

Math 331 Fall 2026, Lecture 6 and 7: Abelian groups, subgroups

website: <https://www.cip.ifi.lmu.de/~grinberg/t/26fa/>
(GPT-6 was used to polish the writing that follows.)

1.8. Orders of elements

1.8.1. Definition and example

The notion of a power gives rise to the notion of an order:

Definition 1.8.1. Let M be a monoid (written multiplicatively). Let $a \in M$.

If there exists a positive integer n such that $a^n = 1$, then the **smallest** such n is called the **order** of a , and is denoted by $o(a)$; and then we say that a has **order** n , and that a has **finite order**.

If no such n exists, then we say that a has **infinite order**, and write $o(a) = \infty$.

Note that if the monoid M is finite and the element $a \in M$ has an inverse, then a always has finite order by Theorem 1.7.15 (a) from Lecture 5. Infinite orders can happen for elements a that have no inverses, as well as for elements of infinite monoids (even for elements of infinite groups).

Example 1.8.2. Recall the symmetric group S_3 . It has six elements $\text{id}, s_1, s_2, w_0, c, c^{-1}$, where s_1 and s_2 are as in Example 1.7.5 in Lecture 5 (so s_1 swaps 1 with 2, whereas s_2 swaps 2 with 3), while w_0 swaps 1 with 3, and finally c is the 3-cycle $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ (so that c^{-1} is the 3-cycle $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$). The orders of these six elements are

$$\begin{aligned} o(\text{id}) &= 1, & o(s_1) &= 2, & o(s_2) &= 2, \\ o(w_0) &= 2, & o(c) &= 3, & o(c^{-1}) &= 3. \end{aligned}$$

For example, $o(c) = 3$ is because $c^3 = \text{id}$ (the neutral element of S_3) while c^1 and c^2 are not id .

Example 1.8.3. Let M be the monoid $(\{0, 1, -1\}, \cdot)$. Its elements have orders

$$o(0) = \infty, \quad o(1) = 1, \quad o(-1) = 2.$$

We observe that $o(g^{-1}) = o(g)$ for any $g \in S_3$. This generalizes to invertible elements of monoids; see homework set #2.

Remark 1.8.4. In group theory, the word “order” has a second meaning: The **order** of a group G means the size (aka cardinality) $|G|$ of the group, that is, the number of all elements of G . This name is a historical fossil, and is nowadays more confusing than it is useful (in particular, as Example 1.8.2 shows, the order of a group doesn’t have to equal the order of any of its elements!); thus, we will avoid it, instead simply saying “size” or “cardinality”.

There is actually a connection between the two orders: For any finite group G and any $a \in G$, we have $o(a) \mid |G|$ (that is, the order of a divides the order of G); equivalently, $a^{|G|} = 1$ (if we write G multiplicatively). This will fall out of something we will do in the next weeks: the notion of a coset and Lagrange’s theorem for subgroups. Note that this is not true for monoids, as Example 1.8.3 shows.

1.8.2. The structure of the powers of a finite-order element

The order of a finite-order element $a \in M$ determines what its powers look like:

Theorem 1.8.5. Let M be a monoid (written multiplicatively). Let $a \in M$ be an element that has finite order. Let $n = o(a)$. Then:

- (a) The powers $a^0, a^1, \dots, a^{n-1}$ are distinct.
- (b) The element a has an inverse, namely a^{n-1} .
- (c) Each power of a (including the negative powers) is one of $a^0, a^1, \dots, a^{n-1}$.
- (d) If M is finite, then $n \leq |M|$.

Proof. We have $n = o(a)$, so that n is a positive integer such that $a^n = 1$. We shall prove part (b) first, so that cancellation will be justified in part (a).

- (b) By the rules of exponents (specifically, Theorem 1.7.3 (b)), we have $aa^{n-1} = a^n = 1$ (since $n = o(a)$) and $a^{n-1}a = a^n = 1$. This shows that a^{n-1} is an inverse of a . Thus, part (b) is proved.
- (a) Assume the contrary. Thus, two of these powers are equal. That is, $a^p = a^q$ for some $n > p > q \geq 0$. Consider these p and q . Then, the rules of exponents yield

$$a^{p-q}a^q = a^{(p-q)+q} = a^p = a^q = 1a^q. \quad (1)$$

But part (b) shows that a has an inverse. Hence Lemma 1.7.10 from Lecture 5 shows that a^q has an inverse as well. Thus, we can cancel the a^q factor from (1) (using Theorem 1.6.1 (d)), obtaining $a^{p-q} = 1$. Note also that $p - q$ is a positive integer (since $p > q$) and satisfies $p - q \leq p < n$ (since $q \geq 0$ and $n > p$).

