ways into irreducibles, say $p_1 \cdots p_m = a = q_1 \cdots q_n$, where $m \leq n$. If $n = 1$, then $m = 1$ and $p_1 = q_1$. Suppose $n > 1$. Since p_1 is prime, it divides one of the q_i, say q_1. But q_1 is irreducible and since p_1 is not a unit, q_1 and p_1 are associates, say $p_1 = uq_1$, where u is a unit. Replacing q_2 by $u^{-1}q_2$ we have $p_1 = q_1$. We then cancel p_1 and proceed by induction.

Conversely, suppose that there is an irreducible element p that is not prime. Then there are elements a and b such that p divides the product $r = ab$, say $r = pc$, but P does not divide a or b. We factor a, b, and c into irreducibles and obtain inequivalent factorizations of r.

(2) Let R be a PID. Every irreducible element of R is prime and so we need only prove that factoring terminates, or, equivalently, that there exists no infinite strictly increasing chain of principal ideals. Suppose $(a_1) \subset (a_2) \subset (a_3) \subset \cdots$ is an increasing chain (not necessarily strict) and let J be the union of these ideals. As R is a PID, $J = (b)$. But then (b) is contained in one of these ideals, say $b \in (a_n)$. Then $(b) \subset (a_n)$ but $(a_n) \subset (a_{n+1}) \subset (b)$, so $(b) = (a_n) = (a_{n+1})$. Thus, the chain is not strictly increasing.

(3) This follows from (2) because all of these rings are Euclidean domains. □

Disclaimer/Warning: The converse of (2) fails (e.g., $\mathbb{Z}[x]$ is a UFD but *not* a PID).

The irreducible polynomials in $\mathbb{C}[x]$ are linear. Thus, as a consequence of the above result, every polynomial $f(x) \in \mathbb{C}[x]$ has a factorization of the form

$$f(x) = (x - \alpha_1) \cdots (x - \alpha_n),$$

where the α_i are the roots of $f(x)$.

3. Gauss's Lemma

We now proceed to studying factorization in $\mathbb{Z}[x]$ and in the process prove that this ring is a UFD. For a prime p, let $\psi_p : \mathbb{Z}[x] \to \mathbb{F}_p[x]$ denote reduction of coefficients mod p. That is, if $f(x) \in \mathbb{Z}[x]$ is written $f(x) = a_n x^n + \cdots + a_1 x + a_0$, then $\psi_p(x) = \overline{a_n} x^n + \cdots + \overline{a_1} x + \overline{a_0}$ where $\overline{a_i}$ is the residue of a_i mod p. It is left as an exercise to show that ψ_p is a homomorphism.

Lemma 10. Let $f(x) = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$, and let $p \in \mathbb{Z}$ be prime. The following are equivalent.

(1) p divides every coefficient a_i of f in $\mathbb{Z}$,
(2) p divides f in $\mathbb{Z}[x]$,
(3) f is in the kernel of ψ_p.

Proof. Easy exercise. □

A polynomial $f(x) = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{Q}[x]$ is called primitive if $f(x) \in \mathbb{Z}[x]$, the gcd of its coefficients is 1, and its leading coefficient a_n is positive.

Lemma 11. Let $f \in \mathbb{Z}[x]$ have positive degree with positive leading coefficient. The following are equivalent.

(1) f is primitive.

(2) f is not divisible by any integer prime p,
(3) for every integer prime p, $\psi_p(f) \neq 0$.

Proposition 12. (1) An integer is a prime element of $\mathbb{Z}[x]$ if and only if it is a prime integer. So a prime integer p divides a product fg of integer polynomials if and only if p divides f or p divides g.
(2) (Gauss's Lemma) The product of primitive polynomials is primitive.

Proof. (1) It is clear that a prime (hence irreducible) element of $\mathbb{Z}[x]$ that is an integer must be a prime integer. Let p be a prime integer and denote $\psi_p(f)$ by $\overline{f}$. The p divides fg if and only if $\overline{fg} = 0$. Since $\mathbb{F}_p[x]$ is a domain, then this holds if and only if $\overline{f} = 0$ or $\overline{g} = 0$ if and only if p divides f or p divides g.

(2) Suppose f and g are primitive polynomials. Then their leading coefficients are positive and hence so is the leading coefficient of fg. Moreover, no prime divides f or g and so by (1), no prime divides fg. Thus, fg is primitive. □

Lemma 13. Every polynomial $f(x) \in \mathbb{Q}[x]$ of positive degree can be written uniquely as a product $f(x) = cf_0(x)$ where $c \in \mathbb{Q}$ and $f_0(x)$ is a primitive polynomial. Moreover, $c \in \mathbb{Z}$ if and only if $f \in \mathbb{Z}[x]$, in which case the gcd of the coefficients of f is $\pm c$.

Proof. Multiply through by an integer, say d to clear denominators and set $f_1 = df \in \mathbb{Z}[x]$. Set c to be the gcd of the coefficients of f_1 and factor out and (after possibly adjusting sign) we have $f = cf_0$ where f_0 is primitive.

For uniqueness, suppose that $cf_0 = f = c'f_0'$ with $c, c' \in \mathbb{Q}$ and f_0, f_0' primitive polynomials. After clearing denominators and adjusting signs we may assume that c, c' are positive integers. If $c \neq 1$, then choose a prime integer p such that $p \mid c$. Then $p \mid c'f_0'$ and so $p \mid c'$ or $p \mid f_0'$. But f_0' is primitive so $p \mid c'$ and we may now cancel p from both sides. We proceed by induction to find that $c = 1$, so $c' = 1$. Thus, $f_0 = f_0'$ and the decomposition is unique. □

Theorem 14. (1) Let f_0 be a primitive polynomial and let $g \in \mathbb{Z}[x]$. If f_0 divides g in $\mathbb{Q}[x]$, then f_0 dives g in $\mathbb{Z}[x]$.
(2) If two polynomials $f, g \in \mathbb{Z}[x]$ have a common nonconstant factor in $\mathbb{Q}[x]$, they have a common nonconstant factor in $\mathbb{Z}[x]$.

Proof. We prove the first part. The second follows directly. Suppose $g = f_0q$ with $q \in \mathbb{Q}[x]$. We claim $q \in \mathbb{Z}[x]$. Write $g = cg_0$ and $q = c'q_0$ with g_0, q_0 primitive. Then $cg_0 = c'f_0q_0$. By Gauss's Lemma, f_0g_0 is primitive so by the uniqueness part of the previous lemma, $c = c'$ and $g_0 = f_0q_0$. But g is an integer polynomial and c is an integer, so $q = cq_0 \in \mathbb{Z}[x]$. □

Proposition 15. (1) Let $f \in \mathbb{Z}[x]$ with positive leading coefficient. Then f is an irreducible element of $\mathbb{Z}[x]$ if and only if it is either a prime integer or a primitive polynomial that is irreducible in $\mathbb{Q}[x]$.

(2) Every irreducible element of $\mathbb{Z}[x]$ is a prime element.

Proof. (1) We have already proved this for constants, so assume f is non-constant, whence f is primitive. Suppose f has a proper factorization in $\mathbb{Q}[x]$, say $f = gh$. Write $g = cg_0$ and $h = c'h_0$ with g_0, h_0 primitive. Then g_0h_0 is primitive and since f is primitive, $f = g_0h_0$. Thus, f has a proper factorization in $\mathbb{Z}[x]$. The converse is clear.

(2) Let f be a primitive irreducible polynomial that divides a product gh of integer polynomials. Then f is irreducible in $\mathbb{Q}[x]$. Since $\mathbb{Q}[x]$ is a PID, f is a prime element of $\mathbb{Q}[x]$, so f divides g or h. Thus, f divides g or h in $\mathbb{Z}[x]$, so f is a prime element. □

The next theorem follows directly from the above results.

Theorem 16. The polynomial ring $\mathbb{Z}[x]$ is a UFD.

We don't attempt a proof of the following theorem, but it is worth noting.

Theorem 17. Let R be a UFD, then the polynomial ring $R[x_1, \dots, x_n]$ is a UFD.

4. Factoring integer polynomials

Let $f(x) \in \mathbb{Z}[x]$ and write

$$(*) \qquad f(x) = a_nx^n + \cdots + a_1x + a_0$$

with $a_n \neq 0$. In this section we will study methods for factoring such polynomials.

Lemma 18. Let $f \in ZZ[x]$ written as in $(*)$.

(1) If $b_1x + b_0 \in \mathbb{Z}[x]$ divides f in $\mathbb{Z}[x]$, then b_1 divides a_n and b_0 divides a_0.

(2) A primitive polynomial $b_1x + b_0 \in \mathbb{Z}[x]$ divides f in $\mathbb{Z}[x]$ fi and only if $-b_0/b_1$ is a root of f.

(3) A rational root of a monic integer polynomial f is an integer.

Proof. (1) Suppose $f = (b_1x + b_0)(q_{n-1}x^{n-1} + \cdots + q_0)$. Then $a_n = b_1q_{n-1}$ an $a_0 = b_0q_0$.

(2) By a theorem from the last section, $b_1x + b_0$ divides f in $\mathbb{Z}[x]$ if and only if it divides f in $\mathbb{Q}[x]$. This is true if and only if $x + (b_0/b_1)$ divides f, i.e., $-b_0/b_1$ is a root.

(3) If $\alpha = a/b$ is a root of f with $b > 0$, and if $\gcd(a, b) = 1$, then $bx - a$ is a primitive polynomial that divides the monic polynomial f, so $b = 1$ and α is an integer. □

To find non-linear factors, we utilize the homomorphism $\psi_p : \mathbb{Z}[x] \to \mathbb{F}_p$ that reduces coefficients modulo a prime p. Recall that, to simplify notation, we denote the image of f under ψ_p by $\overline{f}$.

Proposition 19. Let $f(x) = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$, $a_n \neq 0$, and let $p \in \mathbb{Z}$ be a prime such that $p \nmid a_n$. If $\overline{f}$ is irreducible in $\mathbb{F}_p[x]$, then f is irreducible in $\mathbb{Q}[x]$.

Proof. We prove the contrapositive. Suppose $f = gh$ is a proper factorization of f in $\mathbb{Q}[x]$, and so we may assume that $g, h \in \mathbb{Z}[x]$. Thus, $\deg(g), \deg(h) > 0$ and $\deg(f) = \deg(g) + \deg(h)$. As ψ_p is a homomorphism, $\overline{f} = \overline{g}\overline{h}$, so $\deg(\overline{f}) = \deg(\overline{g}) + \deg(\overline{h})$. Note that, for any $p \in \mathbb{Z}[x]$, $\deg(\overline{p}) \leq \deg(p)$. But by our hypothesis on the leading coefficient, $\deg(\overline{f}) = \deg(f)$. Hence, we must have the same for g and h, so the factorization is proper in $\mathbb{F}_p[x]$. □

Disclaimer/Warning: The converse of the previous proposition is false. For *every* prime p, the polynomial $x^4 - 10x + 1$ factors. However, it is irreducible in $\mathbb{Z}[x]$.

The advantage of using this method is that there are only finitely many irreducible polynomials in $\mathbb{F}_p[x]$ in each degree. Recall the *sieve method* for finding primes in $\mathbb{Z}$. A similar method works in $\mathbb{F}_p[x]$.

Example. We find the irreducible polynomials in $\mathbb{F}_2[x]$ of degree ≤ 3.

In degree 1, the polynomials are x and $x + 1$. These are both irreducible.

In degree 2, the polynomials are x^2, $x^2 + x$, $x^2 + 1$, and $x^2 + x + 1$. The first three have roots in $\mathbb{F}_2$ and so are divisible by x or $x + 1$. The polynomial $x^2 + x + 1$ is the only irreducible polynomial in $\mathbb{F}_2[x]$ of degree 2.

In degree 3, the polynomials are

$$x^3, \quad x^3 + x^2, \quad x^3 + x, \quad x^3 + 1, \quad x^3 + x^2 + x, \quad x^3 + x^2 + 1, \quad x^3 + x + 1, \quad x^3 + x^2 + x + 1.$$

Now zero is a root of

$$x^3, \quad x^3 + x^2, \quad x^3 + x, \quad x^3 + x^2 + x.$$

Of the remaining polynomials, 1 is a root of

$$x^3 + 1, \quad x^3 + x^2 + x + 1.$$

This leaves only the polynomials $x^3 + x^2 + 1$ and $x^3 + x^2 + x + 1$ as irreducible polynomials of degree 3 in $\mathbb{F}_2[x]$.

