Contents

4 Elements of Algebra

Our goal in this chapter is to discuss two foundational objects in algebra: groups and fields. We begin with a broad introduction to groups via the axiomatic definition, and then discuss various fundamental examples of groups such as the integers modulo m, dihedral groups, and symmetric groups. We then establish some basic properties of groups, subgroups, and element orders in groups, culminating in Cauchy's theorem on elements of prime order and Lagrange's theorem on orders of subgroups. We also discuss fields and give various fundamental examples of fields and ordered fields: the integers modulo a prime p, the rational numbers, the real numbers, and the complex numbers.

4.1 Groups

- The set of symmetries of a geometric or algebraic object carries a natural structure under composition.
    - This composition operation is associative (since function composition is associative), there is always an identity element (namely, the identity symmetry that leaves the object unchanged), and every element has an inverse (namely, the "inverse" symmetry that reverses everything).
    - To study the collection of symmetries, therefore, is essentially the same as studying algebraic structures with a single operation that possess three properties of associativity, existence of an identity, and existence of inverses.

4.1.1 The Formal Definition of a Group

- Definition: A group is any set G having a (closed) binary operation $\star$ that satisfies the three axioms [G1]-[G3]:

  [G1] The operation $\star$ is associative: $g \star (h \cdot k) = (g \star h) \star k$ for any elements g, h, k in G.

  [G2] There is a (two-sided) identity element e: $e \star g = g = g \star e$ for any element g in G.

  [G3] Every element has a (two-sided) inverse: for any g in G, there exists g^{-1} in G with $g \star g^{-1} = e = g^{-1} \star g$.

- Note that we do not assume the operation $\star$ in the group is commutative. More precisely:

- Definition: If a group satisfies axiom [G4], we say it is an abelian group[1].

  [G4] The operation $\star$ is commutative: $g \star h = h \star g$ for any elements g, h in G.

- Definition: If G is a group, the order of G, denoted as $|G|$ or $\#G$, is the cardinality of G as a set.

- There are a number of common conventions regarding group notation.

  - Because the group operation is associative, we do not need to specify the order in which the multiplications are performed when we have more than 2 terms, and can simply write expressions like $g \star h \star k$ without needing to use parentheses to distinguish between $(g \star h) \star k$ and $g \star (h \star k)$.[2]
  - If $g \in G$, for any positive integer n we define $g^n = \underbrace{g \star g \star \cdots \star g}_{n \text{ terms}}$, $g^{-n} = \underbrace{g^{-1} \star g^{-1} \star \cdots \star g^{-1}}_{n \text{ terms}}$, and $g^0 = e$.
  - We will frequently omit the symbol for the group operation $\star$ and simply write gh for $g \star h$.
  - We will also often write the operation as $\cdot$ or $+$ when it represents multiplication or addition in a context where those operations are already familiar, and write 1 or 0 for the corresponding identity elements respectively. Also, when the group operation is addition, we write inverses additively, as $-a$ rather than a^{-1}.

- Here are some basic examples (and non-examples) of groups:

- Example: The nonzero rational numbers form an abelian group under multiplication.

  - Explicitly, for [G1] we have $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all rationals a, b, c, for [G2] we have an identity 1 with $1 \cdot a = a = a \cdot 1$ for all rationals a, for [G3] every rational number a has a multiplicative inverse a^{-1} such that $a \cdot a^{-1} = 1 = a^{-1} \cdot a$, and for [G4] we have $a \cdot b = b \cdot a$ for all rationals a, b.

- Example: The integers form an abelian group under addition.

  - Explicitly, for [G1] we have $(a + b) + c = a + (b + c)$ for all integers a, b, c, for [G2] we have an identity 0 with $0 + a = a = a + 0$ for all integers a, for [G3] every integer a has an additive inverse $-a$ such that $a + (-a) = 0 = (-a) + a$, and for [G4] we have $a + b = b + a$ for all integers a, b.

- Non-Example: The integers do not form a group under multiplication, because 0 has no multiplicative inverse.

  - Even if we exclude 0, the nonzero integers still do not form a group, because 2 (and 3, and 4, etc.) all fail to possess multiplicative inverses.

- Non-Example: The positive integers do not form a group under addition.

  - Although [G1] holds, [G2] does not since there is no additive identity inside the positive integers.

- Non-Example: The nonnegative integers do not form a group under addition.

[1] Less commonly, abelian groups are also called commutative groups. A group that is not abelian is called non-abelian. The term "abelian" is named after Neils Henrik Abel, who was a foundational figure in the study of groups; it is stylized in lowercase (rather than in uppercase as "Abelian") in honor of the depth of his contribution.

[2] Technically, this statement requires a proof; it is straightforward though tedious to use induction on the number of terms in the product to establish that all such products are equal to the one where the order is composed left-to-right, as in $((g \star h) * k) \star l$.

- Although [G1] and [G2] both hold (since now the set contains 0, the additive identity), [G3] does not since for example 1 does not possess an additive inverse inside the nonnegative integers.

- Example: For any $m > 1$, the integers modulo m form an abelian group under addition, of order m.
  - Explicitly, for [G1] we have $(\overline{a} + \overline{b}) + \overline{c} = \overline{a} + (\overline{b} + \overline{c})$ for all residue classes $\overline{a}, \overline{b}, \overline{c}$, for [G2] we have an identity $\overline{0}$ with $\overline{0} + \overline{a} = \overline{a} = \overline{a} + \overline{0}$ for all residue classes $\overline{a}$, for [G3] every residue class $\overline{a}$ has an additive inverse $-\overline{a}$ (namely $\overline{-a}$) such that $\overline{a} + (-\overline{a}) = \overline{0} = (-\overline{a}) + \overline{a}$, and for [G4] we have $\overline{a} + \overline{b} = \overline{b} + \overline{a}$ for all residue classes $\overline{a}, \overline{b}$.

- Example: The set $\{1, -1\}$ forms an abelian group under multiplication. This group has order 2.
  - It is easy to see that multiplication here is associative and commutative, that 1 is an identity, and that both elements are their own inverses.

- Example: The set $G = \{e\}$, with operation $e \cdot e = e$, is a group called the trivial group.
  - This group has order 1, and in fact is the only possible group structure for a group of order 1.

- Example: The set $V_4 = \{e, a, b, c\}$ with identity e, and other multiplications given by $a^2 = b^2 = c^2 = 1$, $ab = ba = c$, $ac = ca = b$, and $bc = cb = a$, forms an abelian group of order 4.
  - It is straightforward (although tedious) to verify that multiplication is associative. In this group, every element is its own inverse.
  - This group is called the Klein 4-group (in German, "Viergruppe") and is denoted V_4 or K_4.

- Example: If m is a modulus, the set $(\mathbb{Z}/m\mathbb{Z})^\times$ of residue classes relatively prime to m forms an abelian group under multiplication.
  - As we saw in our discussion of residue class arithmetic, multiplication is associative and commutative, the residue class $\overline{1}$ is a multiplicative identity, and each residue class relatively prime to m has a multiplicative inverse (by the Euclidean algorithm).
  - For example, with $m = 5$ we have four residue classes $\{\overline{1}, \overline{2}, \overline{3}, \overline{4}\}$. We have inverses $\overline{1}^{-1} = \overline{1}$, $\overline{2}^{-1} = \overline{3}$, $\overline{3}^{-1} = \overline{2}$, and $\overline{4}^{-1} = \overline{4}$.

- Example: For any positive integer n, if $\zeta_n = e^{2\pi i/n}$, then the set $G = \{1, \zeta_n, \zeta_n^2, \ldots, \zeta_n^{n-1}\}$ forms a group of order n under multiplication.
  - Explicitly: associativity is inherited from $\mathbb{C}$, the identity element is 1, and $(\zeta_n^k)^{-1} = \zeta_n^{n-k}$ for any $0 \leq k \leq n-1$.
  - This group consists of the solutions to the equation $x^n - 1 = 0$ in $\mathbb{C}$, which are called the nth roots of unity. For this reason the group is often called the group of nth roots of unity.
  - For example, when $n = 4$, we obtain the multiplicative group $G = \{1, i, -1, -i\}$, where $i = \sqrt{-1}$ is the imaginary unit. In this group we have for example $i^2 = -1$, $(-1) \cdot (-i) = i$, $i \cdot (-i) = 1$, and also $i^{-1} = -i$ and $(-i)^{-1} = i$.

- We can deduce a few properties of group arithmetic immediately from the axioms:

- Proposition (Basic Group Arithmetic): Let G be a group. The following properties hold in G:

  1. The identity element e is unique, and $e^{-1} = e$.
     - Proof: For (1), if there were two identity elements e and e', then $e' = e \cdot e' = e$ by the left-identity property of e and the right-identity property of e'. The second statement follows immediately by observing that $ee = e$.
  2. G has left and right cancellation: for any g, h, k in G, either of $gh = gk$ or $hg = kg$ implies $h = k$.
     - Proof: If $gh = gk$ then $h = eh = (g^{-1}g)h = g^{-1}(gh) = g^{-1}(gk) = (g^{-1}g)k = ek = k$. The other statement follows similarly.

3. Inverses are unique. Also, a one-sided inverse of g is automatically a two-sided inverse of g.
   - Proof: If h and k are both inverses of g, then $gh = e = gk$, so by cancellation we see $h = k$.
   - The second statement follows by observing that $gh = e$ implies $h = eh = (g^{-1}g)h = g^{-1}(gh) = g^{-1}e = g^{-1}$, and likewise $hg = e$ also implies $h = g^{-1}$.
4. For any $g, h \in G$, $(gh)^{-1} = h^{-1}g^{-1}$, and $(g^{-1})^{-1} = g$.
   - Proof: We have $(h^{-1}g^{-1})(gh) = h^{-1}(g^{-1}g)h = h^{-1}eh = h^{-1}h = e$ and likewise for the product in the other order.
   - For the second statement note $(g^{-1})^{-1}g^{-1} = e = gg^{-1}$, so cancelling g^{-1} yields $(g^{-1})^{-1} = g$.

- We can also construct new groups using Cartesian products.
  - Recall that if S and T are sets, the Cartesian product $S \times T$ is the set of ordered pairs (s, t) where $s \in S$ and $t \in T$.

- Proposition (Cartesian Products of Groups): If $(G, \star)$ and $(H, \circ)$ are groups, then the Cartesian product $G \times H$ is also a group, with operation performed componentwise: $(g_1, h_1)\triangle(g_2, h_2) = (g_1 \star g_2, h_1 \circ h_2)$. The identity element is $e_{G\times H} = (e_G, e_H)$ and inverses are given by $(g, h)^{-1} = (g^{-1}, h^{-1})$. The group $G \times H$ has order $\#G \cdot \#H$, and is abelian if and only if both G and H are abelian.
  - Proof: Each of the group axioms for $G \times H$ follows immediately from the corresponding axioms in G and H, and the statement about the order follows from the definition of Cartesian product for sets.
  - For the abelian condition, clearly $(g_1, h_1)\triangle(g_2, h_2) = (g_1 \star g_2, h_1 \circ h_2)$ is equal to $(g_2, h_2)\triangle(g_1, h_1) = (g_2 \star g_1, h_2 \circ h_1)$ for all $g_1, g_2 \in G$ and $h_1, h_2 \in H$ if and only if $g_1 \star g_2 = g_2 \star g_1$ and $h_1 \circ h_2 = h_2 \circ h_1$ for all $g_1, g_2 \in G$ and $h_1, h_2 \in H$.

- Example: The Cartesian product $(\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/5\mathbb{Z})$ is an abelian group of order $3 \cdot 5 = 15$.

4.1.2 Dihedral Groups

- As we briefly outlined, groups arise naturally from studying symmetries of objects. Among the simplest objects in geometry are regular n-gons, whose associated symmetry group is called the dihedral group, and denoted[3] $D_{2\cdot n}$.
  - Geometrically, these symmetries are the possible ways to move an n-gon around in space (rotating or reflecting it) and then placing it back on top of itself so that all of the vertices and edges line up.
  - For example, for $n = 4$ (corresponding to the symmetries of a square), one possibility is to rotate the square $\pi/2$ radians counterclockwise in the plane around its center. Another possibility is to reflect the square about one of its diagonals (in fact there are two such maps).