But $n = o(a)$ is defined as the **smallest** positive integer k such that $a^k = 1$. So there cannot be any smaller such k . That is, if k is a positive integer such

that $a^k = 1$, then $k \geq n$. Applying this to $k = p - q$, we obtain $p - q \geq n$ (since $a^{p-q} = 1$), which contradicts $p - q < n$. This contradiction shows that our assumption was wrong, and so part **(a)** is proved.

(c) This is visually clear if you draw the powers of a as a cycle.

For a proper proof, define the set $P = \{a^0, a^1, \dots, a^{n-1}\}$. Then, we must show that $a^m \in P$ for each $m \in \mathbb{Z}$. We do this by two-sided induction on m (see the proof of Theorem 1.7.9 in Lecture 5 for a detailed explanation of this strategy). The *base case* $m = 0$ is obvious (we have¹ $a^0 \in \{a^0, a^1, \dots, a^{n-1}\} = P$). For the *induction step*, we fix an integer m ; then we must prove the equivalence

$$(a^m \in P) \iff (a^{m+1} \in P).$$

We prove its “ $\implies$ ” and “ $\impliedby$ ” directions separately:

$\implies$: If $a^m \in P$, then

$$\begin{aligned} a^{m+1} &= a^m a \\ &\in \{a^0 a, a^1 a, \dots, a^{n-2} a, a^{n-1} a\} \\ &\quad \left(\text{since } a^m \in P = \{a^0, a^1, \dots, a^{n-2}, a^{n-1}\} \right) \\ &= \{a^1, a^2, \dots, a^{n-1}, a^n\} \quad \left(\text{since } a^k a = a^{k+1} \text{ for each } k \in \mathbb{Z} \right) \\ &= \{a^1, a^2, \dots, a^{n-1}, a^0\} \quad \left(\text{since } a^n = 1 = a^0 \right) \\ &= \{a^0, a^1, \dots, a^{n-1}\} = P. \end{aligned}$$

This proves the “ $\implies$ ” direction.

$\impliedby$: Conversely, if $a^{m+1} \in P$, then

$$\begin{aligned} a^m &= a^{m+1} a^{-1} \quad \left(\text{since } a^{m+1} a^{-1} = a^{(m+1)+(-1)} = a^m \right) \\ &\in \{a^0 a^{-1}, a^1 a^{-1}, \dots, a^{n-2} a^{-1}, a^{n-1} a^{-1}\} \\ &\quad \left(\text{since } a^{m+1} \in P = \{a^0, a^1, \dots, a^{n-2}, a^{n-1}\} \right) \\ &= \{a^{-1}, a^0, \dots, a^{n-3}, a^{n-2}\} \quad \left(\text{since } a^k a^{-1} = a^{k-1} \text{ for each } k \in \mathbb{Z} \right) \\ &= \{a^{n-1}, a^0, \dots, a^{n-3}, a^{n-2}\} \quad \left(\text{since } a^{n-1} = a^{-1} \text{ by part (b)} \right) \\ &= \{a^0, a^1, \dots, a^{n-2}, a^{n-1}\} = P. \end{aligned}$$

This proves the “ $\impliedby$ ” direction.

¹We are using the fact that $n = o(a)$ is a positive integer (by its definition), so that $0 \in \{0, 1, \dots, n-1\}$.

So both directions are proved, and the induction step is complete. This proves that $a^m \in P$ for each $m \in \mathbb{Z}$. Thus, part (c) is proved.

- (d) By part (a), the elements $a^0, a^1, \dots, a^{n-1}$ are n distinct elements of the set M . Thus, M contains n distinct elements. Therefore, $n \leq |M|$ if M is finite. $\square$

Remark 1.8.6. Let G be a group. Only one element of G has order 1, namely the identity element 1.