Example. Consider the polynomial $f(x) = x^5 + x^3 + 1$ in $\mathbb{F}_2[x]$. Note that f does not have a linear factor because 0 and 1 are not roots. Thus, if $f(x)$ had a factor (in $\mathbb{F}_2[x]$), then it has a quadratic factor $p(x)$. By the previous example, $p(x) = x^2 + x + 1$. Performing division we find,

$$f(x) = p(x)(x^3 + x^2 + x) + (x + 1).$$

Thus, p does not divide f, so f is irreducible in $\mathbb{F}_2[x]$.

Example. The polynomial $x^5 - 64x^4 + 127x^3 - 200x + 99$ is irreducible in $\mathbb{Q}[x]$ because its residue in $\mathbb{F}_2[x]$ is the irreducible polynomial $f(x)$ from the previous example.

Proposition 20 (Eisenstein Criterion)**.** Let $f(x) = a_n x^n + \cdots + a_0 \in \mathbb{Z}[x]$, $a_n \neq 0$, and let $p \in \mathbb{Z}$ be a prime. Suppose that the coefficients of f satisfy the following conditions:

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

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

Proof. Assume f satisfies the conditions and let $\overline{f} = \psi_p(f)$ as usual. By hypothesis, $\overline{f} = \overline{a_n} x^n$ such that $\overline{a_n} \neq 0$. If f is reducible in $\mathbb{Q}[x]$, it will factor in $\mathbb{Z}[x]$ into factors of positive degree, say $f = gh$, where $g = b_r x^r + \cdots + b_0$ and $h = c_s x^s + \cdots + c_0$. Then $\overline{g}$ divides $\overline{a_n} x^n$, so $\overline{g} = \overline{b_r} x^r$. Every coefficient of g except the leading coefficient is divisible by p. The same is true for h. The constant coefficient of f will be equal to $b_0 c_0$, and since p divides b_0 and c_0, p^2 must divide a_0. This contradicts the third condition. Therefore, f is irreducible. □

Example. Let $f(x) = 16x^5 - 9x^4 + 3x^2 + 6x - 21$ and set $p = 3$. Then clearly p divides all coefficients except for that of x^5 and $p^2 \nmid 21$. Thus, $f(x)$ is irreducible over $\mathbb{Q}$ by Eisenstein's Criterion.

Example. Let p be a prime. Consider the cyclotomic polynomial

$$\phi(x) = x^{p-1} + x^{p-2} + \cdots + x + 1.$$

Its roots are the pth roots of unity, that is, $(e^{2\pi i/p})^k$ for $1 \leq k \leq p-1$. Note that $(x-1)\phi(x) = x^p - 1$. Making the substitution $x = y + 1$ and expanding we find

$$y\phi(y+1) = (y+1)^p - 1 = \left(\sum_{i=0}^{p} \binom{p}{i} y^i \right) - 1 = y \left(\sum_{i=1}^{p} \binom{p}{i} y^{i-1} \right).$$

Cancelling y from both sides shows that

$$\phi(y+1) = \sum_{i=1}^{p} \binom{p}{i} y^{i-1}.$$

The coefficient of y^{p-1} is $\binom{p}{p} = 1$ and hence not divisible by p. On the other hand, every other coefficient is divisible by p. Moreover, the coefficient of 1 is $\binom{p}{p-1} = p$, whence not divisible by p^2. It follows from Eisenstein's Criterion that $\phi(y+1) = \phi(x)$ is irreducible.

Fields

(Last Updated: November 16, 2018)

These notes are derived primarily from *Algebra* by Michael Artin (2ed), though some material is drawn from *Abstract Algebra, Theory and Applications* by Thomas Judson (16ed).

1. Examples of fields

Given a pair of fields $F \subset K$, we say that K is a field extension of F (or an extension field). We denote this relationship by K/F.

Before we begin, we will review the important classes of fields that we will study.

- A number field is a subfield of $\mathbb{C}$. We will focus primarily on algebraic number fields in which all elements are algebraic numbers[1].
- A finite field is a field that contains finitely many elements. Any such field is an extension field of $\mathbb{F}_p$ for some prime p.
- Extensions of the field $\mathbb{C}(t)$ are called function fields.

2. Algebraic and Transcendental elements

Let K be an extension field of some field F and let $\alpha \in K$. We say that α is algebraic over F it is a root of a monic polynomial $f(x) \in F[x]$. An element is transcendental over F if it is not algebraic over F. One can also phrase these definitions in terms of the substitution homomorphism $\phi : F[x] \to K$ given by $x \mapsto a$. The element $a \in K$ is transcendental if ϕ is injective and it is algebraic otherwise. Recall that $F[x]$ is a PID and so $\ker \phi$ is a principal ideal, generated by a monic polynomial $f(x) \in F[x]$.

Proposition 1. Let α be an element of an extension field K of a field F that is algebraic over F. Let $f(x) \in F[x]$. The following are equivalent.

(1) f is the monic polynomial of lowest degree in $F[x]$ that has α as a root.

(2) f is an irreducible element of $F[x]$, and α is a root of f.

(3) f has coefficients in F, α is a root of f, and the principal ideal of $F[x]$ that is generated by f is a maximal ideal.

(4) α is a root of f, and if $g \in F[x]$ is any polynomial that has α as a root, then f divides g.

[1] Recall that a number $\alpha \in R$ is algebraic if it satisfies a polynomial equation $f(\alpha) = 0$ for some $f(x) \in R[x]$.

We call the unique monic polynomial that satisfies any of the above equivalent conditions the irreducible polynomial for α over F. The degree of this polynomial is called the degree of α over F. Of course, both of these points (irreducibility and degree) depend heavily on the field F itself.

Let K be an extension field of F and let $\alpha \in K$. We denote by $F(\alpha)$ the subfield of K generated by F and α. That is, $F(\alpha)$ is the smallest subfield of K that contains F and α. Obviously, if $\alpha \in F$, then $F(\alpha) = F$. On the other hand, $\mathbb{C}$ is an extension field of $\mathbb{R}$ and $\mathbb{R}(i) = \mathbb{C}$. This generalizes to adding several elements: $F(\alpha_1, \ldots, \alpha_k)$ is the smallest of subfield of K containing F and the α_i.

We need to be careful with notation. The *ring* generated by α over F is still denoted $F[\alpha]$, which is the image of the homomorphism $\phi : F[x] \to K$ that sends $x \mapsto \alpha$. The field $F(\alpha)$ is isomorphic to the field of fractions of $F[\alpha]$. If α is transcendental, the map $F[x] \to F[\alpha]$ is an isomorphism and so $F(\alpha)$ is isomorphic to $F(x)$.

Proposition 2. Let α be element of an extension field K/F that is algebraic over F, and let f be the irreducible polynomial for α over F. The canonical map $F[x]/(f) \to F[\alpha]$ is an isomorphism, and $F[\alpha]$ is a field. Thus, $F[\alpha] = F(\alpha)$.

Proof. Let $\phi : F[x] \to K$ be the substitution homomorphism sending $x \mapsto \alpha$. The ideal (f) is maximal, $f(x)$ generates the kernel, and $F[x]/(f)$ is isomorphic to $\operatorname{im} \phi = F[\alpha]$. Moreover, $F[x]/(f)$ is a field, so $F[\alpha]$ is a field. Since $F(\alpha)$ is the fraction field of $F[\alpha]$, then $F[\alpha] = F(\alpha)$. □

Now if $\alpha_1, \ldots, \alpha_k$ are algebraic elements of an extension field K/F, then $F[\alpha_1, \ldots, \alpha_k] = F(\alpha_1, \ldots, \alpha_k)$.

Proposition 3. Let α be an algebraic element over F, and let $f(x)$ be the irreducible polynomial for α over F. If $\deg f(x) = n$, then $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is a basis for $F(\alpha)$ as a vector space over F.

Proof. This is a special case of a proposition in the ring case. We recall the proof here briefly. Note that $F(\alpha) = F[\alpha]$ is a quotient of $F[x]$ and so every element β of $F[\alpha]$ is a residue of a polynomial $g(x) \in F[x]$. Since f is monic, we can perform division with remainder to get $g(x) = f(x)q(x)+r(x)$ so $\beta = r(\alpha)$ and $\deg r < \deg f = n$. □

Example. Let $\omega = e^{2\pi i/3}$. The minimal polynomial of ω over $\mathbb{Q}$ is $x^2 + x + 1$. Since the degree of this polynomial is 2, then $\{1, \omega\}$ is a basis for $Q(\omega)$ over Q.

Proposition 4. Let F be a field, and let α, β be elements of field extensions K/F and L/F, respectively. Suppose that α and β are algebraic over F. There is an isomorphism of field extensions $\sigma : F(\alpha) \to F(\beta)$ that is the identity on F and that sends $\alpha \mapsto \beta$ if and only if the irreducible polynomials for α and β are equal.

Proof. Since α and β are algebraic over F, $F[\alpha] = F(\alpha)$ and $F[\beta] = F(\beta)$. Suppose f is the irreducible polynomial of both α and β. Hence, there are isomorphisms $\phi : F[x]/(f) \to F[\alpha]$ and $\psi : F[x]/(f) \to F[\beta]$ and so $\sigma = \psi\phi^{-1}$ is the required isomorphism $F(\alpha) \to F(\beta)$.

Conversely, suppose σ is an isomorphism that is the identity on F and sends $\alpha \mapsto \beta$. Let f be polynomial with coefficients in F such that $f(\alpha) = 0$. Then also $f(\beta) = 0$. Hence, the irreducible polynomials of α and β are equal. □

Definition. Let K and K' be extensions of a field F. An isomorphism $\phi : K \to K'$ that restricts to the identity on F is called an F-isomorphism (or an isomorphism of field extensions). In this case, we say K and K are isomorphic field extensions.

Let $f \in \mathbb{R}[x]$ and let $\alpha \in \mathbb{C}$ be a root of f. Then a standard result says that $\overline{\alpha}$ is also a root of f. If $\alpha \in \mathbb{R}$ then there is nothing to prove, but if $\alpha \notin \mathbb{R}$, then one need only observe that $\overline{\alpha}$ is the unique complex number such that $f(x) = (x - \alpha)(x - \overline{\alpha}) \in \mathbb{R}[x]$. Note that $f(x)$ is the irreducible polynomial of α over $\mathbb{R}$. The next proposition generalizes this idea.

Proposition 5. Let $\phi : K \to K'$ be an isomorphism of field extensions of F, and let f be a polynomial with coefficients in F. Let α be a root of f in K, and let $\beta = \phi(\alpha)$ be its image in K'. Then β is also a root of f.

Proof. Write $f(x) = a_n x^n + \cdots + a_1 x + a_0$. Since ϕ is an F-isomorphism and since the $a_i \in F$, then $\phi(a_i) = a_i$. Since ϕ is a homomorphism,

$$\begin{aligned} 0 = \phi(0) = \phi(f(\alpha)) &= \phi(a_n \alpha^n + \cdots + a_1 \alpha + a_0) \\ &= \phi(a_n)\phi(\alpha^n) + \cdots + \phi(a_1)\phi(\alpha) + \phi(a_0) \\ &= a_n \beta^n + \cdots + a_1 \beta + a_0 = f(\beta). \end{aligned}$$

Thus, β is a root of f. □

3. The degree of a field extension

Let K be a field extension of F. Then K is an F-vector space and we denote by $[K : F]$ the dimension of K over F when regarded as such. We call $[K : F]$ the degree of the field extension. We say a field extension is finite if $[K : F] < \infty$. A quadratic extension is an extension of degree two, and a cubic extension is an extension of degree three.

For example, since $\mathbb{C} = \mathbb{R}(i)$, then a basis for $\mathbb{C}$ over $\mathbb{R}$ is $\{1, i\}$, so $[\mathbb{C} : \mathbb{R}] = 2$.

Lemma 6. (1) A field extension K/F has degree 1 if and only if $F = K$.
(2) An element α of a field extension K has degree 1 over F if and only if α is an element of F.

Proof. (1) If K/F has degree 1, then a basis for K over F is $\{1\}$. Thus, $K = F$.