- If we label the vertices of the n-gon $1, 2, \ldots, n$, then we can identify all of these symmetries as functions acting on the vertices.
  - For example, if we label the vertices of the square as $1, 2, 3, 4$ counterclockwise, then a counterclockwise rotation of $\pi/2$ radians would correspond to the function σ with $\sigma(1) = 2$, $\sigma(2) = 3$, $\sigma(3) = 4$, and $\sigma(4) = 1$.
  - The collection of symmetries $D_{2\cdot n}$ of the regular n-gon can then be made into a group as follows: if g and h are both elements of $D_{2\cdot n}$, we define the composition $g \cdot h$ to be the symmetry obtained by first applying h, and then g (i.e., by function composition).
  - This operation is associative since function composition is associative, the identity element is the identity transformation (i.e., the symmetry leaving all vertices fixed), and the inverse of a symmetry g is the symmetry g^{-1} that reverses all of the rigid motions of g.

[3] Many authors denote the symmetry group of the n-gon as D_n (emphasizing the geometric flavor of the group), but in group theory literature the notation D_{2n} (emphasizing the elements of the group) is more common. We adopt the notation $D_{2\cdot n}$ as a sort of compromise between these two.

- Proposition (Order of $D_{2\cdot n}$): For any integer $n \geq 3$, the dihedral group $D_{2\cdot n}$ has order $2n$.
  - Proof: Under a symmetry, the vertex labeled 1 can be moved to any of the n vertices, and then the vertex labeled 2 must go to one of the 2 vertices adjacent to it. But once we have fixed the locations of vertices 1 and 2, then all of the other vertices' locations are determined uniquely (since vertex 3 must go to the unique vertex adjacent to the new position of vertex 2 that is not already occupied by vertex 1, and so forth).
  - Thus there are at most $2n$ possible symmetries of a regular n-gon, so $\#D_{2\cdot n} \leq 2n$.
  - On the other hand, we can explicitly list $2n$ distinct symmetries: there are the n possible rotations counterclockwise about the center by $2\pi k/n$ radians for $0 \leq k \leq n-1$, and there are also n possible reflections about a line through the center of the n-gon.
  - Explicitly: if n is odd, these are the n lines passing through one vertex and the center, while if n is even there are $n/2$ lines passing through a pair of opposite vertices and $n/2$ others that bisect a pair of opposite sides.
  - Each of these symmetries is different, so $D_{2\cdot n}$ has order $2n$ as claimed.
- We can give a more concrete description of the elements in $D_{2\cdot n}$ in terms of particular rotations and reflections.
  - Explicitly, let r represent the counterclockwise rotation of the n-gon by $2\pi/n$ radians: as a function on vertices, we have $r(1) = 2$, $r(2) = 3$, ... , $r(n-1) = n$, and $r(n) = 1$. Then r^k represents a counterclockwise rotation by $2\pi k/n$ radians, so the elements $\{e, r, r^2, \ldots, r^{n-1}\}$ are distinct, and $r^n = e$.
  - Also, let s represent the reflection of the n-gon across the line through vertex 1 and the center of the n-gon. As a permutation, we have $s(1) = 1$, $s(2) = n$, $s(3) = n-1$, ... , and $s(n) = 2$. It is then easy to see that s^2 is the identity element, and that $s \neq r^i$ for any i, since the only power of r that fixes vertex 1 is the identity element.
  - From this we can conclude that all of the elements $\{s, sr, sr^2, \ldots, sr^{n-1}\}$ are distinct, since $sr^i = sr^j$ would imply $r^{i-j} = e$ by cancellation, and they are also all distinct from the elements $\{e, r, r^2, \ldots, r^{n-1}\}$ since $sr^i = r^j$ would imply $s = r^{j-i}$ by cancellation.
  - Hence we see that $D_{2\cdot n} = \{e, r, r^2, \ldots, r^{n-1}, s, sr, sr^2, \ldots, sr^{n-1}\}$.
  - To describe the multiplication of any two elements in this list, we first observe that $rs = sr^{-1}$ (so in particular, $D_{2\cdot n}$ is always non-abelian). This relation can be visualized geometrically, since rotating and then reflecting is equivalent to reflecting and then rotating in the opposite direction.
  - Alternatively, we can compute $rs(1) = r(1) = 2$ and $rs(2) = r(n) = 1$, and also $sr^{-1}(1) = s(n) = 2$ and $sr^{-1}(2) = s(1) = 2$. Then since rs and sr^{-1} agree on vertices 1 and 2, they agree on all vertices, so they are equal.
  - Then by an easy induction, we see that $r^i s = sr^{-i}$ for all i.
- To summarize the discussion, $D_{2\cdot n} = \{e, r, r^2, \ldots, r^{n-1}, s, sr, sr^2, \ldots, sr^{n-1}\}$, where r and s are elements satisfying the relations $r^n = s^2 = e$ and $rs = sr^{-1}$.
  - Using these relations (and the ancillary fact that $r^i s = sr^{-i}$ for any i) we can compute the product of any two elements in $D_{2\cdot n}$.
  - For example, in $D_{2\cdot 7}$, we have $(sr^5)(r^4) = sr^9 = sr^2$, $(r^4)(sr^5) = sr^{-4}(r^5) = sr$, and $(sr^2)(sr) = s(r^2 s)r = s(sr^{-2})r = s^2 r^{-1} = r^6$.

4.1.3 Symmetric Groups and Permutations

- Another natural class of groups arises from "symmetries" of sets.
  - To illustrate the idea, observe that the set S_3 of permutations of the set $A = \{1, 2, 3\}$ (formally, the set of bijections of S with itself) forms a group under composition.
  - Note that there are a total of $3! = 6$ such bijections. A somewhat-convenient way to represent these maps is to write a list of the elements of the domain and target vertically: thus the map f with $f(1) = 2$, $f(2) = 3$, and $f(3) = 1$ would be written as $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$.

- In this notation, the 6 elements of S_3 are $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$.
- To compute the product of two elements in S_3, we can simply trace the behavior of each element of $\{1, 2, 3\}$ under the corresponding composition of functions.
- Thus, for example, if $g = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$ and $h = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, to compute the product gh we observe that (i) h sends 1 to 3, and g sends 3 to 3, so gh sends 1 to 3, (ii) h sends 2 to 1, and g sends 1 to 2, so gh sends 2 to 2, and (iii) h sends 3 to 2, and g sends 2 to 1, so gh sends 3 to 1.
- Thus, $gh = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$. In a similar way we can compute $hg = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$, so we see in particular that S_3 is non-abelian.
- It is very tedious to verify that these operations actually form a group using this explicit description (checking associativity, for example, requires 6^3 individual calculations), and the notation is also quite cumbersome.

- We can clarify matters by generalizing this idea to arbitrary sets.
- Proposition (Symmetric Groups): If A is any set, the set of bijections from A to itself forms a group under function composition. This group is the symmetric group on the set A and is denoted S_A. When $\#A = n$ is finite we have $\#S_A = n!$, and when A is infinite, S_A is infinite.
  - Proof: The group operation is well-defined because the composition of two bijections is also a bijection. Property [G1] follows because function composition is associative, property [G2] follows because the identity map is a bijection, and property [G3] follows because the inverse of a bijection is also a bijection.
  - For the statement about the cardinality, suppose first that $\#A = n$. Then, as we showed using pigeonhole ideas, a function $f : A \to A$ is a bijection if and only if f is one-to-one. But there are $n!$ possible one-to-one functions from A to A, since the first element of A has n possible destinations, the second then has $n - 1$ possible destinations, and so forth, yielding a total number of $n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1 = n!$ possible f.
  - Finally, if A is infinite, for any fixed $x \in A$ and any $y \in A$ consider the map f_y that interchanges x and y and leaves all other elements alone. Then f_y is a bijection for each $y \in A$, so since there are infinitely many $y \in A$ this already yields infinitely many bijections.
- We will primarily be interested in the case where $A = \{1, 2, \ldots, n\}$, in which case we will write the group as S_n, the symmetric group on n objects. The elements of this group are called permutations because they rearrange the elements of the set.
  - First, we would like a more convenient way to describe the elements in S_n. We can achieve this by writing permutations in terms of cycles $(a_1\, a_2\, \ldots\, a_k)$.
  - Explicitly, the cycle $(a_1\, a_2\, \ldots\, a_k)$ is the permutation σ with $\sigma(a_1) = a_2$, $\sigma(a_2) = a_3$, ... , $\sigma(a_{k-1}) = a_k$, and $\sigma(a_k) = a_1$, where all other elements are mapped to themselves. This permutation "cycles" the elements $a_1, a_2, \ldots, a_k$ one step forward (whence the name).
  - Thus, for example, inside S_4 the cycle $(2\,1\,4)$ is the permutation with $\sigma(2) = 1$, $\sigma(1) = 4$, $\sigma(4) = 2$, and $\sigma(3) = 3$.
  - Not every permutation can be written as a single cycle, but it is not hard to see that every permutation can be written as a product of disjoint cycles (i.e., cycles having no elements in common) such as $(1\,3)\,(2\,4)$, which represents the permutation with $\sigma(1) = 3$, $\sigma(3) = 1$, $\sigma(2) = 4$, and $\sigma(4) = 2$. Such a representation is called the cycle decomposition of σ.
  - Explicitly, to determine all of the cycles in the cycle decomposition of a permutation σ, we start with the smallest number x not contained in one of the cycles we have identified, and repeatedly apply σ until we obtain a repeated element. In other words, we evaluate $a_1 = x$, $a_2 = \sigma(a_1)$, $a_3 = \sigma(a_2)$, $a_4 = \sigma(a_3)$, ... until the list repeats.

- It is easy to see that the first repeated value will always be x (since $a_i = a_j$ implies $\sigma(a_{i-1}) = \sigma(a_{j-1})$ so that $a_{i-1} = a_{j-1}$ since σ is a bijection), and so we obtain a cycle $(x\, a_2 \ldots a_k)$ containing x. We repeat this process until we have identified the cycles containing every element in $\{1, 2, \ldots, n\}$.

- Example: Find the cycle decomposition of the permutation $\sigma \in S_6$ with $\sigma(1) = 3$, $\sigma(2) = 5$, $\sigma(3) = 4$, $\sigma(4) = 1$, $\sigma(5) = 2$, and $\sigma(6) = 6$.
  - We start with $n = 1$: we compute $\sigma(1) = 3$, $\sigma(3) = 4$, and $\sigma(4) = 1$. This gives the cycle $(1\,3\,4)$.
  - The smallest number not yet used is $n = 2$: then $\sigma(2) = 5$ and $\sigma(5) = 2$, so we obtain the cycle $(2\,5)$.
  - The smallest number not yet used is $n = 6$: since $\sigma(6) = 6$ we obtain the cycle (6).
  - Since we have used all 6 elements in cycles, we see that the cycle decomposition of σ is $\boxed{(1\,3\,4)(2\,5)(6)}$.

- Example: Find the cycle decomposition of the permutation $\sigma \in S_7$ with $\sigma(1) = 1$, $\sigma(2) = 3$, $\sigma(3) = 4$, $\sigma(4) = 7$, $\sigma(5) = 5$, $\sigma(6) = 6$, and $\sigma(7) = 2$.
  - Since $\sigma(1) = 1$ we obtain the cycle (1). Then since $\sigma(2) = 3$, $\sigma(3) = 4$, $\sigma(4) = 7$, and $\sigma(7) = 2$ we obtain the cycle $(2\,3\,4\,7)$.
  - Then since $\sigma(5) = 5$ we obtain the cycle (5). Finally since $\sigma(6) = 6$ we obtain the cycle (6).
  - Since we have used all 7 elements in cycles, the cycle decomposition of σ is $\boxed{(1)(2\,3\,4\,7)(5)(6)}$.