A non-identity element $a \neq 1$ of G has order 2 if and only if $a^{-1} = a$ (because $a^{-1} = a$ is equivalent to $aa = 1$, that is, $a^2 = 1$). Note that a group G can have infinitely many elements of order 2.

Example 1.8.7. Let A be a set. In the group $(\mathcal{P}(A), \Delta)$, each nonempty subset S of A has order 2. This is because $S \Delta S = \emptyset$. (The empty subset has order 1.)

Example 1.8.8. In the group $(\mathbb{Z}, +)$, the element 0 has order 1. All other elements $a \in \mathbb{Z}$ have order ∞ . In fact, the n -th power “ a^n ” of an element $a \in \mathbb{Z}$ in the group $(\mathbb{Z}, +)$ is just na , which is nonzero when n is positive and a is nonzero. (This is a good example of when not to use multiplicative notation.)

Example 1.8.9. In the monoid $(\mathbb{Z}, \cdot)$, the elements have the following orders:

$$\begin{aligned} o(1) &= 1, & o(-1) &= 2, \\ o(a) &= \infty & \text{for all } a \notin \{1, -1\}. \end{aligned}$$

The same applies to $(\mathbb{Q}, \cdot)$ and $(\mathbb{R}, \cdot)$.

But the monoid $(\mathbb{C}, \cdot)$ has many more elements of finite order. They are called **roots of unity**, and have the form

$$e^{2\pi i q} = \cos(2\pi q) + i \sin(2\pi q) \quad \text{for rational } q$$

(that is, geometrically speaking, they lie on the unit circle, and their arguments are $2\pi q$ for rational q ; in other words, they are vertices of regular n -gons inscribed into the unit circle with one vertex at 1). Still, “most” complex numbers have infinite order in $(\mathbb{C}, \cdot)$.

1.8.3. Finite groups of size 4

As an application of orders, let us continue on our “what are the finite groups of size k ” quest. Recall that

1. there is only one group of size 1, namely the trivial group $(\{1\}, \cdot)$ (because it must have a neutral element, and then, having size 1, it cannot have any further elements).

2. there is only one group of size 2, namely $(\{1, f\}, \cdot)$ with $ff = 1$ (as we saw in §1.2 in Lecture 2).
3. there is only one group of size 3, namely $(\{1, f, f^2\}, \cdot)$ with $f^3 = 1$ (as we saw in Example 1.6.5 in Lecture 4; note that we denoted f^2 by g , but it is still f^2).

We are now ready for the next step: How many groups are there of size 4 ?

Let G be a group of size 4, written multiplicatively. Its neutral element 1 has order $o(1) = 1$; its other three elements have orders between 2 and 4 inclusively (since Theorem 1.8.5 (d) tells us that each order is ≤ 4 , but only the neutral element has order 1). We distinguish between three cases:

Case 1: Some element $a \in G$ has order 4. Then, $a^4 = 1$. Moreover, Theorem 1.8.5 (a) tells us that a^0, a^1, a^2, a^3 are distinct. Since G has only 4 elements, this means that $G = \{a^0, a^1, a^2, a^3\}$. Then the multiplication table of G is given by $a^i a^j = a^{i+j}$, where $a^4 = 1 = a^0$ and $a^5 = a \underbrace{a^4}_{=1} = a$ and

$a^6 = a \underbrace{a^5}_{=a} = aa = a^2$. That is, this table is

	a^0	a^1	a^2	a^3
a^0	a^0	a^1	a^2	a^3
a^1	a^1	a^2	a^3	a^0
a^2	a^2	a^3	a^0	a^1
a^3	a^3	a^0	a^1	a^2

(where the format of the table is the same as before: we put xy into the cell in the x -th row and the y -th column). This table has a very simple form: Its first row is (a^0, a^1, a^2, a^3) , and each further row is obtained by cyclically rotating the previous row one step to the left. We will soon see that this really does define a group, called the **cyclic group of order 4**. (If you are impatient, you can check this by hand... though this, too, requires some patience.)

Case 2: No element of G has order 4, but some element a has order 3. Arguing as in Case 1, we see that a^0, a^1, a^2 are three distinct elements of G , so there must be one more element $b \in G$ that is not a power of a (since $|G| = 4$). Consider this b . What is ab ?