(2) The degree of α over F is the degree of the monic irreducible polynomial for α over F. If α has degree 1, this polynomial is $x - \alpha$. But then if $x - \alpha$ has coefficients in F, then $\alpha \in F$. □

The characteristic of a field F is the smallest nonzero integer n such that $1+1+\cdots+1=0$ (add 1 n times), if such an n exists. If no such n exists, we say that that the characteristic is 0. We denote this by $\operatorname{char} F$.

Proposition 7. Let F be a field with $\operatorname{char} F \neq 2$. Then K is a quadratic extension of F if and only if $K = F(\delta)$ where $\delta^2 = d$ for some $d \in F$, $\delta \notin F$.

Proof. Suppose K is a quadratic extension. Then $\{1, \alpha\}$ is a basis of K over F for some α not in F. It follows that α^2 is a linear combination of $\{1, \alpha\}$ with coefficients in F. That is, $\alpha^2 = b\alpha + c$ for some $b, c \in F$. Then α is a root of $f(x) = x^2 - bx - c$. But since $\alpha \notin F$, then this is the monic irreducible polynomial of α over F.

The discriminant of f is $D = b^2 + 4c$. Since $\operatorname{char} F \neq 2$, $\frac{1}{2}(b + \sqrt{D})$ is a solution to $x^2 - bx - c = 0$. There are two choices for $\sqrt{D}$ and so let δ be one of them. Then $\delta \in K$, $\delta^2 \in F$, and and because $\alpha \in F(\delta)$, then δ generates K over F.

Conversely, if $\delta^2 \in F$ but $\delta \notin F$, then $\{1, \delta\}$ is a basis for $F(\delta)$ over F. Thus, $[F(\delta) : F] = 2$. □

Disclaimer/Warning: This proposition *does not* extend to cubic extensions.

However, this proposition does illustrate an important connection between degree of an extension and degree of an irreducible polynomial. In fact, this is why the term degree is used for extension rather than dimension. The next result follows from Proposition 3.

Proposition 8. Let α be an element of an extension field over F. Then $[F(\alpha) : F]$ is finite if and only if α is algebraic. Moreover, if α is algebraic, then the degree of $[F(\alpha) : F]$ is equal to the degree of α over F.

Theorem 9 (Multiplicative property of the degree)**.** Let $F \subset K \subset L$ be fields (could write $L/K/F$). Then $[L : F] = [L : K][K : F]$.

Proof. We consider the case when $[L : K]$ and $[K : F]$ are finite. The infinite case is similar.

Let $U = \{u_1, \ldots, u_n\}$ be a basis for L as a K-vector space and let $V = \{v_1, \ldots, v_m\}$ be a basis for K as an F-vector space. We claim that the mn products $W = \{u_i v_j\}$ for a basis for L as a F-vector space.

Let $\gamma \in L$. Then we can write $\gamma = a_1 u_1 + \ldots + a_n u_n$ with $a_i \in K$. But then each a_i can be written as $a_i = b_{i1} v_1 + \ldots + b_{im} v_m$. Hence, $\gamma = \sum b_{ij} u_i v_j$. Thus, W spans L. Now suppose $\sum c_{ij} u_i v_j = 0$. Because V is a basis for K over V, this implies that $\sum c_{ij} u_i = 0$ for each i. By the same logic, since U is a basis for L over K, then $c_{ij} = 0$ for each i, j. Thus, W is linearly independent. □

An immediate consequence of the multiplicative property is that $[L : K]$ and $[K : F]$ divide $[L : F]$.

Corollary 10. (1) Let $F \subset K$ be a finite field extension of degree n, and let α be an element of K. Then α is algebraic over F and its degree over F divides n.

(2) Let $F \subset F' \subset L$ be fields. If an element α of L is algebraic over F, it is algebraic over F'. If α has degree d over F, its degree over F' is at most d.

(3) A field extension K that is generated over F by finitely many algebraic elements is a finite extension. A finite extension is generated by finitely many elements.

(4) If K is an extension field of F, the set of elements of K that are algebraic over F is a subfield of K.

Proof. (1) We have $F \subset F(\alpha) \subset K$. Thus, $[K : F] = [K : F(\alpha)][F(\alpha) : F]$ by the multiplicative property. If $[K : F]$ is finite, then so is $[F(\alpha) : F]$ and this is the degree of α over F. Hence, this degree divides n.

(2) Let f denote the irreducible polynomial for α over F. Since $F \subset F'$, then f is also an element of $F'[x]$. As α is a root of f, the irreducible polynomial g for α over F' divides f, so the degree of g is at most equal to the degree of f.

(3) Let $\alpha_1, \ldots, \alpha_k$ be elements that generate K and are algebraic over F. Let $F_i = F(\alpha_1, \ldots, \alpha_i)$ for $i = 1, \ldots, k$. These fields form a chain $F = F_0 \subset F_1 \subset \cdots \subset F_k = K$. Since α_i is algebraic over F, it is also algebraic over F_{i-1}. Therefore, $[F_i : F_{i-1}]$ is finite for every i. By the multiplicative property, $[K : F]$ is finite.

(4) Let $\alpha, \beta \in K$ be algebraic over F. By (1) and (2), $\alpha + \beta$, $\alpha\beta$, and α/β are algebraic over F because they are elements of the field extension $F(\alpha, \beta)$. □

Corollary 11. Let K be an extension field of F of prime degree p. If an element α of K is not in F, then $K = F(\alpha)$.

Corollary 12. Let $\mathcal{K}$ be an extension field of F, let K and F' be subfields of $\mathcal{K}$ that are finite extensions of F, and let K denote the subfield of $\mathcal{K}$ generated by K and F'. Let $[K' : F] = N$, $[K : F] = m$, and $[F' : F] = n$. Then m and n divides N and $N \leq nm$.

It follows that N is divisible by the lcm of m and n. If m and n are relatively prime, then $N = mn$.

Example. (1) Let $\alpha = \sqrt[3]{2}$ and $\omega = e^{2\pi i/3}$. The three complex roots of $x^3 - 2$ are $\alpha_1 = \alpha$, $\alpha_2 = \omega\alpha$, and $\alpha_3 = \omega^2\alpha$. Each root has degree 3 over $\mathbb{Q}$. But $\mathbb{Q}(\alpha_1, \alpha_2) = \mathbb{Q}(\alpha_1, \omega)$.

(2) Let $\alpha = \sqrt[3]{2}$ and let β be a root of the irreducible polynomial $x^4 + x + 1$ over $\mathbb{Q}$. Because 3 and 4 are relatively prime, $\mathbb{Q}(\alpha, \beta)$ has degree 12 over $\mathbb{Q}$ and $\alpha \notin \mathbb{Q}(\beta)$.

(3) Let $K = \mathbb{Q}(\sqrt{2}, i)$. Both i and $\sqrt{2}$ have degree 2 over $\mathbb{Q}$ and since i is imaginary, $i \notin \mathbb{Q}(\sqrt{2})$. Thus, $[\mathbb{Q}(\sqrt{2}, i) : \mathbb{Q}] = 4$. Therefore, the degree of i over $\mathbb{Q}(\sqrt{2})$ is 2. Note that $\sqrt{-2}$ and i also generate K, and so $i \notin \mathbb{Q}(\sqrt{-2})$.

4. Finding the irreducible polynomial

Let γ be an algebraic element over a field F. We consider two methods for finding the irreducible polynomial of γ. The first is to compute powers of γ and look for relations between the powers.

Example. Let $\gamma = \sqrt{2} + \sqrt{3}$. Then $\gamma^2 = 5 + 2\sqrt{6}$ and $\gamma^4 = 49 + 20\sqrt{60}$. Hence, $\gamma^4 - 10\gamma^2 + 1 = 0$. Thus, γ is a root of the polynomial $g(x) = x^4 - 10x^2 + 1$.

We will show that g is irreducible over $\mathbb{Q}$. The only possible rational roots of g would be ± 1 (Rational Root Theorem). Clearly these are not roots so, were g to factor, it would factor as integer quadratics, say

$$x^4 - 10x^2 + 1 = (x^2 + ax + b)(x^2 + cx + d) = x^4 + (a + c)x^3 + (ac + b + d)x^2 + (ad + bc)x + bd.$$

Matching coefficients, we have $bd = 1$, so $d = b^{-1}$. Since d and b are integers, then $d = b = \pm 1$. We also have $a + c = 0$, so $c = -a$. Finally, $10 = ac + b + d = -a^2 \pm 2$, so $-a^2 = 8$ or 12, which is absurd.

Lemma 13. (1) A linear dependence relation $c_n\gamma^n + \cdots + c_1\gamma + c_0 = 0$ among powers of an element γ means that γ is a root of the polynomial $c_nx^n + \cdots + c_1x + c_0 = 0$.

(2) Let α and β be algebraic elements of an extension field of F, and let their degrees over F be d_1 and d_2, respectively. The d_1d_2 monomial $\alpha^i\beta^j$, with $0 \le i \le d_1$ and $0 \le j \le d_2$, span $F(\alpha, \beta)$ as an F-vector space.

Another method is to guess[2] at the other roots of the irreducible polynomial.

Example. Consider $\gamma = \sqrt{2} + \sqrt{3}$. We guess that the roots of the irreducible polynomial of γ over $\mathbb{Q}$ are $\gamma_1 = \gamma$, $\gamma_2 = -\gamma$, $\gamma_3 = -\sqrt{2} + \sqrt{3}$, and $\gamma_4 = -\gamma_3$. Then we can check that

$$(x - \gamma_1)(x - \gamma_2)(x - \gamma_3)(x - \gamma_4) = x^4 - 10x + 1.$$

Another way to check whether the polynomial g is irreducible is to determine a basis of K over $\mathbb{Q}$. That is, we want to make sure that we have not introduced additional linear dependences.

Example. Consider the examples above. Set $\alpha = \sqrt{2}$ and $\beta = \sqrt{3}$. Then $d_1 = d_2 = 2$. The elements $\alpha^i\beta^j$ with $0 \le i, j \le 2$ are $1, \sqrt{2}, \sqrt{3}, \sqrt{6}$. These elements form a basis of $K = \mathbb{Q}(\alpha, \beta)$. Hence, the polynomial $g(x) = x^4 - 10x + 1$ is irreducible.

Example. Let $\alpha = \sqrt[3]{2}$ and $\omega = e^{2\pi i/3}$. The three complex roots of $f(x) = x^3 - 2$ are $\alpha_1 = \alpha$, $\alpha_2 = \omega\alpha$, and $\alpha_3 = \omega^2\alpha$. Each root has degree 3 over $\mathbb{Q}$. Set $F = \mathbb{Q}$, $L = \mathbb{Q}(\alpha_1)$, and $K = \mathbb{Q}(\alpha_1, \alpha_2)$. The nine monomials $\alpha_1^i\alpha_2^j$ with $0 \le i, j \le 3$ span K over F. However, f has a root α_1 in L, so in L we have $f(x) = (x - \alpha_1)q(x)$. Then α_2 has degree at most 2 over L. The set $\{1, \alpha_2\}$ is a basis of K over L, and so the six monomials $\alpha_1^i\alpha_2^j$ with $0 \le i \le 3$ and $0 \le j \le 1$ form a basis for K over F.

[2] It's sort of educated guessing, really.

6. Adjoining roots

We have previously seen the process of adjoining elements to a ring. We do not recall that procedure here but note how it can be applied to the problem of finding extension fields where roots of polynomials exist. This is especially useful when we are not working in subfields of $\mathbb{C}$.

Recall that, for a field F, $f \in F[x]$ is irreducible if and only if (f) is a maximal ideal in $F[x]$ if and only if $F[x]/(f)$ is a field.

Lemma 14. Let F be a field, and let f be an irreducible polynomial in $F[x]$. Then the ring $K = F[x]/(f)$ is an extension field of F, and the residue $\overline{x}$ of x is a root of $f(x)$ in K.

Proof. The ring K is a field because (f) is a maximal ideal in $F[x]$. Moreover, the homomorphism $F \to K$ that sends elements of F to the residues of the constant polynomials is injective because F is a field. Hence, we may identify F with its image in K. (That is, F is isomorphic to the subfield of constant polynomials in K.) In this way, K becomes an extension field of F. Since $\overline{x}$ satisfies the equation $f(\overline{x}) = 0$, it is a root of f. □

A polynomial f splits completely in a field K if it factors into linear factors in K. There is also a notion of a *splitting field* but it requires more.

Proposition 15. Let F be a field, and let $f(x)$ be a monic polynomial in $F[x]$ of positive degree. There exists a field extension K of F such that $f(x)$ splits completely in K.