- Definition: The length of a cycle is the number of elements it contains. A cycle of length k is called a k-cycle, and 2-cycles are often called transpositions.
  - The notation for cycle decompositions is not unique. For example, the cycle $(1\,3\,4)$ corresponds to the same permutation as the cycle $(3\,4\,1)$, and the cycle decomposition $(1\,3\,4)(2\,5)(6)$ is the same as $(2\,5)(6)(1\,3\,4)$.
  - We adopt the convention of writing the cycles with the smallest element first, and ordering the cycles in increasing order of their first element. Under this convention, it follows by a straightforward induction argument that the cycle decomposition is unique, and that the algorithm we described above will compute it.
  - It is also common to omit 1-cycles when we write cycle decompositions, with the convention always being that any unlisted elements are fixed (i.e., mapped to themselves). Thus, we would simply write $(1\,3\,4)(2\,5) \in S_6$ and omit the 1-cycle (6). This convention is useful when describing permutations that fix most of the elements in the set.

- We can also compute products using cycle decompositions, with the important remark that the products of cycles are read right-to-left, since they are representing compositions of functions.
  - We can compute the cycle decomposition of the product by tracing what happens to each element $1, 2, \ldots, n$ under each of the cycles from right-to-left, and then using the cycle decomposition algorithm.

- Example: If $g = (1\,3\,4)(2\,5)$ and $h = (1\,2)(3\,5)$ inside S_5, compute the cycle decomposition of gh.
  - Since h sends 1 to 2, and g sends 2 to 5, the composition gh sends 1 to 5.
  - To compute the next element in the cycle containing 1 we need to determine where gh sends 5. Since h sends 5 to 3, and g sends 3 to 4, we see that gh sends 5 to 4.
  - Continuing, we see $gh(4) = g(4) = 1$, which completes a cycle $(1\,5\,4)$.
  - Also, since $gh(2) = g(1) = 3$ and $gh(3) = g(5) = 2$, we get the other cycle $(2\,3)$. Thus the cycle decomposition of gh is $\boxed{(1\,5\,4)(2\,3)}$.

- Example: The six elements in S_3 have respective cycle decompositions e, $(1\,2)$, $(1\,3)$, $(2\,3)$, $(1\,2\,3)$, $(1\,3\,2)$.
  - We can compute, for example, $(1\,2)(1\,3) = (1\,3\,2)$, by tracing what happens to each element from right to left in each of the cycles. (Explicitly, these tracings would look something like $1 \to 3 \to 3$, $3 \to 1 \to 2$, and $2 \to 2 \to 1$.)

- Similarly, $(1\,3)(1\,2) = (1\,2\,3)$, $(1\,3\,2)(1\,2) = (2\,3)$, and $(1\,2)(1\,3\,2)(1\,3) = (2\,3)$ as well.

- Since a cycle $(a_1\,a_2\,\ldots\,a_k)$ represents the permutation that shifts a_1 to a_2, a_2 to a_3, ... , and a_k to a_1, the inverse of the cycle simply shifts in reverse: it sends a_k to a_{k-1}, a_{k-1} to a_{k-2}, ... , a_3 to a_2, a_2 to a_1, and a_1 to a_k.

  - This describes to the cycle $(a_k\,a_{k-1}\,\ldots\,a_3\,a_2\,a_1)$ obtained by reversing the order of the elements. Rearranging it to put the smallest element a_1 first yields the equivalent description $(a_1\,a_k\,a_{k-1}\,\ldots\,a_3\,a_2)$.

- Example: Find the inverses of $(1\,2\,3\,4\,5)$ and $(1\,5)(2\,4\,3)$ in S_5 and verify that the inverses compose with the originals to yield the identity.

  - First, we have $(1\,2\,3\,4\,5)^{-1} = (5\,4\,3\,2\,1) = \boxed{(1\,5\,4\,3\,2)}$. Indeed, $(1\,5\,4\,3\,2)(1\,2\,3\,4\,5) = (1)(2)(3)(4)(5) = e$ by tracing the results from right to left: $1 \to 2 \to 1$, $2 \to 3 \to 2$, $3 \to 4 \to 3$, $4 \to 5 \to 4$, and $5 \to 1 \to 5$.
  - For the inverse of $(1\,5)(2\,4\,3)$ we simply reverse the order of each cycle, and the order in which the cycles are multiplied, and then rearrange as needed: $[(1\,5)(2\,4\,3)]^{-1} = (5\,1)(3\,4\,2) = \boxed{(1\,5)(2\,3\,4)}$.
  - Indeed, $(1\,5)(2\,3\,4)\cdot(1\,5)(2\,4\,3) = (1)(2)(3)(4)(5) = e$ by tracing what happens to each of $1, 2, 3, 4, 5$ from right to left, as above.

4.1.4 Subgroups and Orders

- We have a natural notion of subgroup:

- Definition: If G is a group, we say a subset S of G is a subgroup if it also possesses the structure of a group, under the same operations as G.

  - Example: The set $(2\mathbb{Z}, +)$ of even integers under addition is a subgroup of $(\mathbb{Z}, +)$ because $(2\mathbb{Z}, +)$ is also a group: addition of even integers is associative, there is an additive identity 0, and the additive inverse of an even integer is also even.
  - Observe that if S is a subset of a group, in order for the operation $\star$ to be well-defined inside S, we must have $g \star h \in S$ for any $g, h \in S$.
  - Then axiom [G1] automatically holds in S, since it holds in G. In order for [G2] to hold in S, there must be an identity element e_S in S with the property that $ge_S = g$ for every $g \in S$. However, by the cancellation law in G, since $ge_S = g = ge_G$, we see that $e_S = e_G$: in other words, S must contain the identity element of G.
  - Finally, in order for [G3] to hold in S, we require that every $g \in S$ must have an inverse g_S^{-1}. Since $gg_S^{-1} = e_S = e_G = gg_G^{-1}$ by cancellation in G we must have $g_S^{-1} = g_G^{-1}$, which is to say, the inverse of g must be in S.

- Proposition (Subgroup Criterion): A subset S of G is a subgroup if and only if S contains the identity of G and is closed under the group operation of G and inverses. Equivalently, S is a subgroup if and only if $e_G \in S$ and for any $g, h \in S$, the element $gh^{-1} \in S$.

  - Proof: If S is a subgroup, then as noted above S must contain the identity of G and be closed under the group operation and inverses. Conversely, if S contains the identity of G and is closed under the group operation and inverses, then it is also a group.
  - For the second statement, if S is a subgroup then $e_G \in S$ and for any $g, h \in S$ we must have $h^{-1} \in S$ and then $gh^{-1} \in S$.
  - Conversely, if $e_G \in S$ and $gh^{-1} \in S$ for any $g, h \in S$, setting $g = e_G$ implies that $h^{-1} \in S$ so S is closed under inverses.
  - Then for any $k \in S$, setting $h = k^{-1}$ and using the fact that $(k^{-1})^{-1} = k$ implies that $gh^{-1} = gk \in S$ so S is closed under the group operation, hence is a subgroup.

- Using the subgroup criterion, we can construct additional examples of groups.

- Example: For any group G, the sets $\{e\}$ and G are always subgroups of G. The subgroup $\{e\}$ is called the trivial subgroup.
- Example: The set $\{3n : n \in \mathbb{Z}\} = \{\dots, -6, -3, 0, 3, 6, \dots\}$ under addition is a subgroup of $(\mathbb{Z}, +)$ since it satisfies the subgroup criterion.
- Example: The set of positive rational numbers under multiplication is a subgroup of $(\mathbb{C}\backslash\{0\}, \cdot)$ since it satisfies the subgroup criterion.
- Example: The set $\{2^n : n \in \mathbb{Z}\} = \{\dots, 2^{-2}, 2^{-1}, 1, 2, 4, 8, \dots,\}$ under multiplication is a subgroup of $(\mathbb{Q}^+, \cdot)$ since it satisfies the subgroup criterion.
- Example: The set $\{e, (1\,2)\}$ is a subgroup of S_3 since it satisfies the subgroup criterion. The set is closed under multiplication since $(1\,2)(1\,2) = e$ and it is closed under inverses since $(1\,2)^{-1} = (1\,2)$.
- Non-Example: The set $\{e, (1\,2\,3)\}$ is not a subgroup of S_3 since it is not closed under multiplication: the product $(1\,2\,3)(1\,2\,3) = (1\,3\,2)$ is not in the set. The set is also not closed under inverses, since $(1\,2\,3)^{-1} = (3\,2\,1) = (1\,3\,2)$ is also not in the set.
- Example: The set $\{e, (1\,2\,3), (1\,3\,2)\}$ is a subgroup of S_3 since it satisfies the subgroup criterion.
- Non-Example: The set $(\mathbb{Z}_{\geq 0}, +)$ of nonnegative integers under addition is not a subgroup of $(\mathbb{Z}, +)$ since it is not closed under additive inverses.
- Non-Example: The set of odd integers together with 0, under addition, is not a subgroup of $(\mathbb{Z}, +)$ since it is not closed under the group operation of addition.

- If g is an element of G, the set of powers of g, namely $\{\dots, g^{-2}, g^{-1}, e, g, g^2, \dots\}$ play an important role in understanding the behavior of multiplication by g.
- Definition: If g is an element of the group G, the subgroup generated by g is the set $\langle g\rangle = \{\dots, g^{-2}, g^{-1}, e, g, g^2, \dots\}$ of powers of g. The order of g, written $|g|$, is the order of this subgroup. Equivalently, the order of g is the smallest positive integer n such that $g^n = e$, if such an n exists, and otherwise (when $g^n \neq e$ for any positive integer n) the order of g is ∞.
  - If G is a finite group, then every element of G has finite order, since the set of powers $\{e, g, g^2, \dots\}$ must be finite, and if $g^a = g^b$ with $a < b$ then cancelling g^a yields $g^{b-a} = e$.
  - More generally, if $g^n = e$ for some $n > 0$, then the order of g divides n by an application of the division algorithm.
  - Example: The order of the identity element in any group is always 1.
  - Example: Inside $G = \{1, i, -1, -i\}$, the element -1 has order 2 since $(-1)^2 = 1$ but $-1 \neq 1$. Similarly, both i and $-i$ have order 4.
  - Example: Inside $(\mathbb{Z}, +)$, the order of every nonidentity element is ∞.
  - Example: Inside $(\mathbb{Z}/7\mathbb{Z}, +)$, the order of every nonidentity element is 7.
  - Example: Inside $(\mathbb{Z}/6\mathbb{Z}, +)$, the order of $\overline{2}$ is 3 since $\overline{2} + \overline{2} + \overline{2} = \overline{0}$ but $\overline{2} \neq \overline{0}$ and $\overline{2} + \overline{2} \neq \overline{0}$. In a similar way, the orders of $\overline{0}$, $\overline{1}$, $\overline{2}$, $\overline{3}$, $\overline{4}$, $\overline{5}$ are respectively 1, 6, 3, 2, 3, and 6.
  - Example: Inside $(\mathbb{Z}/11\mathbb{Z})^\times$, the powers of $\overline{2}$ are $\{\overline{2}, \overline{4}, \overline{8}, \overline{5}, \overline{10}, \overline{9}, \overline{7}, \overline{3}, \overline{6}, \overline{1}\}$. We see that $\overline{2}^{10} = \overline{1}$ but no lower power is equal to $\overline{1}$, so the order of $\overline{2}$ is 10 inside $(\mathbb{Z}/11\mathbb{Z})^\times$.
  - Example: Every nonidentity element in the group $(\mathbb{Z}/p\mathbb{Z})^n$, the Cartesian product of n copies of $\mathbb{Z}/p\mathbb{Z}$, has order p.
  - Example: In the dihedral group $D_{2\cdot n}$, since $r^n = e$ but $r^k \neq e$ for $0 < k < n$, we see that $|r| = n$. One may make a similar calculation to see more generally that the order of r^k is $n/\gcd(k, n)$.
  - Example: In $D_{2\cdot n}$, since $(sr^k)^2 = s(r^k s)r^k = s(sr^{-k})r^k = s^2 = e$, we see that the order of sr^k is 2 for any k.
  - Example: In the symmetric group S_n, the order of any n-cycle $\sigma = (a_1\, a_2\, \dots\, a_n)$ is n, since $\sigma^n = 1$, but $\sigma^k(a_1) = a_k$ (so $\sigma^k \neq e$) for $1 \leq k \leq n-1$.
  - More generally, in S_n, if a lies in a k-cycle for the permutation τ, then $\tau^n(a) = a$ only when k divides n by the same argument as above. Thus, the order of τ is the least common multiple of the lengths of the cycles in its cycle decomposition.