- a) If $ab = 1$, then $b = a^{-1} = a^2$ (since $a^3 = 1$), which contradicts b not being any of a^0, a^1, a^2 .
 - b) If $ab = a$, then (cancelling a) we obtain $b = 1 = a^0$, which leads to a similar contradiction.
-

- c) If $ab = a^2$, then (cancelling a) we find $b = a = a^1$, which leads to a similar contradiction as well.
- d) If $ab = b$, then $a = 1$ (by cancelling b), which is also false (since a has order 3, not 1).

So we get a contradiction, which means that Case 2 is impossible.

Case 3: All three non-identity elements of G have order 2. Call them a, b, c . Then we can easily reconstruct the multiplication table of G : For instance, $aa = a^2 = 1$ since a has order 2. Likewise, $bb = 1$ and $cc = 1$. Furthermore, ab cannot be a (since $b \neq 1$), nor can it be b (since $a \neq 1$), nor can it be 1 (since $ab = 1 = aa$ would yield $b = a$). Hence, ab must be c . Similarly, $ba = c$ and $bc = a$ and $cb = a$ and $ca = b$ and $ac = b$. Thus, the multiplication table of G looks as follows:

	1	a	b	c
1	1	a	b	c
a	a	1	c	b
b	b	c	1	a
c	c	b	a	1

Is there such a group? Yes: it is precisely the group $(\mathcal{P}(A), \Delta)$ for $A = \{1, 2\}$, with its four elements $\emptyset, \{1\}, \{2\}, \{1, 2\}$ renamed as $1, a, b, c$. This is known as the **Klein four-group**.

So we conclude that there are two essentially different (as we will later say: non-isomorphic) groups of size 4, namely the cyclic group and the Klein four-group.

What about size 5? We will answer this later, once we have proved Lagrange's theorem.

1.9. The opposite group

There are several ways to build new groups from old ones. The following is perhaps the simplest:

Proposition 1.9.1. Let $(S, *)$ be a semigroup. Then, we can define a new binary operation $\#$ on S by setting

$$s \# t := t * s \quad \text{for all } s, t \in S.$$

So $\#$ is the same operation as $*$ but with its two arguments swapped. Then, $(S, \#)$ is again a semigroup.

The same holds if we replace "semigroup" by "monoid" or "group".

Proof. Let us first check associativity.

For any $a, b, c \in S$, we have

$$a \# (b \# c) = (b \# c) * a = (c * b) * a$$

and similarly

$$(a \# b) \# c = c * (b * a).$$

Since $*$ is associative, the right-hand sides of these two equalities are equal; therefore so are the left-hand sides. That is, $a \# (b \# c) = (a \# b) \# c$ for all $a, b, c \in S$. In other words, $\#$ is associative. This shows that $(S, \#)$ is a semigroup.

If $(S, *)$ is a monoid with neutral element e , then each $s \in S$ satisfies

$$e \# s = s * e = s \quad \text{and} \quad s \# e = e * s = s.$$

Thus, e is also neutral for $\#$, and therefore $(S, \#)$ is a monoid.

If $(S, *)$ is a group and t is the inverse of s with respect to $*$, then

$$s \# t = t * s = e \quad \text{and} \quad t \# s = s * t = e.$$

Thus, the inverses are the same for the two operations $*$ and $\#$, so $(S, \#)$ is a group. $\square$

Definition 1.9.2. If $(S, *)$ is a semigroup, and $\#$ is the operation defined in Proposition 1.9.1, then the semigroup $(S, \#)$ is called the **opposite semigroup** of $(S, *)$, and is denoted by $(S, *)^{\text{op}}$.

Likewise, if $(S, *)$ is a monoid or a group, then the monoid $(S, \#)$ is called the **opposite monoid** or **opposite group** (respectively) of $(S, *)$.

In general, we cannot use multiplicative notation for both $(S, *)$ and $(S, \#)$ simultaneously, since this would cause two different operations to both be denoted by the same $\cdot$ symbol.

1.10. Abelian groups

As we said, commutativity is not a given in semigroups, but it happens often and deserves further study. We recall its definition:

Definition 1.10.1. Let S be a semigroup (written multiplicatively). Two elements a and b of S are said to **commute** if $ab = ba$.