Proof. We induct on the degree of f. Suppose f has a root α in F, so that $f(x) = (x - \alpha)q(x)$ in $F[x]$. Then we may replace f by q and we are done by induction. Now let g be an irreducible factor of f. By the lemma, there is a field extension F_1 of F in which g has a root of α. Then α is a root of f too. Replace F by F_1 and repeat. □

The next proposition generalizes some of the story between $\mathbb{Z}[x]$ and $\mathbb{Q}[x]$ that we learned previously, though we note that $\mathbb{Z}[x]$ is not a field so the results we state don't actually descend.

Proposition 16. Let F be a field and $f, g \in F[x]$, $f \neq 0$, and let K be an extension field of F.

(1) $F[x]$ is a subring of $K[x]$.
(2) Division with remainder of g by f gives the same answer in $F[x]$ and $K[x]$.
(3) f divides g in $F[x]$ if and only if f divides g in $K[x]$.
(4) The monic gcd of f and g is the same in $F[x]$ and $K[x]$.
(5) If f and g have a common root in K, there are not relative prime in $F[x]$. If f and g are not relatively prime in $F[x]$, there exists an extension field in which they have a common root.
(6) If f is an irreducible element of $F[x]$ and if f and g have a common root in K, then f divides g in $F[x]$.

Write $f(x) = a_n x^n + \cdots + a_1 x + a_0$. The derivative of f is defined as

$$f'(x) = n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \cdots + a_1.$$

Note that, while the rules for differentiation of polynomials in polynomial rings over fields all follow as they would in $\mathbb{R}[x]$, there are subtleties, especially when in positive characteristic. For example, If $f(x) = x^p - 1 \in \mathbb{F}_p[x]$, then $f'(x) = 0$.

Recall that α is a root of a polynomial $f(x)$ if $x - \alpha$ divides f (in some field). We say that α is a multiple root if $(x - \alpha)^2$ divides f.

Lemma 17. Let K be an extension field of a field F and $f \in F[x]$. An element $\alpha \in K$ is a multiple root of f if and only if α is a root of f and f'.

Proof. If α is a root of f, then $x - \alpha$ divides f, say $f(x) = (x - \alpha)g(x)$. Then α is a multiple root of f if and only if it is a root of g. By the product rule,

$$f'(x) = (x - \alpha)g'(x) + g(x).$$

Substituting $x = \alpha$ we have $f'(\alpha) = 0$ if and only if $g(\alpha) = 0$. □

Lemma 18. Let f be a polynomial with coefficients in a field F There exists a field extension K of F in which f has a multiple root if and only if f and f' are not relatively prime.

Proof. If f has a multiple root in K, then f and f' have a common root in K, so they are not relatively prime in K (or in F). Conversely, if f and f' are not relatively prime, then they have a common root in some field extenstion K, hence f has a multiple root there. □

Proposition 19. Let f be an irreducible polynomial in $F[x]$.

(1) f has no multiple root in any field extension of F unless the derivative f' is the zero polynomial.
(2) If F is a field of characteristic zero, then f has no multiple root in any field extension of F.

Proof. (1) Suppose $f' \neq 0$. We claim that f and f' are relatively prime. Since f is irreducible, f will have a nonconstant factor in common with another polynomial g only if f divides g. If f divides g, then unless $g = 0$, the degree of g will be at least as large as the degree of f. But $f' \neq 0$ so its degree is less than the degree of f, and then f and f' have no common nonconstant factor.

(2) The derivative of an irreducible polynomial over a field of characteristic zero is never zero. □

7. Finite fields

At this point, we are fairly familiar with the field $\mathbb{F}_p$, p a prime. While these fields are fundamental in the study of finite fields, they do not constitute all such fields.

Let K be a finite field. Then K has finite characteristic p where p is a prime, so K contains $F = \mathbb{F}_p$. (Otherwise, K would contain zero divisors, which are non-invertible elements.) Moreover, as K is finite, it is finite dimensional over F. Set $[K : F] = r$. As an F-vector space, K is isomorphic to the space of F^r column vectors, which contains p^r elements. Thus, $q = |K| = p^r$. Throughout this section, q denotes a power of a prime integer p. Fields of order q are denoted $\mathbb{F}_q$. We will show that, up to isomorphism, there is only one such field.

Example. Let $F = \mathbb{F}_2$ and let $f = x^2 + x + 1 \in F[x]$. Note that f is irreducible in $F[x]$. We let K be the field obtained by adjoining a root α of f to F, so $K \cong F[x]/(x^2 + x + 1)$. The set $\{1, \alpha\}$ forms a basis of K over F, so $K = \{0, 1, \alpha, 1 + \alpha\}$. Note that $1 + \alpha$ is the other root of f. The field K is (isomorphic to) the field $\mathbb{F}_4$.

Disclaimer/Warning: The field $\mathbb{F}_4$ *is not* isomorphic to $\mathbb{Z}/(4)$, which is not a field.

We won't bother with the proof of the following theorem (see Section 15.7 in the text) because it isn't our primary concern.

Theorem 20. *Let p be a prime integer, and let $q = p^r$ be a positive power of p.*

(1) *Let K be a field of order q. The elements of K are roots of the polynomial $x^q - x$.*
(2) *The irreducible factors of the polynomial $x^q - x$ over the field $F = \mathbb{F}_p$ are the irreducible polynomials in $F[x]$ whose degrees divide r.*
(3) *Let K be a field of order q. The multiplicative group $K^\times$ of nonzero elements of K isa. cyclic group of order $q - 1$.*
(4) *There exists a field of order q, and all fields of order q are isomorphic.*
(5) *A field of order p^r contains a subfield of order p^k if and only if k divides r.*

Corollary 21. *For every positive integer r, there exists an irreducible polynomial of degree r over the prime field $\mathbb{F}_p$.*

Example. The field $\mathbb{F}_4$ has degree 2 over $\mathbb{F}_2$. Its elements are the roots of the polynomial

$$x^4 - x = x(x - 1)(x^2 + x + 1).$$

The roots of x and $x - 1$ are 0 and 1. The other two roots are α and $1 + \alpha$ of $x^2 + x + 1$ that we saw in the previous example.

Example. The field $\mathbb{F}_8$ has degree 3 over the field $\mathbb{F}_2$. Its elements are the eight roots of the polynomial

$$x^8 - x = x(x - 1)(x^3 + x + 1)(x^3 + x^2 + 1).$$

Let β be a root of one of the two irreducible polynomials, say $x^3 + x + 1$. Then $\{1, \beta, \beta^2\}$ is a basis of $\mathbb{F}_8$ over $\mathbb{F}_2$. The elements of $\mathbb{F}_8$ are the eight linear combinations:

$$\{0, 1, \beta, 1 + \beta, \beta^2, 1 + \beta^2, \beta + \beta^2, 1 + \beta + \beta^2\}.$$

One can check that this includes the roots of $x^3 + x^2 + 1$, as well as the other roots of $x^3 + x + 1$. Computations can be done by noting that $1 + 1 = 0$ and $\beta^3 + \beta + 1 = 0$.

Note that $[\mathbb{F}_8 : \mathbb{F}_2] = 3$ and $[\mathbb{F}_4 : \mathbb{F}_2] = 2$. Since 2 does not divide 3, $\mathbb{F}_4$ is not a subfield of $\mathbb{F}_8$.

Lemma 22. Let p be a prime and let $q = p^r$ be a positive power of p. Let L be a field of characteristic p, and let K be the set of roots of $x^q - x$ in L. Then K is a subfield of L.

Let K and K' be two fields of the same order $q = p^r$. We claim that they are isomorphic. Let α be a generator for the cyclic group $K^\times$. Then $K = F(\alpha)$, so the irreducible polynomial f for α over F has degree equal to $[K : F] = r$. Then f generates the ideal of polynomials in $F[x]$ with root α. Since α is also a root of $x^q - x$, f divides $x^q - x$. But $x^q - x$ splits completely in K', so f has a root α' in K' as well. Thus, $F(\alpha)$ and $F(\alpha')$ are both isomorphic to $F[x]/(f)$. Counting degrees shows that $F(\alpha') = K'$, so K and K' are isomorphic.

8. Primitive elements

Let K be a field extension of a field F. An element $\alpha \in K$ such that $K = F(\alpha)$ is called a primitive element for the extension K/F.

Theorem 23 (Primitive element theorem)**.** Every finite extension K of a field F of characteristic zero contains a primitive element.

To prove the lemma, we will use induction and the following lemma.

Lemma 24. Let F be a field of characteristic zero, and let K be an extension field of F that is generated over F by two elements α and β. For all but finitely many $c \in F$, $\gamma = \beta + c\alpha$ is a primitive element for K over F.

Proof. Let $f(x)$ and $g(x)$ be the irreducible polynomials for α and β, respectively, over F. Let $\mathcal{K}$ be a field extension of K in which f and g split completely and call their respective roots in L $\alpha_1, \ldots, \alpha_m$ and $\beta_1, \ldots, \beta_n$, with $\alpha = \alpha_1$ and $\beta = \beta_1$.

Since $\operatorname{char} F = 0$, the roots α_i are distinct, as are the roots β_j. Let $\gamma_{ij} = \beta_j + c\alpha_i$ with $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Note that, if $(i, j) \neq (k, \ell)$, then $\gamma_{i,j} = \gamma_{k,\ell}$ holds for *at most one* value of c. (To see this, just solve the corresponding equation for c.) Suppose c is not one of these values, then we claim that $\gamma = \gamma_{11} = \beta_1 + c\alpha_1$ is a primitive element for K over F.

Set $L = F(\gamma)$. We will show that $\alpha = \alpha_1 \in L$. It then follows that $\beta = \beta_1 = \gamma - c\alpha_1 \in L$ and $L = K$. Note that α_1 is a root of $f(x)$, and it is also a root of $h(x) = g(\gamma - cx)$ and $h(x) \in L[x]$. (To see this, note that $h(\alpha_1) = g(\gamma - c\alpha_1) = g(\beta_1) = 0$.)

Recall that the gcd d of f and h is the same in $L[x]$ and $\mathcal{K}[x]$. Since $f(x) = (x-\alpha_1)\cdots(x-\alpha_m)$ in $\mathcal{K}[x]$, d is the product of factors $(x-\alpha_i)$ that also divide h. Of course, one common root is $(x-\alpha_i)$. We claim this is the only common factor. It then follows that $d = x - \alpha_i \in L[x]$.

Suppose $i > 1$, then $h(\alpha_i) = g(\gamma - c\alpha_i)$. The roots of g are $\beta_1, \dots, \beta_n$, so it suffices to show that $\gamma - c\alpha_i \neq \beta_j$ for any j. Equivalently, $\beta_1 + c\alpha_1 \neq \beta_j + c\alpha_i$. This is true because of the choice of c and this completes the proof. □

Proof of the Primitive Element Theorem. Since the extension K/F is finite, $K = F(\alpha_1, \dots, \alpha_k)$ for some k. If $k = 1$, then we are done. Otherwise, for $k > 1$ we proceed by induction. Assume the theorem is true for $K_1 = F(\alpha_1, \dots, \alpha_{k-1})$. Then by induction we may assume that $K_1 = F(\beta)$ for some primitive element β. But then K will be generated by two elements α_k and β. Thus, the result follows by the lemma. □

Disclaimer/Warning: The primitive element theorem holds for a finite field F but with a different proof.

Galois Theory

(Last Updated: December 5, 2018)

These notes are derived primarily from *Algebra* by Michael Artin (2ed), though some material is drawn from *Abstract Algebra, Theory and Applications* by Thomas Judson (16ed).

1. Symmetric Functions

The primary motivation for Galois' work was whether a quintic equation with integer coefficients was solvable by radicals. That is, can all of the roots of an integer polynomial be expressed using only integers, nth roots, and basic arithmetic operations? (Spoiler alert: no.) We will develop this theory and (hopefully) see this result toward the end.

Throughout, we assume that the roots of an irreducible polynomial over a field are distinct, and that a finite extension field K/F has a primitive element.