- Example: The six elements e, $(1\,2)$, $(1\,3)$, $(2\,3)$, $(1\,2\,3)$, $(1\,3\,2)$ in S_3 have respective orders 1, 2, 2, 2, 3, 3.
- Example: The element $\tau = (1\,3\,5)(2\,6)$ in S_6 has order 6. Indeed, the powers of τ are $\tau^2 = (1\,5\,3)$, $\tau^3 = (2\,6)$, $\tau^4 = (1\,3\,5)$, $\tau^5 = (1\,5\,3)(2\,6)$, and $\tau^6 = 1$, so τ indeed has order 6.

- The existence of elements having a particular order in G can be a bit difficult to characterize. Even when the order of G is composite it is possible that all its nonidentity elements have prime order, such as the case of S_3 above, so the most we could hope for in general is for the existence of elements of prime order. In fact, we do have such a result:

- Theorem (Cauchy's Theorem): Suppose G is a group and p is a prime dividing $\#G$. Then there exists an element of G of order p.
  - Proof: Consider the set S of ordered p-tuples of elements $(g_1, g_2, \ldots, g_p)$ in G such that $g_1 g_2 \cdots g_p = e$. Since such a tuple is characterized by having $g_p = (g_{p-1} \cdots g_2 g_1)^{-1}$, we can choose $g_1, g_2, \ldots, g_{p-1}$ arbitrarily and then g_p is determined.
  - Therefore there are exactly $(\#G)^{p-1}$ such p-tuples, so in particular the cardinality of S is divisible by p.
  - Now we define an equivalence relation on these p-tuples by saying that $(g_1, \ldots, g_p) \sim (h_1, \ldots, h_p)$ if we may apply a cyclic permutation to $(g_1, \ldots, g_p)$ that yields $(h_1, \ldots, h_p)$.
  - Indeed, if $(g_1, g_2, \ldots, g_p) \in S$ then any cyclic permutation, such as $(g_2, \ldots, g_p, g_1)$, is also in S. If not all the elements in the tuple are equal, then there are p distinct cyclic permutations of this tuple in S, while if all elements are equal there is only 1, namely $(g, g, \ldots, g)$.
  - Thus, since $\#S$ is divisible by p, and the number of tuples of the first type is divisible by p, the number of tuples of the second type must be divisible by p. In particular, there must be at least one tuple $(g, g, \ldots, g)$ with $g \neq e$: then $g^p = e$ so g is an element of order p.

4.1.5 Group Isomorphisms

- Some of the groups we have already described have very similar-looking structures, even though the actual sets and operations themselves are different. For example, compare the structure of the additive group $\mathbb{Z}/2\mathbb{Z}$ to the multiplicative group $\{1, -1\}$:

$+$	$\overline{0}$	$\overline{1}$
$\overline{0}$	$\overline{0}$	$\overline{1}$
$\overline{1}$	$\overline{1}$	$\overline{0}$

$\cdot$	1	-1
1	-1	1
-1	1	-1

  - Both groups have an identical structure: in each group we have an identity element and a nonidentity element, and the composition rules are also the same (the identity composed with any element gives that element back, and the nonidentity composed with itself gives the identity).
  - We can see that if we "relabel" the two residue classes $\overline{0}$ and $\overline{1}$ with the numbers 1 and -1, and correspondingly convert the operation $+$ on $\mathbb{Z}/2\mathbb{Z}$ to the operation $\cdot$ on $\{1, -1\}$, then the first group becomes the second group. Likewise, by using the inverse labeling, we can convert the second group into the first group.
  - More formally, we can phrase this using the language of functions: for the function $f : (\mathbb{Z}/2\mathbb{Z}) \to \{-1, 1\}$ with $f(\overline{0}) = 1$ and $f(\overline{1}) = -1$, we have $f(\overline{a} + \overline{b}) = f(\overline{a}) \cdot f(\overline{b})$ for all residue classes $\overline{a}, \overline{b} \in \mathbb{Z}/2\mathbb{Z}$. Note here that f is a bijection, as well, so that each element of the first group corresponds to a unique element of the second group.

- For another example, consider the dihedral group $D_{2\cdot 3}$ and the symmetric group S_3 (the multiplications are done as (row label) times (column label)):

$D_{2\cdot 3}$	e	r	r^2	s	sr	sr^2
e	e	r	r^2	s	sr	sr^2
r	r	r^2	e	sr^2	s	sr
r^2	r^2	e	r	sr	sr^2	s
s	s	sr	sr^2	e	r	r^2
sr	sr	sr^2	s	r^2	e	r
sr^2	sr^2	s	sr	r	r^2	e

S_3	e	$(1\,2\,3)$	$(1\,3\,2)$	$(1\,2)$	$(2\,3)$	$(1\,3)$
e	e	$(1\,2\,3)$	$(1\,3\,2)$	$(1\,2)$	$(2\,3)$	$(1\,3)$
$(1\,2\,3)$	$(1\,2\,3)$	$(1\,3\,2)$	e	$(1\,3)$	$(1\,2)$	$(2\,3)$
$(1\,3\,2)$	$(1\,3\,2)$	e	$(1\,2\,3)$	$(2\,3)$	$(1\,3)$	$(1\,2)$
$(1\,2)$	$(1\,2)$	$(2\,3)$	$(1\,3)$	e	$(1\,2\,3)$	$(1\,3\,2)$
$(2\,3)$	$(2\,3)$	$(1\,3)$	$(1\,2)$	$(1\,3\,2)$	e	$(1\,2\,3)$
$(1\,3)$	$(1\,3)$	$(1\,2)$	$(2\,3)$	$(1\,2\,3)$	$(1\,3\,2)$	e

- One may check (carefully!) that if we replace e with e, r with $(1\,2\,3)$, r^2 with $(1\,3\,2)$, s with $(1\,2)$, sr with $(2\,3)$, and sr^2 with $(1\,3)$, then the multiplication table for the dihedral group $D_{2\cdot 3}$ on the left turns into the multiplication table for S_3 on the right.
- In other words, this bijection $f : D_{2\cdot 3} \to S_3$ preserves the composition operation in the two groups, meaning that $f(g_1 g_2) = f(g_1)f(g_2)$ for all elements g_1 and g_2 in $D_{2\cdot 3}$.
- In fact, there is a pleasant way to understand where this bijection comes from: as we saw earlier, the symmetry group of the equilateral triangle is the dihedral group $D_{2\cdot 3}$.
- Now consider the actions of the various symmetries on the set $\{1, 2, 3\}$ of the three vertices of the triangle: certainly each symmetry yields a permutation of the vertices, and different symmetries permute the vertices differently, so we equally well view the set of symmetries as a subgroup of S_3. But since there are six symmetries and six permutations in S_3, in fact the set of symmetries coincides with the group S_3 as well.
- We obtain the explicit bijection f above simply by writing down the permutation on the vertices obtained by applying each of the possible symmetries of the triangle.

- Motivated by the examples above, we can now formalize the notion of when two groups have identical structures:

- Definition: Let $(G, \star)$ and $(H, \circ)$ be groups. A group isomorphism φ from G to H is a bijective function $\varphi : G \to H$ such that $\varphi(g_1 \star g_2) = \varphi(g_1) \circ \varphi(g_2)$ for all g_1 and g_2 in G. If there is an isomorphism $\varphi : G \to H$, we say G and H are isomorphic, and write $G \cong H$.
  - We will often suppress the notation for the group operations and write the condition simply as $\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2)$.
  - Example: For $G = (\mathbb{R}, +)$ and $H = (\mathbb{R}^+, \cdot)$, the map $\varphi : G \to H$ defined via $\varphi(x) = e^x$ is an isomorphism from G to H. The map respects the group operation since $e^{x+y} = e^x e^y$, and it is a bijection since it has an inverse map $\varphi^{-1}(x) = \ln(x)$.
  - Example: For $G = \mathbb{Z}/6\mathbb{Z}$ and $H = (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$, the map $\varphi : G \to H$ defined via $\varphi(n \bmod 6) = (n \bmod 2,\, n \bmod 3)$ is an isomorphism of groups.

- Here are some fundamental properties of isomorphisms:

- Proposition (Properties of Isomorphisms): If G, H, K are any groups, the following hold:

1. The identity map $i_G : G \to G$ defined by $i_G(g) = g$ for all $g \in G$ is an isomorphism from G to G.
   - Proof: Clearly i_G is a bijection and respects the group operation.
2. If $\varphi : G \to H$ is an isomorphism, then the inverse map $\varphi^{-1} : H \to G$ is also an isomorphism.
   - Proof: We have previously shown that the inverse function of a bijection is also a bijection.
   - Now suppose $\varphi^{-1}(h_1) = g_1$ and $\varphi^{-1}(h_2) = g_2$, so that $\varphi(g_1) = h_1$ and $\varphi(g_2) = h_2$.
   - Then $\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2) = h_1 h_2$, meaning that $\varphi^{-1}(h_1 h_2) = g_1 g_2 = \varphi^{-1}(h_1)\varphi^{-1}(h_2)$, so φ^{-1} is also an isomorphism.
3. If $\varphi : G \to H$ and $\psi : H \to K$ are isomorphisms, then the composition $\psi\varphi : G \to K$ is also an isomorphism.
   - Proof: The composition of two bijections is a bijection, and also $(\psi\varphi)(g_1 g_2) = \psi(\varphi(g_1 g_2)) = \psi(\varphi(g_1)\varphi(g_2)) = \psi\varphi(g_1)\psi\varphi(g_2)$, so $\psi\varphi$ is an isomorphism.
4. The isomorphism relation $\cong$ on groups is an equivalence relation.
   - Proof: (1) gives reflexivity, (2) gives symmetry, and (3) gives transitivity.
5. Isomorphisms map the identity to the identity: if $\varphi : G \to H$ is an isomorphism then $\varphi(e_G) = e_H$.
   - Proof: We have $\varphi(e_G)\varphi(g) = \varphi(e_G g) = \varphi(g) = e_H \varphi(g)$, so cancelling $\varphi(g)$ yields $e_H = \varphi(e_G)$.
6. Isomorphisms preserve powers and orders: if $\varphi : G \to H$ is an isomorphism and $g \in G$, then $\varphi(g^n) = \varphi(g)^n$ and $|g| = |\varphi(g)|$.

- Proof: For $n \geq 0$ the statement $\varphi(g^n) = \varphi(g)^n$ follows by induction using $\varphi(g^n) = \varphi(g^{n-1}g) = \varphi(g^{n-1})\varphi(g) = \varphi(g)^n$.
- For the second statement, by (5) and the above we see that $g^n = e_G$ if and only if $\varphi(g)^n = e_H$, so g and $\varphi(g)$ must have the same order.

- In order to show that two given groups are isomorphic, we essentially need to construct an isomorphism between them, which can often be difficult to do[4]. Even if we are handed an isomorphism, actually verifying that it is an isomorphism can be very time-consuming.
  - On the other hand, it is often easier to show that two given groups cannot be isomorphic to one another, if one of the properties of isomorphisms above fails.
  - For example, the group $D_{2\cdot 4}$ is not isomorphic to S_3, because the former has order 8 and the latter has order 6, and so there cannot even exist a bijection between their underlying sets of elements.
  - In a similar way we can see that $D_{2\cdot 4}$ is not isomorphic to $\mathbb{Z}/8\mathbb{Z}$, because the latter is abelian and the former is not; likewise, S_3 is not isomorphic to $\mathbb{Z}/6\mathbb{Z}$.
  - Also, $\mathbb{Z}/8\mathbb{Z}$ is not isomorphic to $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/4\mathbb{Z})$, because the former has an element of order 8 (namely $\overline{1}$) while the latter does not have any elements of order 8.

- A fundamental goal of group theory is to classify (up to isomorphism) all of the groups of a given order.
  - One can show, for example, that the two groups $D_{2\cdot 3} \cong S_3$ and $\mathbb{Z}/6\mathbb{Z} \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$ are the only groups of order 6, up to isomorphism.
  - But even to prove this fact only from the results we have developed so far is rather difficult. To progress further, we need some additional tools.

4.1.6 Cosets of Subgroups and Lagrange's Theorem

- Definition: If H is a subgroup of G and $a \in G$, the set $aH = \{ah : h \in H\}$ is called a left coset of H. We also define the index of H in G, denoted $[G : H]$, to be the number of distinct left cosets of H in G.
  - We also have a symmetric notion of $Ha = \{ha : h \in H\}$, which is called a right coset of H. If G is abelian, then left and right cosets are the same, but when G is non-abelian, this need not be the case. We will see in a moment that the definition of the index is independent of whether we use left or right cosets.
  - Example: If $H = \{e, r^2\}$ in $G = D_{2\cdot 4}$, then there are four left cosets of H in G, namely $eH = r^2H = \{e, r^2\}$, $rH = r^3H = \{r, r^3\}$, $sH = sr^2H = \{s, sr^2\}$, and $srH = sr^3H = \{sr, sr^3\}$.
  - Example: If $H = \{1, (1\,2\,3), (1\,3\,2)\}$ in $G = S_3$, then there are two left cosets of H in G, so $[G : H] = 2$. Explicitly, these cosets are $1H = (1\,2\,3)H = (1\,3\,2)H = \{1, (1\,2\,3), (1\,3\,2)\}$ and $(1\,2)H = (1\,3)H = (2\,3)H = \{(1\,2), (1\,3), (2\,3)\}$.
  - Example: If $H = \{1, (1\,3)\}$ in $G = S_3$, then there are three left cosets of H in G, so $[G : H] = 3$. Explicitly, these cosets are $1H = (1\,3)H = \{1, (1\,3)\}$, $(1\,2)H = (1\,3\,2)H = \{(1\,2), (1\,3\,2)\}$, and $(2\,3)H = (1\,2\,3)H = \{(2\,3), (1\,2\,3)\}$.
  - Example: If $H = 2\mathbb{Z} = \{\ldots, -2, 0, 2, 4, \ldots\}$ in $G = \mathbb{Z}$, then there are two (left) cosets of H in G, so $[G : H] = 2$. These cosets are $0 + H = \{\ldots, -2, 0, 2, 4, \ldots\}$ and $1 + H = \{\ldots, -3, 1, 3, 5, \ldots\}$.

- In each of the examples above, all of the left cosets have the same size (which is then the same size as $eH = H$), and the left cosets form a partition of G. This is true in general:

- Proposition (Properties of Cosets): Let H be a subgroup of G. Then the following hold:

1. For any $a \in G$, the map $f : H \to aH$ defined by $f(h) = ah$ is a bijection between H and gH.

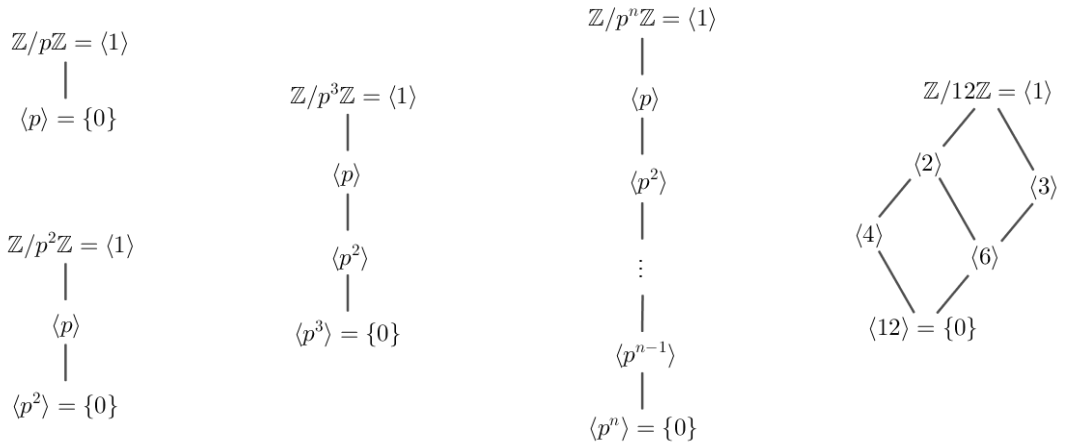
[4] In fact, it has been shown that the isomorphism problem for groups (given two groups, decide whether or not they are isomorphic) is undecidable, in the sense that there exists no algorithm that is guaranteed to determine whether two arbitrary groups are isomorphic in a finite amount of time. Of course, the isomorphism problem can always be solved for two finite groups in finite time, since there are only finitely many possible bijections from one to the other, but the general philosophy is still that this is a hard question!

- Proof: By definition of aH, the map f is surjective. On the other hand, $f(h_1) = f(h_2)$ is equivalent to $ah_1 = ah_2$, which by cancellation implies $h_1 = h_2$: thus, f is also injective, hence it is a bijection.

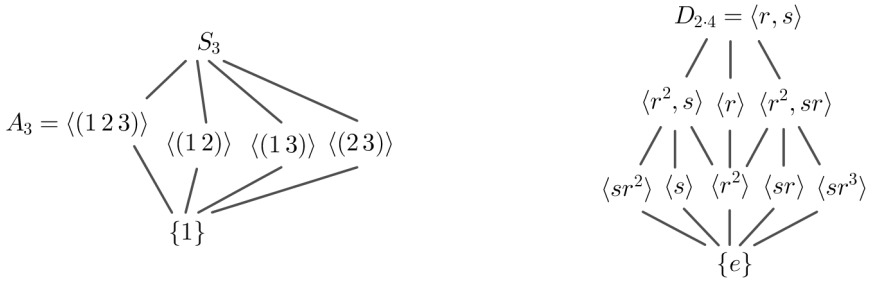
2. For any $a \in G$, the only left coset of H containing a is aH.
   - Proof: Clearly aH is a left coset of H containing a since $e \in H$, so we need to show it is the only one.
   - If $a \in bH$ then by definition $a = bh$ for some $h \in H$.
   - Then for any $h' \in H$, since $hh' \in H$ because H is a subgroup, we see that $ah' = b(hh') \in bH$. Thus bH contains aH.
   - On the other hand, for any $bh'' \in bH$, since $b = ah^{-1}$ we can write $bh'' = a(h^{-1}h'') \in aH$ because $h^{-1}h'' \in H$ again because H is a subgroup. Thus, aH contains bH, so they are equal.
3. Any two left cosets of H in G are either disjoint or identical. Thus, the left cosets of H in G partition G.
   - Proof: Suppose aH and bH are left cosets of H. If they are disjoint we are done, so suppose they have some common element g.
   - But then by (2), this means $aH = gH = bH$, so $aH = bH$. The other statement is immediate since any $g \in G$ is contained in the left coset gH.
4. For any $a, b \in G$, we have $aH = bH$ if and only if $a^{-1}b \in H$.
   - Proof: If $aH = bH$ then since $b \in aH$ this means $b = ah$ for some $h \in H$: then $a^{-1}b = a^{-1}ah = h \in H$.
   - Conversely, if $a^{-1}b \in H$, then $b = ah$ for some $h \in H$, and so $b \in aH$. Then by (2), this means $bH = aH$.

- These properties seem rather simple, but we can deduce a very important consequence from them:
- Theorem (Lagrange's Theorem): If H is a subgroup of G, then $\#G = \#H \cdot [G : H]$, where if one side is infinite then both are. In particular, if G is a finite group, then the order of any subgroup H divides the order of G.
  - Proof: By our properties of cosets, each left coset of H has a bijection with H, and so all of the left cosets have the same cardinality.
  - Since the left cosets form a partition of G, we may partition the $\#G$ elements into a total of $[G : H]$ left cosets each of which has size $\#H$.
  - Thus, $\#G = \#H \cdot [G : H]$. The second statement follows immediately from this relation, since $[G : H]$ is an integer.
  - Remark: If we work with right cosets instead of left cosets, we obtain the same formula: thus, the number of left cosets is equal to the number of right cosets.
- Corollary (Orders of Elements): If G is a finite group of order n, then for every $g \in G$ the order of g divides n, and $g^n = e$.
  - Proof: Suppose g has order k and let $H = \{\dots, g^{-2}, g^{-1}, e, g, g^2, \dots\} = \{e, g, g^2, \dots, g^{k-1}\}$ be the subgroup of G consisting of all powers of g. Then H is a subgroup of G since it is closed under multiplication and inverses, and it has order k.
  - Thus by Lagrange's theorem, k, the order of H, divides n. The second statement follows immediately.
- Although its proof is seemingly easy, Lagrange's theorem is an extremely important tool in unraveling the structure of groups (particularly, finite groups) since it substantially narrows the possible orders for elements and subgroups of G.
  - A convenient way to organize this information is by drawing the subgroup lattice of G (more formally called the Hasse diagram of G): we arrange all of the subgroups of G starting with the smallest subgroups at the bottom, and then draw paths to indicate immediate containments.

- For $\mathbb{Z}/n\mathbb{Z}$ the subgroups are in bijection with the divisors of n, and $\langle a \rangle$ is contained in $\langle b \rangle$ precisely when a divides b. Here are a few examples of the resulting subgroup lattices:



- Here are subgroup lattices for some of the other small groups we have described:



4.2 Fields

- We now give a brief discussion of fields with the goal of describing the special properties of the real numbers and the complex numbers.

4.2.1 The Formal Definition of a Field

- Definition: A field is any set F having two (closed) binary operations $+$ and $\cdot$ that satisfy the nine axioms [F1]-[F9]:

  [F1] The operation $+$ is associative: $a + (b + c) = (a + b) + c$ for any elements a, b, c in F.

  [F2] The operation $+$ is commutative: $a + b = b + a$ for any elements a, b in F.

  [F3] There is an additive identity 0 satisfying $a + 0 = a$ for all a in F.

  [F4] Every element a in F has an additive inverse $-a$ satisfying $a + (-a) = 0$.

  [F5] The operation $\cdot$ is associative: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for any elements a, b, c in F.

  [F6] The operation $\cdot$ is commutative: $a \cdot b = b \cdot a$ for any elements a, b in F.

  [F7] There is a multiplicative identity $1 \neq 0$, satisfying $1 \cdot a = a = a \cdot 1$ for all a in F.

  [F8] Every nonzero a in F has a multiplicative inverse a^{-1} satisfying $a \cdot a^{-1} = 1$.

  [F9] The operation $\cdot$ distributes over $+$: $a \cdot (b + c) = a \cdot b + a \cdot c$ for any elements a, b, c in F.

  - For convenience, in a field F we can also define the operations of subtraction via $a - b = a + (-b)$ and division via $a/b = a \cdot b^{-1}$ (the latter whenever $b \neq 0$).