Let R be a ring and let $R[u]$ denote the polynomial ring $R[u_1, \ldots, u_n]$ in n variables over R. A permutation $\sigma \in \mathcal{S}_n$ operates on polynomials by permuting the variables:

$$f = f(u_1, \ldots, u_n) \mapsto f(u_{\sigma(1)}, \ldots, u_{\sigma(n)}) = \sigma(f).$$

Then σ defines an automorphism of $R[u]$. By an abuse of notation, we denote this automorphism also by σ. This automorphism acts trivially on the scalars R in $R[u]$, and so we call σ an R-automorphism. A symmetric polynomial is one that is fixed by every permutation $\sigma \in \mathcal{S}_n$.

A polynomial g is symmetric if two monomials that are in the same orbit have the same coefficient in g. We call the sum of the monomials in an orbit an orbit sum. The orbit sums form a basis for the (vector) space of symmetric polynomials.

Example. The orbit sums of degree at most 3 in three variables are

$$\begin{gathered} 1, \quad u_1 + u_2 + u_3, \quad u_1^2 + u_2^2 + u_3^2, \quad u_1u_2 + u_1u_3 + u_2u_3, \\ u_1^3 + u_2^3 + u_3^3, \quad u_1u_2^2 + u_1^2u_2 + u_1u_3^2 + u_1^2u_3 + u_2u_3^2 + u_2^2u_3, \quad u_1u_2u_3. \end{gathered}$$

The elementary symmetric functions are defined as

$$\begin{aligned} s_1 &= \sum_i u_1 = u_1 + u_2 + \cdots + u_n \\ s_2 &= \sum_{i<j} u_i u_j = u_1 u_2 + u_1 u_3 + \cdots \\ s_3 &= \sum_{i<j<k} u_i u_j u_k = u_1 u_2 u_3 + \cdots \\ &\vdots \quad \vdots \\ s_n &= u_1 u_2 \cdots u_n = u_1 u_2 \cdots u_n. \end{aligned}$$

The elementary symmetric functions are the coefficients of the polynomial with variable roots $u_1, \ldots, u_n$:

$$\begin{aligned} P(x) &= (x - u_1)(x - u_2) \cdots (x - u_n) \\ &= x^n - s_1 x^{n-1} + s_2 x^{n-2} - \cdots + (-1)^n s_n. \end{aligned}$$

Suppose that a polynomial

$$f(x) = x^n - a_1 x^{n-1} + a_2 x^{n-2} - \cdots + (-1)^n a_n$$

splits completely in a field K with roots $\alpha_1, \ldots, \alpha_n$. That is, in K,

$$f(x) = (x - \alpha_1) \cdots (x - \alpha_n).$$

It follows by substituting $u_i = \alpha_i$ that the coefficients of f are obtained by evaluating the elementary symmetric functions. That is

$$a_i = s(\alpha_1, \ldots, \alpha_n).$$

We will show that the ring symmetric functions in $R[u_1, \ldots, u_n]$ is the ring $R[s_1, \ldots, s_n]$. Every symmetric function can be written as a polynomial in the elementary symmetric functions. This fact is not obvious. Before proving we first consider some examples.

Example. Let $f(u) = u_1^2 + \cdots + u_n^2$. Then $f(u) = c_1 s_1^2 + c_2 s_2$ for some scalars c_1, c_2. Substituting $u = (1, 0, \ldots, 0)$ gives $c_1 = 1$. Substituting $u = (1, -1, 0, \ldots)$ shows that $c_2 = -2$. Hence, $u_1^2 + \cdots + u_n^2 = s_1^2 - 2s_2$.

Example. Consider

$$g(u) = u_1 u_2^2 + u_2 u_1^2 + u_1 u_3^2 + u_3 u_1^2 + u_2 u_3^2 + u_3 u_2^2$$

in the three variables u_1, u_2, u_3. Set $u_3 = 0$ and define $g^\circ = u_1 u_2^2 + u_2 u_1^2$ in the variables $u_1 u_2$. Let s_i° denote the elementary symmetric functions in u_1, u_2, so $s_1^\circ = u_1 + u_2$ and $s_2^\circ = u_1 u_2$. Then $g^\circ = s_1^\circ s_2^\circ$.

Back in the polynomial ring on three variables:

$$s_1 s_2 = (u_1 + u_2 + u_3)(u_1 u_2 + u_1 u_3 + u_2 u_3).$$

This product has nine terms. One of the terms in this expansion is $u_1 u_2^2$, and so it must contain the orbit sum of $u_1 u_2^2$, which has six terms. The remaining three terms are all equal to $u_1 u_2 u_3$. Hence, $g = s_1 s_2 - 3s_3$.

Theorem 1 (Symmetric Functions Theorem)**.** Every symmetric polynomial $g(u_1, \dots, u_n)$ with coefficients in a ring R can be written in a unique way as a polynomial in the elementary symmetric functions $s_1, \dots, s_n$.

Proof. Let $g(u_1, \dots, u_n) \in R[u_1, \dots, u_n]$ be symmetric. We claim $g(u_1, \dots, u_n) = G(s_1, \dots, s_n)$ for some $G \in R[s_1, \dots, s_n]$.

If $n = 1$, then $u_1 = s_1$ and we are done. Assume the theorem holds for $n - 1$ variables. Let g° be the polynomial obtained from g by setting $u_n = 0$. Then g° is a symmetric in $u_1, \dots, u_{n-1}$ and so by the inductive hypothesis, $g^\circ = Q(s_1^\circ, \dots, s_{n-1}^\circ)$.

Let $p(u_1, \dots, u_n) = g(u_1, \dots, u_n) - Q(s_1, \dots, s_{n-1})$. As p is a difference of symmetric polynomials, it is symmetric. Moreover, $p(u_1, \dots, u_{n-1}, 0) = g^\circ - Q(s^\circ) = 0$. Thus, u_n divides p, but because p is symmetric, every u_i divides p. Therefore, s_n divides p and so $p = s_n h$ where h is symmetric.

It now follows that $g = Q(s_1, \dots, s_{n-1}) + s_n h$ where h is symmetric and of lower degree. Hence, we may apply the inductive hypothesis to h for the result. The uniqueness also follows from the inductive process. □

Corollary 2. Suppose that $f(x) = x^n - a_1 x^{n-1} + \cdots \pm a_n$ has coefficients in a field F, and that it splits completely in an extension field K, with roots $\alpha_1, \dots, \alpha_n$. Let $g(u_1, \dots, u_n)$ be a symmetric polynomial in $u_1, \dots, u_n$ with coefficients in F. Then $g(\alpha_1, \dots, \alpha_n) \in F$.

Proof. By the Symmetric Function Theorem, $g(u_1, \dots, u_n) = G(s_1, \dots, s_n)$ for some polynomial G. We observed above that $s_i(\alpha) = a_i \in F$. □

Example. Let $p_1 = u_1^2 + u_2 u_3$ be a polynomial in three variables. The orbit of p_1 consists of three polynomials:

$$p_1 = u_1^2 + u_2 u_3, \quad p_2 = u_2^2 + u_3 u_1, \quad p_1 = u_3^2 + u_1 u_2.$$

Then a symmetric polynomial in the p_i will be a symmetric polynomial in the u_i. For example, $p_1 p_2 + p_2 p_3 + p_3 p_1$ is symmetric in u_1, u_2, u_3 (check!).

Proposition 3. Let $p_1 = p_1(u_1, \dots, u_n)$ be a polynomial and let $\{p_1, \dots, p_k\}$ be its orbit for the operation of the symmetric group on the variables. If $h(p_1, \dots, p_k)$ is a symmetric polynomial in the p_i, the it is a symmetric polynomial in the u_i.

3. Splitting fields

Let F be a field and $f \in F[x]$ (f need not be irreducible). A splitting field for f over F is an extension field K/F such that f splits completely in K and K is generated over F by the roots of f. That is, $f(x) = (x - \alpha_1)\dots(x - \alpha_n)$ with $\alpha_i \in K$ and $K = F(\alpha_1, \dots, \alpha_n)$. The second condition implies that for every $\beta \in K$, there exists $p(u_1, \dots, u_n) \in F[x]$ such that $p(\alpha_1, \dots, \alpha_n) = \beta$. Previously we saw that there is always a field in which a polynomial f splits completely. In some sense, a splitting field is the minimal field in which f splits completely.

Lemma 4. (1) If $F \subset L \subset K$ are fields, and if K is a splitting field of a polynomial f over F, then K is also a splitting field of the same polynomial over L.
(2) Every polynomial $f(x) \in F[x]$ has a splitting field.
(3) An extension field of F is finite if and only if it is contained in splitting field.

Proof. (1) The statement is clear. (2) Construct an extension field K' of F in which f splits completely as we did previously. The roots of f all live in K' and are algebraic over F. Take K to be the field generated over F by the roots of f. Then K is also a splitting field for f.

(3) A splitting field is a finite extension. Suppose L/F is a finite extension, say by $\gamma_1, \dots, \gamma_k$, each algebraic over F. Let g_i be the irreducible polynomial for γ_i over F, and let $f = g_1 \cdots g_k$. We extend L to a splitting field K of f over L. Thus, K is a splitting field over F as well. □

Theorem 5 (Splitting Theorem)**.** Let K be an extension of a field F that is a splitting field of a polynomial $f(x) \in F[x]$. If an irreducible polynomial $g(x) \in F[x]$ has a root in K, then it splits completely in K.

Proof. Let f and g be as in the statement of the theorem. Let $\beta_1 \in K$ be a root of g in K. Since g is irreducible, it is the irreducible polynomial of β_1. Since K is the splitting field of f over F, then $K = F(\alpha_1, \dots, \alpha_n)$ where the α_i are the roots of f. Every element of K can be written as a polynomial in α with coefficients in F. Choose a polynomial $p_1(u_1, \dots, u_n)$ such that $p_1(\alpha) = \beta_1$.

Let $\{p_1, \dots, p_k\}$ be the orbit of $p_1(u)$ for the action of the symmetric group $\mathcal{S}_n$ on $F[u_1, \dots, u_n]$, and let $\beta_j = p_j(\alpha)$. Thus, $\beta_1, \dots, \beta_k$ are elements of K. We claim that $h(x) = (x - \beta_1) \cdots (x - \beta_k)$ has coefficients in F. Once shown, then since β_1 is a root of h, the irreducible polynomial, g, of β_1 divides h. Since h splits completely in K then so does g.

Say $h(x) = x^k - b_1x^{k-1} + b_2x^{k-2} - \cdots + (-1)^k b_k$. The coefficients $b_1, \dots, b_k$ are obtained by evaluating elementary symmetric functions at $\beta = \beta_1, \dots, \beta_k$. These are the elementary symmetric functions in k variables. Let $s_1'(p), \dots, s_k'(p)$ be the elementary symmetric functions in the p_i. Then $s_j'(p(u))$ is a symmetric polynomial in u and so $b_j = s_j'(\beta_j) = s_j'(p(\alpha)) \in F$. □

Moral: An irreducible polynomial over F with one root in K splits completely in K.

4. Isomorphisms of field extensions

Disclaimer/Warning: Henceforth, we assume that all fields have characteristic zero and all extensions are assumed to be finite.

Recall that an F-isomorphism between extension fields K/F and K'/F is a (field) isomorphism whose restriction to F is the identity map. An F-automorphism of an extension field K/F is an F-isomorphism from K to K. The F-automorphisms of K are the symmetries of the field extension.

Definition. The Galois group of a finite extension K/F is the group of F-automorphisms of K. A finite extension K/F is a Galois extension if the order of its Galois group: $G(K/F)$ is equal to the degree of the extension: $|G(K/F)| = [K : F]$.

Example. Consider $\mathbb{C}$ as an extension of $\mathbb{R}$ ($\mathbb{C} = \mathbb{R}(i)$). The only nontrivial $\mathbb{R}$-automorphism of $\mathbb{C}$ is the map given by $\phi(i) = -i$ (that is, ϕ is just complex conjugation). Thus, $G(K/F) = 2$ and so $\mathbb{C}$ is a Galois extension. We get an analogous result by adjoining a square root to a field F.

Lemma 6. Let K and K' be extensions of a field F.

(1) Let $f(x)$ be a polynomial with coefficients in F, and let σ be an F-isomorphism from K to K'. If α is a root of f in K, then $\sigma(\alpha)$ is a root of f in K'.

(2) Suppose that K is generated over F by some elements $\alpha_1, \ldots, \alpha_n$. Let σ and σ' be F-isomorphisms $K \to K'$. If $\sigma(\alpha_i) = \sigma'(\alpha_i)$ for $i = 1, \ldots, n$, then $\sigma = \sigma'$. If an F-automorphism σ of K fixes all of the generators, it is the identity map.