- Example: The set $\mathbb{Q}$ of rational numbers is a field.
  - We established all of these properties of $\mathbb{Q}$ when we described the elements of $\mathbb{Q}$ as equivalence classes of fractions a/b for integers a and b with $b \neq 0$.
- Non-Example: The set $\mathbb{Z}$ of integers is not a field.
  - Although eight of the nine properties hold for $\mathbb{Z}$, property [F8] does not, because there are many nonzero elements of $\mathbb{Z}$, such as 2 and 3, that do not have a multiplicative inverse in $\mathbb{Z}$.
- Example: The set $\mathbb{R}$ of real numbers is a field, as is the set $\mathbb{C}$ of complex numbers.
  - Again, as with $\mathbb{Q}$, the real numbers and complex numbers have the property that every nonzero element has a multiplicative inverse.
- Example: If p is a prime number, the set $\mathbb{Z}/p\mathbb{Z}$ of residue classes modulo p is a field.
  - Unlike the other examples of fields above, this field only has finitely many elements: they are the p residue classes $\overline{0}$, $\overline{1}$, ... , $\overline{p-1}$.
- Example: The set $S = \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$ forms a field, denoted $\mathbb{Q}(\sqrt{2})$ (typically read as "$\mathbb{Q}$ adjoin $\sqrt{2}$").
  - The arithmetic in $\mathbb{Q}(\sqrt{2})$ is as follows: $(a+b\sqrt{2})+(c+d\sqrt{2}) = (a+c)+(b+d)\sqrt{2}$, and $(a+b\sqrt{2})(c+d\sqrt{2}) = (ac+2bd)+(ad+bc)\sqrt{2}$.
  - The associative and commutative properties of addition and multiplication, and the distributive law, are all inherited from $\mathbb{R}$. The additive identity is $0 = 0 + 0\sqrt{2}$, the multiplicative identity is $1 = 1 + 0\sqrt{2}$, and the additive inverse of $a + b\sqrt{2}$ is $-a - b\sqrt{2}$.
  - Finally, we need to show that every nonzero element has a multiplicative inverse. We can do this by rationalizing the denominator: explicitly, we have $\dfrac{1}{a+b\sqrt{2}} = \dfrac{1}{a+b\sqrt{2}} \cdot \dfrac{a-b\sqrt{2}}{a-b\sqrt{2}} = \dfrac{a-b\sqrt{2}}{a^2-2b^2}$.
  - Since $\sqrt{2}$ is irrational, as long as one of a, b is nonzero, the expression $a^2 - 2b^2$ is a nonzero rational number, so we obtain an inverse $(a + b\sqrt{2})^{-1} = \dfrac{a}{a^2-2b^2} - \dfrac{b}{a^2-2b^2}\sqrt{2}$.
- Notice that if F is a field, $(F, +)$ is an abelian group, as is $(F\backslash\{0\}, \cdot)$.
  - Therefore, all of the basic properties of abelian groups yield basic properties of field arithmetic. Here are some such properties:
- Proposition (Basic Field Arithmetic): Let F be a field. The following properties hold in F:
  1. The additive identity 0 is unique, as is the multiplicative identity 1.
  2. Addition has a cancellation law: for any $a, b, c \in F$, if $a + b = a + c$, then $b = c$.
  3. Additive inverses are unique.
  4. For any $a \in F$, $0 \cdot a = 0 = a \cdot 0$.
  5. For any $a \in F$, $-(-a) = a$.
  6. For any $a \in F$, $(-1) \cdot a = -a = a \cdot (-1)$.
  7. For any $a, b \in F$, $-(a + b) = (-a) + (-b)$.
  8. For any $a, b \in F$, $(-a) \cdot b = -(a \cdot b) = a \cdot (-b)$, and $(-a) \cdot (-b) = a \cdot b$.
  9. For any $a, b \in F$, if $a \cdot b = 0$ then $a = 0$ or $b = 0$.
     - Proofs: These follow from the field axioms using similar calculations to the ones we gave for properties of arithmetic in $\mathbb{Z}$ and in groups.

4.2.2 Ordered Fields

- The rational numbers and real numbers have a familiar ordering, which we can formalize by identifying the special subset consisting of "positive elements" of the field.
- Definition: An ordered field is a field $(F, +, \cdot)$ along with a subset P (the "positive elements" of F) with the following properties:

  [O1] For every $a \in F$, precisely one of the following holds: $a \in P$, $a = 0$, or $(-a) \in P$.

  [O2] The set P is closed under addition: if $a, b \in P$ then $a + b \in P$.

  [O3] The set P is closed under multiplication: if $a, b \in P$ then $a \cdot b \in P$.

- Example: The rational numbers $\mathbb{Q}$ are an ordered field upon taking P to be the set of positive rational numbers.
  - More explicitly, using the definition of $\mathbb{Q}$ as collections of equivalence classes of fractions $[a/b]$, we can define P to be the set of equivalence classes of fractions $[a/b]$ where both a and b are both positive integers.
- Example: The real numbers $\mathbb{R}$ are an ordered field upon taking P to be the set of positive real numbers.
- Non-Example: The complex numbers $\mathbb{C}$ are not an ordered field for any choice of subset P.
  - By [O1], since $i \neq 0$ either i or $-i$ would have to be in P.
  - But because $i \cdot i \cdot i = -i$ and $(-i) \cdot (-i) \cdot (-i) = i$, [O3] would force both i and $-i$ to be in P, but this contradicts [O1].
- Non-Example: The integers modulo p are not an ordered field for any choice of subset P.
  - By [O1], either $\overline{1}$ or $-\overline{1}$ would have to be in P.
  - But then by [O2] either $\underbrace{\overline{1} + \overline{1} + \cdots + \overline{1}}_{p \text{ terms}} = \overline{0}$ or $\underbrace{(-\overline{1}) + (-\overline{1}) + \cdots + (-\overline{1})}_{p \text{ terms}} = \overline{0}$ would be in P, but either way this contradicts [O1].
- Using the ordering on an ordered field, we can define the various inequality symbols:
- Definition: If F is an ordered field and $a, b \in F$, we write $a < b$ (equivalently $b > a$) when $b - a \in P$, and we write $a \leq b$ (equivalently, $b \geq a$) when $b - a \in P$ or $b - a = 0$.
- We then have the following basic properties of inequalities:
- Proposition (Basic Ordered Field Arithmetic): Let F be an ordered field and $a, b \in F$.

  1. Exactly one of $a < b$, $a = b$, $a > b$ is true.
  2. If $a > 0$ and $b > 0$ then $a + b > 0$ and $ab > 0$.
  3. If $a < b$ then $a + c < b + c$ for any $c \in F$.
  4. If $a < b$ and $c > 0$ then $ac < bc$.
  5. If $a < b$ and $b < c$ then $a < c$.
  6. If $a > 0$ then $ab > 0$ if and only if $b > 0$, and if $a < 0$ then $ab > 0$ if and only if $b < 0$.
  7. For any $a \neq 0$ it is true that $a^2 > 0$. In particular, $1 > 0$.
     - Proofs: These follow from the ordered field axioms. For example, (1) is merely a rewriting of [O1], while (2) is a rewriting of [O2] and [O3]. Items (3), (4), and (5) follow from manipulating [O2] and [O3] appropriately, while (6) follows by breaking into cases based on whether $b > 0$, $b = 0$, or $b < 0$.
     - For (7), note that if $a \neq 0$ then by [O1] either $a \in P$ or $(-a) \in P$. But since $a \cdot a = a^2 = (-a) \cdot (-a)$, either way [O3] implies that $a^2 \in P$, meaning that $a^2 > 0$. Then since $1^2 = 1$ we see $1 > 0$.

4.2.3 Least Upper Bounds and the Real Numbers

- Our final goal is to characterize the field of real numbers by an additional special property of their ordering known as the least upper bound axiom.
- Definition: Suppose F is an ordered field and S is a subset of F. We say an element $x \in F$ is an upper bound for S if $s \leq x$ for all $s \in S$. If S has some upper bound $x \in F$, we say that S is bounded above.
  - We remark that an upper bound for S need not be an element of S itself, it only needs to be an element of F that is greater than or equal to all elements of S.
  - Example: In $\mathbb{Q}$, the set $S = \{1, 2, 3, 4, 5\}$ has an upper bound $x = 5$, since $s \leq 5$ is true for all $s \in S$. The element $x = 6$ is also an upper bound for S, since $s \leq 6$ is also true for all $s \in S$.
  - Example: In $\mathbb{Q}$, the set $S = \{\frac{n}{n+1} : n \in \mathbb{Z}_{>0}\} = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots\}$ has an upper bound $x = 1$, since $\frac{n}{n+1} \leq 1$ is true for all positive integers n.
  - Example: In $\mathbb{Q}$, the set $S = \{r \in \mathbb{Q} : r^2 \leq 2\}$ has an upper bound $x = 2$, since if $r > 2$ then $r^2 > 4$, so taking the contrapositive shows that if $r^2 \leq 2$ then $r \leq 2$. In fact, any rational number $x \geq \sqrt{2}$ is an upper bound for S.
  - Example: In $\mathbb{Q}$, the set $S = \mathbb{Z}$ has no upper bound, since there is no rational number x such that $n \leq x$ for all integers $n \in \mathbb{Z}$.
  - Example: In $\mathbb{R}$, the set $S = \{x : 0 < x < 1\}$, the open interval $(0, 1)$, has an upper bound $x = 1$, since $s \leq 1$ is true for all $s \in S$.
- To show that a set S is bounded above, we need only give some upper bound for S. Of course, any larger element is then also an upper bound (by transitivity), so the most useful upper bound on a set would be the smallest possible one.
- Definition: Suppose F is an ordered field and S is a subset of F that is bounded above. We say that $x \in F$ is a least upper bound if x is an upper bound of S, and x is the smallest upper bound: namely, if y is any other upper bound, then $x \leq y$.
  - Equivalently, if we consider the set U of all upper bounds of S, then a least upper bound is a smallest element of U, if one exists. We saw earlier in our discussion of smallest elements that there is at most one smallest element in any partially ordered set.
  - Example: In $\mathbb{Q}$, the set $S = \{1, 2, 3, 4, 5\}$ has least upper bound 5, since 5 is an upper bound, and any other upper bound y must satisfy $5 \leq y$.
  - Example: In $\mathbb{Q}$, the set $S = \{\frac{n}{n+1} : n \in \mathbb{Z}_{>0}\} = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \dots\}$ has least upper bound 1. As noted above, 1 is an upper bound. Any smaller upper bound would necessarily be of the form $1 - r$ for some positive rational number r, say $r = \frac{p}{q}$. But then $1 - r = 1 - \frac{p}{q} \leq 1 - \frac{1}{q} < 1 - \frac{1}{2q} = \frac{2q-1}{2q}$, which is a contradiction because the number $\frac{2q-1}{2q}$ is an element of S that is larger than this purported upper bound $1 - r$. So there is no upper bound of S that is less than 1, so 1 is the least upper bound of S. Note here that 1 is not actually an element of S, but that the elements of S approach 1 "arbitrarily closely" from below.
  - Example: In $\mathbb{Q}$, the set $S = \{r \in \mathbb{Q} : r^2 \leq 2\}$ has no least upper bound. The set of upper bounds of S is all rational numbers x with $x \geq \sqrt{2}$, and this set has no smallest element since $\sqrt{2}$ is irrational (as we proved using prime factorizations), but can be approximated arbitrarily closely from above by rational numbers (using truncations of its decimal expansion rounded upward, for instance).
  - Example: In $\mathbb{R}$, the set $S = \{r \in \mathbb{Q} : r^2 \leq 2\}$ has a least upper bound, namely $x = \sqrt{2}$. Like in the previous example, the set of upper bounds of S is all real numbers x with $x \geq \sqrt{2}$, but now this set does have a smallest element since $\sqrt{2}$ is a real number.
  - Example: In $\mathbb{R}$, the empty set $\emptyset$ is bounded above, since any real number is an upper bound. However, $\emptyset$ has no least upper bound: precisely because *any* real number is an upper bound, there is no smallest upper bound.