(3) Let f be an irreducible polynomial with coefficients in F, and let α and α' be roots of f in K and K', respectively. There is a unique F-isomorphism $\sigma : F(\alpha) \to F(\alpha')$ that sends α to α'. If $F(\alpha) = F(\alpha')$, then σ is an F-automorphism.

Proposition 7. (1) Let f be a polynomial with coefficients in F. An extension field L/F contains at most one splitting field of f over F.

(2) Let f be a polynomial with coefficients in F. Any two splitting fields of f over F are isomorphism extensions of fields.

Proof. (1) If L contains a splitting field of f, then say $f(x) = (x - \alpha_1) \cdots (x - \alpha_n)$ with $\alpha_i \in L$. Let β be a root of f in L, then $\beta = \alpha_i$ for some i. Hence, f has no additional roots in L, and so the only splitting field of f that is contained in L is $F(\alpha_1, \ldots, \alpha_n)$.

(2) Let K_1 and K_2 be splitting fields of f over F. As K_1 is a finite extension of F, it contains a primitive element γ. Let g be the irreducible polynomial for γ over F. Choose an extension field L of K_2 in which g has a root γ', and we let K' denote the subfield $F(\gamma')$ of L generated by γ'. There is an F-isomorphism $\phi : K_1 \to K'$ that sends $\gamma \mapsto \gamma'$. Because K' is F-isomorphic to the splitting field K_1, it is also a splitting field of f. Then both K' and K_2 are splitting fields contained in the field L, and part (1) shows that they are equal. Thus, ϕ is an F-isomorphism from K_1 to K_2. □

5. Fixed Fields

Lemma 8. Let H be a group of automorphisms of a field K. The set

$$K^H := \{\alpha \in K : \sigma(\alpha) = \alpha \text{ for all } \sigma \in H\}$$

is a subfield of K called the fixed field of H.

Proof. Recall that elements of H are field automorphisms. Let $\sigma \in H$. Hence, $\sigma(1) = 1$, and $\sigma(0) = 0$ so $0, 1 \in K^H$. Moreover, for all $a, b \in K^H$, $a + b = \sigma(a) + \sigma(b) = \sigma(a + b)$, so $a + b \in K^H$. Similarly, $ab \in K^H$. Finally, if $c \in K^H$ and $c \neq 0$, then $c^{-1} = \sigma(c)^{-1} = \sigma(c^{-1})$, so $c^{-1} \in K^H$. Thus, K^H is a field and since $K^H \subset K$, then K^H is a subfield of K. □

Let K be an extension of a field F, then an intermediate field L is a field such that $F \subset L \subset K$. We now preview of the main theorem of Galois Theory.

Theorem 9 (Main Theorem of Galois Theory)**.** Let K be a Galois extension of a field F, and let G be its Galois group. There is a bijective correspondence between subgroups of G and intermediate fields. The correspondence is given by the inverse functions

$$\begin{array}{ccc} \{\text{subgroups}\} \to \{\text{intermediate fields}\} & \qquad & \{\text{intermediate fields}\} \to \{\text{subgroups}\} \\ H \mapsto K^H & & L \mapsto G(K/L). \end{array}$$

Our next goal will be to prove that $H = G(K/K^H)$ and a key piece of this is the Fixed Field Theorem. This next theorem follows almost directly from our discussion on symmetric functions.

Theorem 10. Let H be a finite group of automorphisms of a field K and let F denote the fixed field K^H. Let β_1 be an element of K, and let $\{\beta_1, \dots, \beta_r\}$ be the H-orbit of β_1.

(1) The irreducible polynomial for β_1 over F is $g(x) = (x - \beta_1) \cdots (x - \beta_r)$.

(2) The element β_1 is algebraic over F, and its degrees over F is equal to the order of its orbit. Therefore the degree of β_1 over F divides the order of H.

Proof. We will prove (1) and (2) follows directly. Write

$$g(x) = (x - \beta_1) \cdots (x - \beta_r) = x^n - b_1 x^{n-1} + b_2 x^{n-2} - \cdots + (-1)^r b_r.$$

Thus, the coefficients of $g(x)$ are symmetric functions of the orbit $\{\beta_1, \dots, \beta_r\}$. Since the elements of H permute the orbit, they fix the coefficients. It follows that g has coefficients in the fixed field.

Let $h \in F[x]$ be a polynomial with β_1 as a root. For $i = 1, \dots, r$, there exists $\sigma \in H$ such that $\sigma(\beta_1) = \beta_i$. The elements of H are F-automorphisms of K and $h \in F[x]$ so it follows that β_i is a root of h. That is, $x - \beta_i$ divides h. This is true for every i and so g divides h in $K[x]$ and therefore in $F[x]$. Thus, g generates the (principal) ideal of polynomials in $F[x]$ with β_1 as a root. This proves that g is the irreducible polynomial of β_1 over F. □

Lemma 11. Let K be an infinite algebraic extension of a field F (every element of K is algebraic over F) There exist elements in K whose degrees over F are arbitrarily large.

Proof. Choose an element $\alpha_1 \in K \backslash F$ and set $F_1 = F(\alpha_1)$. Then α_1 is algebraic over F and so $[F_1 : F] < \infty$, thus $K \neq F(\alpha_1)$. Now choose an element $\alpha_2 \in K \backslash F_1$ and set $F_2 = F_1(\alpha_2) = F(\alpha_1, \alpha_2)$. Again, α_2 is algebraic over F and so $[F_2 : F] < \infty$, thus $K \neq F_2$. Moreover, there exists a primitive element γ_2 such that $F_2 = F(\gamma_2)$ (we can also retroactively set $\gamma_1 = \alpha_1$). Continue in this way to obtain a proper tower of extension $F \subset F_1 \subset F_2 \subset \cdots$ such that $[F_{i-1} : F] < [F_i : F] < \infty$ and the corresponding primitive elements γ_i become arbitrarily large. □

Theorem 12 (Fixed Field Theorem)**.** Let H be a finite group of automorphisms of a field K and let $F = K^H$. Then K is a finite extension of F, and $[K : F] = |H|$.

Proof. Let $F = K^H$ and set $n = |H|$. The previous theorem shows that K/F is algebraic, and that the degree over F of any element $\beta \in K$ divides n. Therefore $[K : F] < \infty$ by the lemma. Let γ be a primitive element for this extension. Since $\sigma \in H$ acts as the identity on F, then if σ fixes γ it will be the identity map on K. This would imply that $\sigma = 1_H$. Thus, the stabilizer of γ is the the trivial subgroup of H and the orbit of γ has order n. But the theorem shows that γ has degree n over F. Since $K = F(\gamma)$, then $[K : F] = n$ as well. □

Example. Let $K = \mathbb{C}(t)$ (the field of rational functions over $\mathbb{C}$ in one variable). Let σ and τ be automorphisms of K that are the identity on $\mathbb{C}$ and such that $\sigma(gt) = it$ and $\tau(t) = t^{-1}$ (exercise: verify that these are indeed automorphisms). Then $\sigma^4 = 1$, $\tau^2 = 1$, and $\tau\sigma = \sigma^{-1}\tau$. Therefore, σ and τ generate a group of automorphisms H that is isomorphic to D_4.

Consider the rational function $u = t^4 - t^{-4}$ over $\mathbb{C}$. Suppose $g(x) = x^d + c_{d-1}x^{d-1} + \cdots + c_0 \in \mathbb{C}[x]$ is any monic polynomial of degree d. Then by substitution, $t^{4d}g(u)$ is a monic polynomial of degree $8d$ in t. Since t is transcendental over $\mathbb{C}$, then $t^{4d}g(u) \neq 0$, so $g(u) \neq 0$. It follows that u is transcendental over $\mathbb{C}$.

Thus, $\mathbb{C}(u)$ is isomorphic to a field of rational functions in one variable. Note that u is fixed by σ and τ, so $u \in K^H$ and thus $\mathbb{C}(u) \subset K^H$. We claim that $\mathbb{C}(u) = K^H$. By the first theorem, the irreducible polynomial for t over K^H is the polynomial whose roots form its orbit:

$$\{t, it, -t, -it, t^{-1}, -it^{-1}, -t^{-1}, it^{-1}\}.$$

The polynomial whose roots are the elements of this orbit is:

$$(x^4 - t^4)(x^4 - t^{-4}) = x^8 - ux^4 + 1.$$

Thus, t is a root of a polynomial of degree 8 with coefficients in $\mathbb{C}(u)$. Therefore, $[K : \mathbb{C}(u)] \leq 8$. By the Fixed Field Theorem, $[K : K^H] = 8$ and since $\mathbb{C}(u) \subset K^H$, then $\mathbb{C}(u) = K^H$.

6. Galois Extensions

Let K be an extension field of F and L an intermediate field of the extension (so $F \subset L \subset K$). Since every L-automorphism is necessarily an F-automorphism, we have $G(K/L) \subset G(K/F)$.

Lemma 13. (1) The Galois group G of a finite field extension K/F is a finite group whose order divides $[K : F]$.

(2) Let H be a finite group of automorphisms of a field K. Then K is a Galois extension of its fixed field K^H, and H is the Galois group of K/K^H.

Proof. (1) By a prior result we know that G is finite and elements of G act trivially on F, so $F \subset K^G \subset K$. Thus, $[K : K^G]$ divides $[K : F]$. By the Fixed Field Theorem, $|G| = [K : K^G]$.

(2) By definition of K^H, the elements of H are K^H-automorphisms. Thus, H is a subgroup of $G(K/K^H)$. Since $|G(K/K^H)|$ divides $[K : K^H]$ and $|H| = [K : K^H]$, the two groups are equal and K is a Galois extension of K^H. □

Lemma 14. Let γ_1 be a primitive element for a finite field extension K/F. Let $f(x)$ be the irreducible polynomial for γ_1 over F with roots $\gamma_1, \ldots, \gamma_r \in K$. There is a unique F-automorphism σ_i of K such that $\sigma_i(\gamma_1) = \gamma_i$. These are all of the F-automorphisms of K, so $|G(K/F)| = r$.

Proof. There is a unique F-automorphism σ_i of K such that $\sigma_i(\gamma_1) = \gamma_i$. Since $K = F(\gamma_1)$ and each $F(\gamma_i)$ has the same degree over F, then $K = F(\gamma_i)$ for each i. Therefore, σ_i is an F-automorphism of K. Conversely, every F automorphism sends γ_1 to a root of f, so it is one of the σ_i. □

Theorem 15 (Characteristic properties of Galois extensions)**.** Let K/F be a finite extension and let G be its Galois group. The following are equivalent.

(1) K/F is a Galois extension, i.e., $|G| = [K : F]$.

(2) The fixed field K^G is equal to F.

(3) K is a splitting field over F.

Proof. (1) ⇔ (2) By the Fixed Field Theorem, $|G| = [K : K^G]$. Since $F \subset K^G \subset K$, $|G| = [K : F]$ if and only if $F = K^G$.

(1) ⇔ (3) Let $n = [K : F]$ and choose a primitive element γ_1 for K over F. Let f be its irreducible polynomial over F. Since γ_1 is a primitive element, the degree of f is n. Let $\gamma_1, \ldots, \gamma_r$ be the roots of f that are in K. By the lemma, $|G| = r$, so $|G| = [K : F]$ if and only if f splits completely in K. Since $K = F(\gamma_1)$, then K is generated by all of the roots of f and is therefore a splitting over F it and only if f splits completely in K. □

In light of the previous theorem, given the splitting field K of a polynomial f over F, we may refer to $G(K/F)$ as the Galois group of f.

Corollary 16. (1) Every finite extension K/F is contained in a Galois extension.

(2) If K/F is a Galois extension, and if L is an intermediate field, then K is also a Galois group extension of L, and $G(K/L)$ is a subgroup of $G(K/F)$.

Theorem 17. Let K/F be a Galois extension with Galois group G, and let g be a polynomial with coefficients in F that splits completely in K. Let its roots in K be $\beta_1, \dots, \beta_r$.

(1) The group G operates on the set of roots $\{\beta_i\}$.

(2) If K is a splitting field of g over G, the operation on the roots is faithful, and by its operation on the roots, G embeds as a subgroup of the symmetric group $\mathcal{S}_r$.

(3) If g is irreducible over F, the operation on the roots is transitive.

(4) If K is a splitting field of g over F and g is irreducible over F, then G embeds as a transitive subgroup of $\mathcal{S}_r$.