- In the last few examples, we can see an important difference between least upper bounds in $\mathbb{Q}$ and in $\mathbb{R}$: there are some sets of rational numbers that are bounded above but do not have a least upper bound. However, when we pass to $\mathbb{R}$, theose issues disappeared.
    - In fact, we found a subset of $\mathbb{R}$ that was bounded above but had no least upper bound: the empty set.
    - The miraculous fact is that this is the *only* subset of $\mathbb{R}$ that is bounded above with no least upper bound.
    - We formalize this as follows:
- Definition: An ordered field F is complete if it satisfies the following axiom:

  [C] If S is a nonempty subset of F that is bounded above, then S has a least upper bound.
- Theorem (Characterization of $\mathbb{R}$): If F is a complete ordered field, then F is simply the real numbers up to a relabeling of the elements. More precisely, there exists a bijection $f : F \to \mathbb{R}$ that preserves addition, multiplication, and orderings, in the sense that for any $a, b \in F$ it is true that $f(a + b) = f(a) + f(b)$, $f(a \cdot b) = f(a) \cdot f(b)$, and if $a < b$ then $f(a) < f(b)$.
    - In other words, this theorem says that the least upper bound axiom characterizes the real numbers, in that the real numbers are the only ordered field satisfying the least upper bound axiom, up to a relabeling of the elements.
    - The function f is what is known as an isomorphism: a function that preserves all of the relevant algebraic properties of the object under study (in this case, an ordered field).
- The least upper bound axiom is incredibly useful in developing calculus and (more abstractly) mathematical analysis, since it holds the key to understanding the notion of a limit and the closely related notion of a continuous function.
    - Intuitively, the least upper bound axiom ensures that there are no "holes" in the real numbers, in contrast to $\mathbb{Q}$ which is "missing" elements like $\sqrt{2}$ that arise naturally as least upper bounds.
- We can also use the underlying idea of the least upper bound axiom to give a construction of the real numbers from the rational numbers, as follows:
    - For each real number α, consider the set $S_\alpha = \{r \in \mathbb{Q} : r < \alpha\}$ of rational numbers less than α. Then α is the least upper bound of S_α. So if we can characterize these sets S_α in $\mathbb{Q}$, we can reverse the process and use a set S_α to "define" a real number α.
    - Each set S_α is a nonempty proper subset of $\mathbb{Q}$ with no largest element. Also, if x is rational and $x \in S_\alpha$ then for any rational $y < x$ we have $y \in S_\alpha$.
    - In fact, these properties characterize the sets S_α, which are called Dedekind cuts since they "cut" the rational numbers into two pieces (one set S_α consisting of all numbers below the cut, and the other S_α^c consisting of all numbers above the cut).
- Starting with this description of the sets S_α (nonempty proper subsets of $\mathbb{Q}$ that are "closed below"), we can then define how to add, multiply, and order the S_α, which provides an construction of the real numbers from the rational numbers.
    - Explicitly, we define the sum $S_\alpha + S_\beta = \{x + y : x \in S_\alpha \text{ and } y \in S_\beta\}$, along with the additive identity $S_0 = \{x \in \mathbb{Q} : x < 0\}$ and the slightly trickier additive inverse $S_{-\alpha} = \{x \in \mathbb{Q} : -x \notin S_\alpha \text{ and } -x \text{ is not the least element of } S_\alpha^c\}$. One may then directly verify the field axioms [F1]-[F4].
    - Next we define the order relation as $S_\alpha < S_\beta$ when $S_\alpha$is a proper subset of S_β, and verify the order axioms [O1]-[O2].
    - Then we define multiplication by writing $S_\alpha \cdot S_\beta = S_{-\alpha} \cdot S_{-\beta} = \{x \cdot y : x \in S_\alpha \text{ and } y \in S_\beta\} \cup \{z \in \mathbb{Q} : z \leq 0\}$ when $S_0 \leq S_\alpha, S_\beta$, and also set $S_{-\alpha} \cdot S_\beta = S_\alpha \cdot S_{-\beta}$ as the additive inverse set of $S_\alpha \cdot S_\beta$.
    - We also take the multiplicative identity $S_1 = \{x \in \mathbb{Q} : x < 1\}$ and multiplicative inverse $S_{\alpha^{-1}} = \{1/x : x > 0 \text{ and } 1/x \notin S_\alpha \text{ and } 1/x \text{ is not the least element of } S_\alpha^c\} \cup \{z \in \mathbb{Q} : z \leq 0\}$ for $S_0 < S_\alpha$ and $S_{-\alpha^{-1}}$ as the additive inverse set of $S_{\alpha^{-1}}$.

- Using these definitions we can (with suitable tedious casework) we verify the remaining field axioms [F5]-[F9], order axioms [O2]-[O3], and the least upper bound axiom [C], to see that this collection of sets S_α is indeed a complete ordered field.
- Therefore, by our uniqueness theorem, these sets S_α along with these operations of addition, multiplication, and ordering provide an explicit construction of $\mathbb{R}$.

- As a conceptual matter, we emphasize that the underlying details of how to construct $\mathbb{R}$ are not really that important for understanding the real numbers themselves: rather, it is the axiomatic description of $\mathbb{R}$ as a complete ordered field that provides the most useful standpoint for working with properties of real numbers.
  - In fact, another common construction of $\mathbb{R}$ uses equivalence classes of Cauchy sequences.
  - But our theorem characterizing $\mathbb{R}$ in fact dictates that *any* property of $\mathbb{R}$ can be proven using *only* the axiomatic description by itself, without referring to any details about the construction of $\mathbb{R}$.

4.2.4 The Complex Numbers

- Definition: A complex number is a number of the form $a+bi$, where a and b are real numbers and i is the so-called "imaginary unit", defined so that $i^2 = -1$. The real part of $z = a+bi$, denoted $\mathrm{Re}(z)$, is the real number a, while the imaginary part of $z = a+bi$, denoted $\mathrm{Im}(z)$, is the real number b. The set of all complex numbers is denoted $\mathbb{C}$.
  - The notation $\sqrt{-1}$ is also often used to denote the imaginary unit i. In certain disciplines (especially electrical engineering), the letter j may instead be used to denote $\sqrt{-1}$, rather than i (which is instead used to denote electrical current).
  - Examples: Some complex numbers are $4+3i$, $3-\pi i$, $6i = 0+6i$, and $-5 = -5+0i$. Their real parts are 4, 3, 0, and -5 respectively, while their imaginary parts are 3, $-\pi$, 6, and 0 respectively.

- Definition: The complex conjugate of $z = a+bi$, denoted $\overline{z}$, is the complex number $a-bi$. The modulus (also called the absolute value, magnitude, or length) of $z = a+bi$, denoted $|z|$, is the real number $\sqrt{a^2+b^2}$.
  - The notation for conjugate varies among disciplines. The notation z^* is often used in physics and computer programming to denote the complex conjugate (in place of $\overline{z}$) since it is easier to type on a standard keyboard.
  - Example: For $z = 3+4i$ we have $\overline{z} = 3-4i$ and $|z| = \sqrt{3^2+4^2} = 5$.

- Two complex numbers are added (or subtracted) simply by adding (or subtracting) their real and imaginary parts: $(a+bi)+(c+di) = (a+c)+(b+d)i$.
  - Example: The sum of $1+2i$ and $3-4i$ is $\boxed{4-2i}$. The difference is $(1+2i)-(3-4i) = \boxed{-2+6i}$.

- Two complex numbers are multiplied using the distributive law and the fact that $i^2 = -1$: $(a+bi)(c+di) = ac+adi+bci+bdi^2 = (ac-bd)+(ad+bc)i$.
  - Example: The product of $1+2i$ and $3-4i$ is $(1+2i)(3-4i) = 3+6i-4i-8i^2 = \boxed{11+2i}$.
  - Observe in particular that for $z = a+bi$, we have $|z|^2 = a^2+b^2 = z \cdot \overline{z}$.

- For division, we rationalize the denominator using the conjugate: $\dfrac{a+bi}{c+di} = \dfrac{(a+bi)(c-di)}{(c+di)(c-di)} = \dfrac{ac+bd}{c^2+d^2} + \dfrac{bc-ad}{c^2+d^2}\,i$.
  - Example: The quotient of $2i$ by $1-i$ is $\dfrac{2i}{1-i} = \dfrac{2i(1+i)}{(1-i)(1+i)} = \dfrac{-2+2i}{2} = \boxed{-1+i}$.

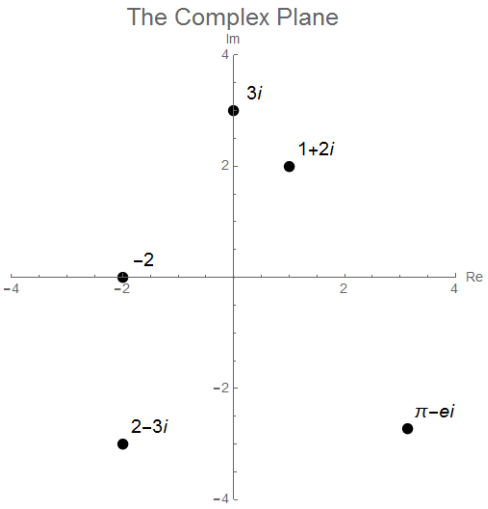
- Here are a few more simple properties of complex number arithmetic:

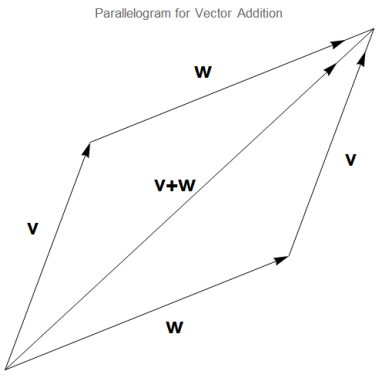
- Proposition (Complex Arithmetic): Suppose z and w are complex numbers.

1. We have $\mathrm{Re}(z) = (z+\overline{z})/2$ and $\mathrm{Im}(z) = (z-\overline{z})/(2i)$.
2. We have $\overline{z+w} = \overline{z} + \overline{w}$, $\overline{zw} = \overline{z} \cdot \overline{w}$, and $\overline{\overline{z}} = z$.
3. We have $|\overline{z}| = |z|$ and $|zw| = |z| \cdot |w|$.
4. We have $z = \overline{z}$ if and only if z is real, while $\overline{z} = -\overline{z}$ if and only if z is purely imaginary (of the form ri where r is real).
5. We have $\mathrm{Re}(z) \leq |z|$ and $\mathrm{Im}(z) \leq |z|$.
6. (Triangle Inequality) We have $|z+w| \leq |z| + |w|$.
   - Proofs: (1)-(5) are easy algebraic calculations.
   - For (6), use (1) and (2) to observe $z\overline{w} + w\overline{z} = 2\mathrm{Re}(z\overline{w})$, and (5) and (3) to observe $2\mathrm{Re}(z\overline{w}) \leq 2\,|z\overline{w}| = 2\,|z|\,|w|$.
   - Then $|z+w|^2 = (z+w)(\overline{z+w}) = z\overline{z}+z\overline{w}+w\overline{z}+w\overline{w} = |z|^2+|w|^2+2\mathrm{Re}(z\overline{w}) \leq |z|^2+|w|^2+2\,|z|\,|w| = (|z|+|w|)^2$. Since both $|z+w|$ and $|z|+|w|$ are nonnegative, taking the square root yields the desired $|z+w| \leq |z| + |w|$.

- We emphasize that (2) above shows that the conjugate is both additive and multiplicative.
  - Example: If $z = 1+2i$ and $w = 3-i$, then $\overline{z} = 1-2i$ and $\overline{w} = 3+i$. We compute $z+w = 4+i$, $\overline{z+w} = 4-i$, $zw = 5+5i$ and $\overline{z} \cdot \overline{w} = 5-5i$, so indeed $\overline{z+w} = \overline{z} + \overline{w}$ and $\overline{zw} = \overline{z} \cdot \overline{w}$.
  - The multiplicativity of the conjugate explains the procedure for performing division: we write $\dfrac{z}{w} = \dfrac{z \cdot \overline{w}}{w \cdot \overline{w}} = \dfrac{z \cdot \overline{w}}{|w|^2}$, where the denominator is now the real number $|w|^2$.

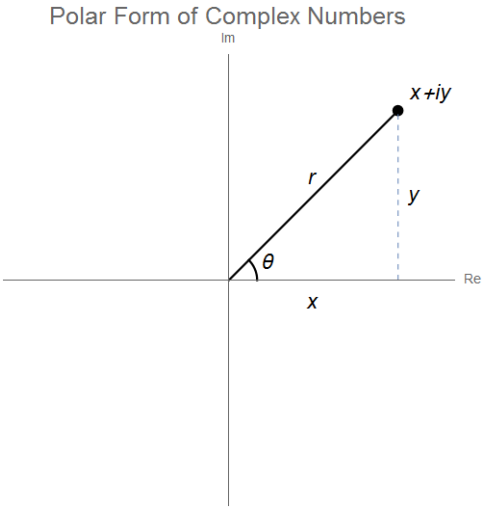
- We often think of the real numbers geometrically, as a line. The natural way to think of the complex numbers is as a plane, with the x-coordinate denoting the real part and the y-coordinate denoting the imaginary part.





  - In this geometric view of the complex numbers, addition corresponds to vector addition in the plane. Explicitly, if we think of the complex number $a+bi$ as a vector, it represents "adding (a,b)" to the coordinates of a point, while $c+di$ represents "adding (c,d)" to the coordinates of a point. The sum $(a+bi)+(c+di)$ is then obtained by adding (a,b) and then (c,d) to the coordinates of a point, which is the same as adding $(a+c, b+d)$ to it.
  - Geometrically, we can think of this "vector addition" using a parallelogram whose pairs of parallel sides are $\mathbf{v} = (a,b)$ and $\mathbf{w} = (c,d)$ and whose diagonal is $\mathbf{v}+\mathbf{w} = (a+c, b+d)$, as shown above.
  - Multiplication of complex numbers also has a very nice geometric interpretation: multiplying a complex number $a+bi$ by a real number r will scale its length by r, while multiplying a complex number $a+bi$ by the imaginary unit i corresponds to a $\pi/2$-radian (90-degree) counterclockwise rotation.

- One may give a more direct geometric description of complex multiplication by rewriting the description of a complex number using basic trigonometry.



  - From the right-triangle diagram above, for $z = x + yi$, if r is the length of the segment joining $x + iy$ to the origin 0, and θ is the angle made by this segment and the positive real axis, then we have $x = r\cos\theta$ and $y = r\sin\theta$, and also $r = \sqrt{x^2 + y^2} = |z|$ and $\tan\theta = y/x$. (The angle θ is called the argument of z and sometimes denoted $\theta = \arg(z)$.)
  - For $z = x + yi$, we thus have $z = r\cos\theta + (r\sin\theta)i = r(\cos\theta + i\sin\theta)$. This expression is called the polar form of the complex number z, contrasting with the rectangular form $z = x + yi$.
  - Example: If $z = 1 + i$, then the corresponding values of r and θ above are $r = |z| = \sqrt{2}$ and $\theta = \tan^{-1}(1) = \dfrac{\pi}{4}$, so we can write z in polar form as $z = \boxed{\sqrt{2}\left[\cos\dfrac{\pi}{4} + i\sin\dfrac{\pi}{4}\right]}$. Indeed, we may check that $\sqrt{2}\left[\cos\dfrac{\pi}{4} + i\sin\dfrac{\pi}{4}\right] = \sqrt{2}\left[\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}\right] = 1 + i$, as it should be.
  - Using the polar form we can give another interpretation of multiplication: if we have complex numbers $z = r[\cos\theta + i\sin\theta]$ and $w = s[\cos\varphi + i\sin\varphi]$, then the product $zw = rs(\cos\theta + i\sin\theta)(\cos\varphi + i\sin\varphi) = rs[(\cos\theta\cos\varphi - \sin\theta\sin\varphi) + (\cos\theta\sin\varphi + \sin\theta\cos\varphi)i] = rs[\cos(\theta + \varphi) + i\sin(\theta + \varphi)]$.
  - So we see that the length of zw is the product of the lengths of z and w, while the argument of zw is the sum of the arguments of z and w: this is often summarized as "lengths multiply, angles add" when multiplying complex numbers.

4.2.5 Solving Polynomial Equations

- Finding formulas for the roots of a polynomial has been a very classical problem in algebra.
  - The methods for finding solutions to quadratic equations $az^2 + bz + c = 0$ have been well understood for thousands of years, as various procedures were described (in essentially as complete detail as the methods and notation of the time would allow) by the ancient Egyptians and Greeks.
  - Explicitly, if a, b, c are real numbers and $a \neq 0$, then we may complete the square in the expression $az^2 + bz + c = 0$ and write it as $a(z + \frac{b}{2a})^2 + \frac{4ac - b^2}{4a} = 0$.
  - Moving the constant term to the right-hand side and dividing by a yields $(z + \frac{b}{2a})^2 = \frac{b^2 - 4ac}{4a^2}$ and now taking the square root and solving for z yields the familiar quadratic formula $z = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.
  - Many textbooks introduce complex numbers as a tool for giving meaning to the formal symbols obtained when using the quadratic formula to "solve" quadratic equations that do not have real solutions, because in the situation where $b^2 - 4ac < 0$, the solutions are not real numbers but rather complex numbers.
  - For example, the formula states that the solutions to $z^2 + 2z + 2 = 0$ are $z = \dfrac{-2 \pm \sqrt{-4}}{2} = -1 \pm i$, and indeed we can check directly that the expression $z^2 + 2z + 2$ evaluates to 0 when $z = -1 + i$ or $-1 - i$.

- The quadratic formula yields the factorization $az^2+bz+c = a(z-r_1)(z-r_2)$ where $r_1 = \dfrac{-b-\sqrt{b^2-4ac}}{2a}$ and $r_2 = \dfrac{-b+\sqrt{b^2-4ac}}{2a}$ are the two roots.

- One reason we are interested in the complex numbers is that all polynomial equations have their solutions inside $\mathbb{C}$:

- Theorem (Fundamental Theorem of Algebra): Suppose $p(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$ is a polynomial with complex coefficients $a_n, a_{n-1}, \ldots, a_0$ of any degree n. Then $p(z)$ can be factored completely over the complex numbers: namely, it can be written as $p(z) = a(z-r_1)(z-r_2)\cdots(z-r_n)$ for some (not necessarily distinct) complex numbers r_i.

  - The statement of this theorem was given as far back as 1629 by Girard, but it was not generally believed to be true for all polynomials until the late 1700s, when attempts to prove it were given by Euler, Lagrange, Laplace, and others. Gauss published an essentially correct proof in 1799 that had a small gap, and the first fully correct proof was given by Argand in 1806. (Gauss eventually gave several other fully rigorous proofs in the 1810s.) Since then, many other different proofs of the fundamental theorem of algebra have been given, some of which require little more than basic facts about polynomials and continuous functions, and the triangle inequality.
  - Here is an outline of one such proof: induct on the degree n. The base case $n = 1$ is trivial.
  - For the inductive step, suppose that $p(z)$ has degree $n+1$. We first show that $|p(z)|$ must achieve its global minimum value at some complex number (this uses some basic facts about continuous functions and the fact that $|p(z)|$ grows to ∞ as $|z|$ grows to ∞), and then we also show that if $|p(\alpha)| > 0$, then there is some value β near α with $|p(\beta)| < |p(\alpha)|$ (this is an estimate that can eventually be obtained using the triangle inequality).
  - Together these two facts imply that the minimum value of $|p(z)|$ must be zero, so $p(z)$ has some factor $z - r$. Dividing $p(z)$ by this factor yields a polynomial of degree n, which by the inductive hypothesis must have a factorization: multiplying it by $z - r$ then yields a factorization of $p(z)$.
  - Interestingly, however, all known proofs of this theorem for general polynomials are non-constructive, in that they do not give explicit formulas for the roots r_1, ... , r_n of the polynomials in terms of the coefficients. (There are proofs that construct the roots as limits of infinite sequences, but these procedures do not give explicit formulas for the roots.)

- The problem of finding a general formula for the roots of a cubic equation, analogous to the quadratic formula, was considered by the ancient Egyptians and Greeks, who were unable to find such a formula. Ultimately, the story of how the cubic formula was eventually discovered and publicized is rather convoluted, and we will briefly summarize it.

  - Minimal progress was made on solving the cubic until the early 1500s, when del Ferro discovered a method for solving cubics of the form $t^3 + pt = q$. However, due to the nature of Renaissance patronage, he did not publicize his method, but only taught it to his student Fior.
  - In 1535, Fior in turn challenged another scholar, Niccolo Fontana (nicknamed Tartaglia due to a physical deformity), who eventually (re)discovered the solution to the cubic, and (again, as was normal at the time) kept it a secret.
  - Eventually, Gerolamo Cardano (an avid astrologer and gambler who at one time was one of the most well-regarded physicians in Europe, who was eventually jailed for heresy and then pardoned by the Pope) was able, after repeated entreaties and vows never to reveal Tartaglia's method, to coax Tartaglia into revealing it.
  - Cardano was then able to extend Tartaglia's method to solve the general cubic equation, and eventually took a student, Ludovico Ferrari, who was able to extend Cardano's techniques to solve degree-4 equations. Cardano and Ferrari eventually discovered that del Ferro had solved the cubic prior to Tartaglia's discovery of the solution, and published his generalization in 1545, giving credit to del Ferro, Fior, and Tartaglia. (Despite receiving proper attribution, Tartaglia nonetheless felt betrayed by Cardano, despite the fact that del Ferro had developed the technique prior to Tartaglia.)

- We will present a solution of the cubic similar to Cardano's (and presumably, also to Tartaglia's).
- Theorem (Cardano's Formulas): If p and q are complex numbers, then one root of the polynomial of $g(t) = t^3 + pt + q$ is the sum $A + B$ where $A = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$ and $B = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$ where the cube roots are chosen so that $AB = -p/3$.
  - Proof: From the algebraic identity $(A + B)^3 - 3AB(A + B) = A^3 + B^3$, we can see that if we take $A + B = t$, $3AB = -p$, and $A^3 + B^3 = -q$, then the identity becomes $t^3 + pt + q = 0$.
  - The equation $3AB = -p$ implies $B = -p/(3A)$, and then $A^3 + B^3 = -q$ becomes $A^3 - p^3/(27A^3) = -q$, whence $A^6 + qA^3 - p^3/27 = 0$.
  - Solving this quadratic in A^3 yields $A^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$ and so $A = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$.
  - One may then check that for $B = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$ we do have $3AB = -p$ and $A^3 + B^3 = -q$, so we obtain the root $t = A + B$ as claimed.
- Example: Solve the cubic equation $t^3 - 15t - 4 = 0$ using Cardano's formulas.
  - Plugging in $p = -15$ and $q = -4$ yields $A = \sqrt[3]{2 + \sqrt{-121}}$ and $B = \sqrt[3]{-2 + \sqrt{-121}}$, yielding the root $r = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{-2 + \sqrt{-121}}$.
  - Notice that the cube roots involve complex numbers. Interestingly, however, it is not hard to check that the three roots of this cubic are the real numbers 4 and $-2 \pm \sqrt{3}$, so if Cardano's formula is correct, there must be a way to simplify it to obtain a real number.
  - In fact, it is precisely this perplexing appearance of square roots of negative numbers in the formulas for real solutions to a cubic equation that led to the initial development of complex numbers in mathematics in the first place!
  - Specifically, in 1572 Bombelli observed that one may formally compute $(2 \pm \sqrt{-1})^3 = \pm 2 + \sqrt{-121}$, and so one may take $A = 2 + \sqrt{-1}$ and $B = 2 - \sqrt{-1}$ to obtain the correct roots $A + B = 4$.
  - In fact, it turns out to be impossible to give general formulas involving only real radicals for the solutions of cubic polynomials with 3 real roots, and so (rather by necessity) resolving this difficulty could only be achieved by working with non-real numbers.
- With the cubic formula, the next question becomes: what about polynomials of larger degree?
  - Indeed, as mentioned earlier, Cardano's student Ferrari found a formula for the solutions of degree-4 polynomials in 1540 and published the solution along with Cardano's solution of the cubic in 1545.
  - But they, and everyone who came after them, were unable to extend their results to give a general solution in radicals for degree-5 polynomials.
  - In fact, there is a very good reason for this: there is no such formula! This nonexistence result is known as the Abel-Ruffini theorem: an incomplete proof was given by Ruffini in 1799 that was refined and completed in 1813, and a more general result was proven by Abel in 1824.
  - As a specific example, the roots of $p(z) = z^5 + z + 1$ cannot be expressed in terms of radicals.
  - In fact, the key idea in the Abel-Ruffini theorem involves relating the behavior of the roots of a degree-5 polynomial to the properties of subgroups of the symmetric group S_5!
  - Precisely, the idea is that the group S_5 acts on the roots of a degree-5 polynomial by permuting them, and by studying this action appropriately, one can relate the existence of a solution in radicals to a certain property of the group of permutations that S_5 lacks, which quite appropriately is called "solvability"! (This approach of using group theory to study the roots of polynomials is known as Galois theory and was pioneered by Évariste Galois, who was, sadly, killed in a duel at the age of 20.)

Well, you're at the end of my handout. Hope it was helpful.