Proof. Parts (1) and (2) are standard while part (4) is a combination of (2) and (3). We are left to prove (3). Suppose g is irreducible over F. Then it is the irreducible polynomial for β_1 over F. Since $F = K^G$, the the roots β_i of g form the G-orbit of β_1. Hence, the operation is transitive. □

7. The Main Theorem

Theorem 18 (Main Theorem of Galois Theory)**.** Let K be a Galois extension of a field F, and let G be its Galois group. There is a bijective correspondence between subgroups of G and intermediate fields. The correspondence is given by the inverse functions

$$\begin{aligned} \{\text{subgroups}\} &\to \{\text{intermediate fields}\} & \qquad \{\text{intermediate fields}\} &\to \{\text{subgroups}\} \\ H &\mapsto K^H & L &\mapsto G(K/L). \end{aligned}$$

Proof. It suffices to show that the composition of two of the maps in either order is the identity. Let H be a subgroup of G and let L be its fixed field. By the Fixed Field Theorem, $G(K/L) = H$. For the opposite direction, let L be an intermediate field and let H be the Galois group of K over L. Then K is a Galois extension of L (Corollary 16 (2)) and so the fixed field of H is L by the Characteristic Properties of Galois extensions. □

Corollary 19. (1) The correspondence above reverses inclusions: If L and L' are intermediate fields and if H and H' are the corresponding subgroups, then $L \subset L'$ if and only if $H' \subset H$.

(2) The subgroup that corresponds to the field F is the whole group $G(K/F)$, and the subgroup that corresponds to K is the trivial subgroup $\{1\}$.

(3) If L corresponds to H, then $[K : L] = |H|$ and $[L : F] = [G : H]$.

Corollary 20. A finite field extension K/F has finitely many intermediate fields $F \subset L \subset K$.

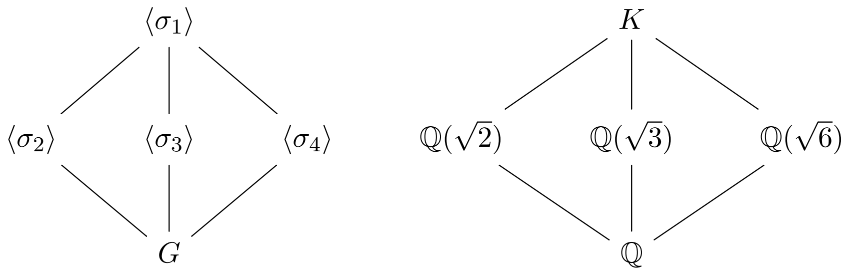
Example. Let $K = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Since $x^2 - 3$ is irreducible over $\mathbb{Q}(\sqrt{2})$, then $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6})$ is a basis of K over $\mathbb{Q}$. Let $\sigma \in \operatorname{Aut}_{\mathbb{Q}} K$, then σ is completely determined by $\sigma(\sqrt{2})$ and $\sigma(\sqrt{3})$. These automorphisms permute the roots of irreducible polynomials. Hence, $\sigma(\sqrt{2}) = \pm\sqrt{2}$ and $\sigma(\sqrt{3}) = \pm\sqrt{3}$. Thus, there are at most four distinct $\mathbb{Q}$-automorphisms of K and each is indeed an automorphism. Define the automorphisms σ_i by $\sigma_1 = \text{id}$ and

$$\sigma_2(\sqrt{2}) = \sqrt{2}, \sigma_2(\sqrt{3}) = -\sqrt{3}, \quad \sigma_3(\sqrt{2}) = -\sqrt{2}, \sigma_3(\sqrt{3}) = \sqrt{3}, \quad \sigma_4(\sqrt{2}) = -\sqrt{2}, \sigma_4(\sqrt{3}) = -\sqrt{3}$$

Each of the σ_i, $i \neq 1$ generates a proper (normal) subgroup of order 2. It follows that $\operatorname{Aut}_{\mathbb{Q}} K \cong C_2 \times C_2$. Since $[K : \mathbb{Q}] = 4$, then K is Galois over $\mathbb{Q}$. The fixed fields are

$$K^{\langle\sigma_2\rangle} = \mathbb{Q}(\sqrt{2}), \quad K^{\langle\sigma_3\rangle} = \mathbb{Q}(\sqrt{3}), \quad K^{\langle\sigma_4\rangle} = \mathbb{Q}(\sqrt{6}).$$

In each case $[K : \mathbb{Q}(\sqrt{k})] = 2$ and $[G : \langle\sigma_i\rangle] = 2$. The correspondence can be visualized as



Suppose we have a tower of fields $F \subset L \subset K$. If K is a Galois extension of F, the K is a Galois extension of L. However, in general, L need not be a Galois extension of F.

Theorem 21. Let K/F be a Galois extension with Galois group G, and let $L = K^H$ with H a subgroup of G. The extension L/F is Galois if and only if H is a normal subgroup of G, in which case $G(L/F) \cong G/H$.

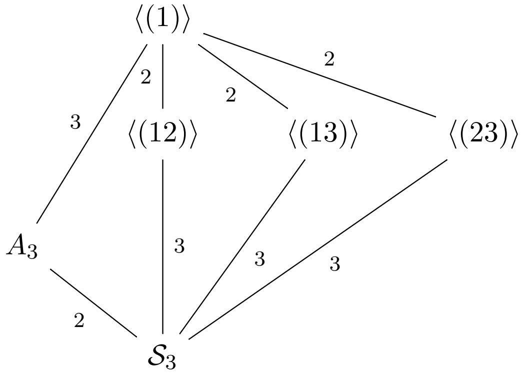
Proof. Let $\epsilon_1 \in L$ be a primitive element for the extension L/F, and let g be the irreducible polynomial for ϵ_1 over F. Since K is Galois over F, it is a splitting field for some polynomial By the Splitting Theorem, g splits completely in K with roots, say, $\epsilon_1, \ldots, \epsilon_r$. Recall that L/F is Galois if and only if L is a splitting field, in which case all $\epsilon_i \in L$.

Let $\sigma \in G$ such that $\sigma(\epsilon_1) = \epsilon_i$. Then $F(\epsilon_i) = L$ if and only if $\epsilon_i \in L$, in which case the stabilizer of ϵ_i is H. On the other hand, the stabilizer of $\sigma(\epsilon_1)$ is $\sigma H \sigma^{-1}$. Therefore, L/F is Galois if and only if $\sigma H \sigma^{-1} = H$.

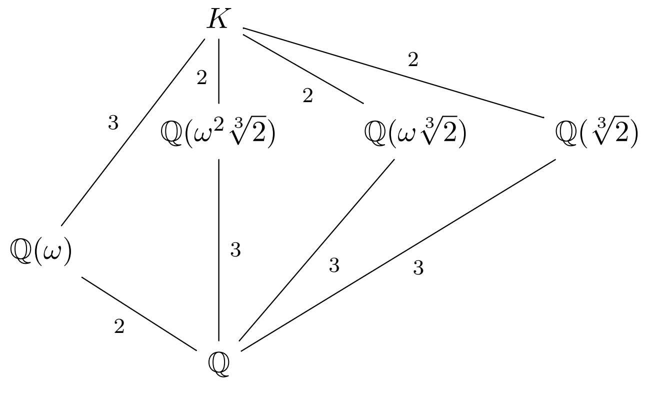
Suppose L/F is Galois, so the $\epsilon_i \in L$. An element $\sigma \in G$ maps ϵ_1 to ϵ_i, and therefore maps $L = F(\epsilon_1)$ to $F(\epsilon_i) = L$. Restricting σ to L defines an F-automorphism of L. This restriction gives a homomorphism $\phi : G \to G(L/F)$ whose kernel is the set of σ that restrict to the identity on L. This is precisely H. Moreover, $|G/H| = [G : H] = |G(L/F)|$. Now, by the First Isomorphism Theorem, $G/H \cong G(L/F)$. □

Example. Set $L = \mathbb{Q}(\sqrt[3]{2})$ and consider the field extension $\mathbb{Q} \subset L$. The irreducible polynomial of $\sqrt[3]{2}$ over $\mathbb{Q}$ is $x^3 - 2$. *However*, the other roots of this polynomial are $\omega\sqrt[3]{2}$ and $\omega^2\sqrt[3]{2}$ where ω is a primitive third root of unity not equal to 1. It's clear that these other two roots (because they are complex) do not live in L. Hence, the only field automorphism of K over $\mathbb{Q}$ is the identity (because it must permute the roots of $x^3 - 2$). It follows that $|G(L/\mathbb{Q})| = 1$ while $[L : \mathbb{Q}] = 3$.

Example. Let K be the splitting field of $x^3 - 2$ over $\mathbb{Q}$. Then $K = \mathbb{Q}(\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2}) = \mathbb{Q}(\sqrt[3]{2}, \omega)$. As K is a splitting field, it is Galois over $\mathbb{Q}$. We have $[K : \mathbb{Q}] = 6$ (convince yourself that $x^2 + x + 1$ is irreducible over $\mathbb{Q}(\sqrt[3]{2})$). Moreover, $G = G(K/\mathbb{Q}) = 6$ because there are six permutations of the three roots of $x^3 - 2$. Thought of another way, if $\sigma \in G$, then $\sigma(\sqrt[3]{2}) \in \{\sqrt[3]{2}, \omega\sqrt[3]{2}, \omega^2\sqrt[3]{2}\}$ and $\omega \in \{\omega, \omega^2\}$ since the minimal polynomial of ω over $\mathbb{Q}$ is $x^2 + x + 1$. Thus, $G \cong \mathcal{S}_3$. The subgroup diagram for $\mathcal{S}_3$ is



Label the roots of $x^3 - 2$ as $1 = \sqrt[3]{2}$, $2 = \omega\sqrt[3]{2}$, and $3 = \omega^2\sqrt[3]{2}$ Then (12), thought of as an element of G, fixes $\omega^2\sqrt[3]{2}$. Hence, $K^{\langle(12)\rangle} = \mathbb{Q}(\omega^2\sqrt[3]{2})$. Similarly, $K^{\langle(13)\rangle} = \mathbb{Q}(\omega\sqrt[3]{2})$ and $K^{\langle(23)\rangle} = \mathbb{Q}(\sqrt[3]{2})$. As ω has degree 2 over $\mathbb{Q}$, then it stands to reason that $K^{A_3} = \mathbb{Q}(\omega)$. Thus, the corresponding picture for the intermediate fields is


K
2
2
2
3
$\mathbb{Q}(\omega^2\sqrt[3]{2})$
$\mathbb{Q}(\omega\sqrt[3]{2})$
$\mathbb{Q}(\sqrt[3]{2})$
$\mathbb{Q}(\omega)$
3
3
3
2
$\mathbb{Q}$

8. Cubic equations

Let $f(x) = x^3 - a_1x^2 + a_2x - a_3$ be an irreducible (cubic) polynomial over a field F and let K be the splitting field of f. Let $\alpha_1, \alpha_2, \alpha_3 \in K$ be the roots of f. Then $f(x) = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$. Since $\alpha_1 + \alpha_2 + \alpha_3 = a_1 \in F$, then $F(\alpha_1, \alpha_2) = F(\alpha_1, \alpha_2, \alpha_3) = K$. Set $L = F(\alpha_1)$, so $[L : F] = 3$. In $L[x]$, $f(x) = (x - \alpha_1)q(x)$ where q is the quadratic polynomial with roots α_2, α_3. Thus, K is obtained from L by adjoining a root of q. We have two cases:

(1) If q is irreducible over L, then $[K : L] = 2$ and so $[K : F] = 6$.

(2) If q is reducible over L, then $\alpha_2, \alpha_3 \in L$ so $L = K$ and $[K : F] = 3$.

The action of the Galois group $G = G(K/F)$ is isomorphic to a transitive subgroup of $\mathcal{S}_3$: $\mathcal{S}_3$ (in case $[K : F] = 6$) or A_3 (in case $[K : F] = 3$). Thus, the goal is to determine whether $q(x)$ is irreducible over $L = F(\alpha_1)$. Alternatively, one can use the discriminant.

A digression on discriminants: Let $P(x)$ be a polynomial of degree n with variable roots $u_1, \dots, u_n$. The discriminant is defined as

$$D(u) = (u_1 - u_2)^2(u_1 - u_3)^2 \cdots (u_{n-1} - u_n)^2 = \prod_{i<j}(u_i - u_j)^2.$$

Note that $D(u)$ is a symmetric polynomial with integer coefficients and if $\alpha_1, \dots, \alpha_n$ are elements of a field, then $D(\alpha) = 0$ if and only if two of the α_i are equal.

Back to the cubic. Let δ be the square root of the discriminant of f, so

$$\delta = (\alpha_1 - \alpha_2)(\alpha_1 - \alpha_3)(\alpha_2 - \alpha_3).$$

Observe that $\delta \in K$ (clearly), $\delta \neq 0$ (the roots are distinct because f is irreducible), and a permutation of the roots of f multiplies δ be the sign of the permutation.

Theorem 22 (Galois theory for a cubic)**.** Let K be the splitting field of an irreducible polynomial f over a field F, let D be the discriminant of f, and let $G = G(K/F)$.

(1) If D is a square in F, then $[K : F] = 3$ and $G \cong A_3$.

(2) If D is not a square in F, then $[K : F] = 6$ and $G \cong \mathcal{S}_3$.

Proof. Suppose $\delta \in F$. Because $F = K^G$ (K is the splitting field of f so K/F is Galois) then δ is fixed by every permutation of G. But $(12)\delta = -\delta$ (similarly for (13) and (23)), and so G does not contain an odd permutation (more generally $\sigma\delta = \operatorname{sgn}(\sigma)\delta$). Thus, $G \cong A_3$. On the other hand, if $\delta \notin F$ then it is *not* fixed by all of G. Thus, G must contain an odd permutation, so $G \cong \mathcal{S}_3$. □

If $G \cong A_3$, then G has no proper subgroups, whence no intermediate fields for F/K (this is also clear because $[K : F] = 3$, a prime). On the other hand, there $\mathcal{S}_3$ has four proper subgroups corresponding to the intermediate fields $F(\delta)$, $F(\alpha_1)$, $F(\alpha_2)$, $F(\alpha_3)$.

12. Quintic equations

Here we see the real connection between Galois Theory and solvability of polynomials over radicals.

Let $F \subset \mathbb{C}$ and $\alpha \in F$. We say that α is solvable over F if there is a chain of subfields $F = F_0 \subset F_1 \subset \cdots \subset F_r = K$ of $\mathbb{C}$ such that $\alpha \in K$ and for $j = 1, \ldots, r$, $F_j = F_{j-1}(\beta_j)$ where $\beta_j^n \in F_{j-1}$ for some n. This definition says that F_j is generated over F_{j-1} by attaching an nth root of some element. There is an equivalent characterization in that each F_{j+1}/F_j is a Galois extension of prime degree. We will prove that these are equivalent at the end of the section. If α is a root of a polynomial of degree at most four with coefficients in a field F, then α is solvable over F (exercise).

Theorem 23. Let f be an irreducible polynomial of degree 5 over a subfield F of $\mathbb{C}$ whose Galois group is either $\mathcal{S}_5$ or A_5. Then the roots of f are not solvable over F.

Proof. If $G = \mathcal{S}_5$, then we replace F by the quadratic extension $F(\delta)$ where δ is square root of the discriminant of F. If we can solve over F, then we can solve over $F(\delta)$, so we may assume that $G = A_5$. Recall that A_5 is a simple group. We will use the second characterization of solvability. Let F'/F be a Galois extension of F of prime degree p with Galois group G' (so G' is a cyclic group). As F'/F is Galois, F' is the splitting field for some irreducible polynomial g over F.

Let K be the splitting field of f. Let K' be the splitting field, over F, of the fg. It is generated by the roots $\alpha_1, \ldots, \alpha_5$ and $\beta_1, \ldots, \beta_p$ of f and g, respectively. We have the diagram

$$\begin{tikzcd}
 & K' \arrow[dl, dash, "H'"'] \arrow[dr, dash, "H"] \arrow[dd, dash, "\mathcal{G}"] & \\
K \arrow[dr, dash, "G"'] & & F' \arrow[dl, dash, "G'"] \\
 & F &
\end{tikzcd}$$

Since K/F is Galois, $G \cong \mathcal{G}/H'$ and similarly $G' \cong \mathcal{G}/H$. We claim $H \cong A_5$.

Suppose the claim holds. Then replacing F with a Galois extension F' does not change the Galois group. Doing this enough times, one must eventually get a root of f in K/F in F. But A_5 contains an element of order 5, and so cannot be the Galois group of a *reducible* polynomial of degree 5.

The group H' consists of F-automorphisms of K' that fix the α_i, while H consists of F-automorphisms that fix the β_j. An F-automorphism that fixes both is the identity on K'. Thus, $H \cap H'$ is the trivial group. We restrict the canonical map $\mathcal{G} \to \mathcal{G}/H \cong G'$ to H'. The kernel of this map is $H \cap H'$, so the restriction is injective and maps H' isomorphically to a subgroup of G', which is of prime order. Thus, either H' is the trivial group or else H' has order p. If H' is trivial, then the surjective map $\mathcal{G} \to \mathcal{G}/H' \cong G \cong A_5$ is an isomorphism. Then $\mathcal{G}$ is simple and so there is no surjective map $\mathcal{G} \to \mathcal{G}/H \cong G'$. Thus, H' is cyclic of order p, so $|\mathcal{G}| = |G||H'| = p|G|$ and $|\mathcal{G}| = |G'||H| = p|H|$.

Thus, G and H both have order 60. We restrict the canonical map $\mathcal{G} \to \mathcal{G}/H' \cong G$ to H and the kernel is again the trivial group $H \cap H'$, so the restriction is injective. Thus, H is isomorphic to a subgroup of G, when $H \cong G \cong A_5$. □

Now we produce an example of a polynomial that is not solvable over $\mathbb{Q}$.

Corollary 24. Let $f(x)$ be an irreducible polynomial of degree 5 over $\mathbb{Q}$. If f has exactly three real roots, its Galois group G is the symmetric group, and hence its roots are not solvable.

Proof. Let the roots be $\alpha_1, \dots, \alpha_5$ with the first three real and the last two complex. Let K be the splitting field of f. The only permutations of the roots that fix the first three are the identity and the transposition (45). Since $F(\alpha_1, \alpha_2, \alpha_3) \neq K$, that transposition must be in G. But G operates transitively on the roots, so it contains an element of order 5. These elements generate $\mathcal{S}_5$ (exercise), so $G \cong \mathcal{S}_5$. □

Example. The polynomial $x^5 - 16x + 2$ satisfies the above corollary.

We return to the notion of solvability. Our goal is to show that the two definitions of solvability are, in fact, equivalent. First we consider a case in which roots are *always* solvable.

Lemma 25. Let K/F be a Galois extension whose Galois group G is abelian. There is a chain of intermediate fields $F = F_0 \subset F_1 \subset \cdots F_m = K$ such that F_i/F_{i-1} is a Galois extension of prime degree for each i.

Proof. The abelian group G contains a subgroup H of prime order H corresponds to an intermediate field L, and K is a Galois extension of L with $H = G(K/L)$. As G is abelian, H is normal and so L is a Galois extension of F with abelian Galois group G/H. Now G/H has smaller order than G and so we may apply induction to complete the proof. □

A cyclotomic field is a subfield F of the complex numbers generated over $\mathbb{Q}$ be an nth root of unity $\zeta_n = e^{2\pi i/n}$. When $n = p$ a prime, the irreducible polynomial for ζ_p is $f(x) = x^{p-1} + \cdots + x + 1$. Since the roots of f are the $p-1$ powers of ζ_p, then ζ_p generates the splitting field of f, so $K = F(\zeta_p)$ is a degree $p-1$ Galois extension of F.

Proposition 26. Let p be a prime and $\zeta = e^{2\pi i/p}$. The Galois group $G = G(\mathbb{Q}(\zeta)/\mathbb{Q})$ is a cyclic group of order $p-1$ isomorphic to $\mathbb{F}_p^\times$. Moreover, for any subfield F' of $\mathbb{C}$, the Galois group $G' = G(F'(\zeta)/F')$ is a cyclic group.

Proof. Set $F = \mathbb{Q}$ and let $G = G(F(\zeta)/F)$. An element $\sigma \in G$ is determined by the image $\sigma(\zeta)$, which must be one of the roots of $f(x) = x^{p-1} + \cdots + x + 1$. Thus, there are $p-1$ maps σ_i determined by $\sigma_i(\zeta) = \zeta^i$. Define a map $\epsilon : G \to \mathbb{F}_p^\times$ by $\sigma_i \mapsto i$. This map is well defined because the exponent of ζ is taken modulo p. Then $\sigma_i\sigma_j(\zeta) = \sigma_i(\zeta^j) = (\zeta^j)^i = \zeta^{ij} = \sigma_{ij}(\zeta)$. Hence, the map is a homomorphism and clearly bijective. Since $\mathbb{F}_p^\times$ is cyclic, G is as well.

For an arbitrary subfield F', an element $\sigma \in G' = G(F'(\zeta)/F')$ will send ζ to a power of itself. Thus, by the above, G' is isomorphic to a subgroup of $\mathbb{F}_p^\times$, so it is also cyclic. □

Extensions of the type described below are called Kummer extensions.

Theorem 27. Let F be a subfield of $\mathbb{C}$ that contains the pth root of unity $\zeta = e^{2\pi i/p}$, p prime, and let K/F be a Galois extension of degree p. Then $K = F(\beta)$ with $\beta^p \in F$.

Proof. Set $G = G(K/F)$. Since $|G| = p$, then G is cyclic so let $\sigma \in G$ be a generator. For all $c \in F$, $\sigma(c) = c$, and for $\alpha, \beta \in K$, $\sigma(\alpha + \beta) = \sigma(\alpha) + \sigma(\beta)$ and $\sigma(c\alpha) = \sigma(c)\sigma(\alpha) = c\sigma(\alpha)$. These all follow because σ, by definition, is a field automorphism of K. However, viewing K as a vector space over F, this shows that σ is in fact a F-linear operator on K.

As $\sigma^p = 1$, every eigenvalue λ satisfies $\lambda^p = 1$, so λ is a power of ζ. By hypothesis, every eigenvalue is in F. Because every matrix over the complex numbers is diagonalizable, then σ must have an eigenvalue different from 1 (otherwise, σ would just be the identity). Let β be an eigenvector of σ with eigenvalue $\lambda \neq 1$ and set $b = \beta^p$. Then $\sigma(\beta) = \lambda\beta$, so $\sigma(b) = (\lambda\beta)^p = b$. As σ generates G, $b \in K^G = F$. But $F \subset F(\beta) \subset K$ and $\beta \notin F$. Since $[K : F]$ is prime, then $F(\beta) = K$. □

Proposition 28. Let $F \subset K$ be subfields of $\mathbb{C}$ and let $\alpha \in K$. The following are equivalent:

(1) $F = F_0 \subset F_1 \subset \cdots \subset F_r = K$ such that $F_j = F_{j-1}(\beta_j)$ for $j = 1, \ldots, r$ where $\beta_j^n \in F_{j-1}$ for some n.

(2) $F = F_0 \subset F_1 \subset \cdots \subset F_s = K$ such that F_{j+1}/F_j is Galois of prime degree for $j = 1, \ldots, s$.

Proof. (2) ⇒ (1) Suppose we have a chain as in (2) and consider one of the (Galois) extensions, say $F_{i-1} \subset F_i$, so $[F_i : F_{i-1}] = p$ for some prime p. By Theorem 27, this extension is obtained by adjoining a pth root, so long as that pth root is in F_{i-1}. If necessary, we enlarge the chain by first adjoining the necessary pth roots and so this new chain will satisfy (1).

(1) ⇒ (2) Suppose we have a chain as in (1). Possibly after enlarging our chain, we may suppose that the roots that occur in our chain are pth roots for various primes p. Suppose these primes are listed, in order, as $p_1, \ldots, p_k$.

Adjoin to F a p_ith root of unity for $i = 1, \ldots, k$, successively. Each extension is Galois with cyclic Galois group (Proposition 26). By Lemma 25, each contains a chain whose layers are Galois extensions of prime degree. Let F' be the resulting field. We continue adjoining the given roots to F'. Each root adjunction is Galois (Theorem 27) with cyclic Galois group of prime order, unless the extension is trivial. The field K' be obtain at the end of our new chain necessarily contains K, the last field of the original chain. Hence, $\alpha \in K'$. This new chain has the form (1). □