Example 1.10.2. (a) For a non-example, consider the symmetric group S_3 again. The permutations s_1 and s_2 in S_3 (see Example 1.8.2) do not commute (since $s_1 s_2$ is the cycle $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, whereas $s_2 s_1$ is the cycle $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$).

(b) For a positive example, consider any semigroup S . Then, any two powers of a single element $a \in S$ commute. That is, if $a \in S$ and $n, m > 0$, then a^n and a^m commute (since $a^n a^m = a^{n+m} = a^m a^n$). The same holds for $n, m \in \mathbb{N}$ when S is a monoid and for $n, m \in \mathbb{Z}$ when S is a group (or a has an inverse). (Recall that $\mathbb{N} = \{0, 1, 2, \dots\}$.)

Definition 1.10.3. A semigroup (or monoid or group) S (written multiplicatively) is said to be **abelian** (or **commutative**) if all $a, b \in S$ commute, i.e., if we have $ab = ba$ for all $a, b \in S$.

Thus, a semigroup is abelian if and only if its multiplication table is symmetric across its main diagonal.

Here are some examples and non-examples:

- The semigroup $(\mathbb{Z}_{>0}, +)$, the monoid $(\mathbb{N}, +)$, and the groups $(\mathbb{Z}, +)$, $(\mathbb{C}, +)$ and $(\mathbb{R}^{n \times n}, +)$ are abelian.
- Any of the monoids $(\mathbb{Z}_{>0}, \cdot)$, $(\mathbb{N}, \cdot)$, $(\mathbb{Z}, \cdot)$ and $(\mathbb{C}, \cdot)$ is abelian, but the monoids $(\mathbb{R}^{n \times n}, \cdot)$ and $(\text{GL}_n(\mathbb{R}), \cdot)$ for $n > 1$ are not (since matrices do not commute, even invertible matrices).
- For any set A , the group $(\mathcal{P}(A), \Delta)$ and the monoids $(\mathcal{P}(A), \cup)$ and $(\mathcal{P}(A), \cap)$ are abelian.
- All groups of size ≤ 4 are abelian (as we have seen when we described them). This still holds for groups of size ≤ 5 (as we will soon see). But not for groups of size 6: the symmetric group S_3 is not abelian (s_1 and s_2 do not commute). Non-abelian finite groups exist in sizes 6 and 8 but not 7 and 9; we will later learn why.

Generally, the symmetric group S_n (that is, the group of all the $n!$ many permutations of $\{1, 2, \dots, n\}$) is not abelian whenever $n \geq 3$. But the symmetric group S_2 has size 2, so it is abelian.

- In contrast, there are non-abelian semigroups of any size ≥ 2 . Indeed, on any set S , we can define a binary operation $*$ by

$$a * b := b \quad \text{for all } a, b \in S$$

(that is, the operation that forgets its first argument and returns its second); this operation is associative, thus defines a semigroup, and this semigroup is non-abelian as soon as S has two distinct elements.

We have been often using multiplicative notation for arbitrary semigroups. For abelian semigroups, it has a rival: the (analogous) additive notation, in which the operation is written as $+$. In more detail:

Definition 1.10.4. Abelian semigroups (thus also abelian monoids and abelian groups) are often **written additively** (aka in **additive notation**); this means that the binary operation is denoted by $+$. Correspondingly, the neutral element (when it exists) is denoted by 0 (not by 1). The inverse of an element a (when it exists) is then denoted by $-a$ (not by a^{-1}). The n -th power of an element a is then denoted by na (not by a^n).

All this notation is merely generalizing the example of $(\mathbb{Z}, +)$, in which the meanings of $a + b$, 0 , $-a$ and na we just defined are precisely the usual meanings of these expressions known from middle school. Of course, you should not use additive notation when the symbol $+$ already has a different meaning on your set.

Also, it is **not recommended** to use additive notation for non-abelian semigroups (or monoids, or groups). Most mathematicians instinctively reorder sums without thinking, relying instinctively on the commutativity of addition; thus, writing a non-commutative operation as addition risks this kind of re-ordering being used by accident when it is unjustified.

1.11. Subsemigroups, submonoids, subgroups

We shall now define **subobjects** of the objects we have been studying: subsemigroups of semigroups, submonoids of monoids, subgroups of groups. The definitions are fairly unsurprising if you have seen your share of mathematical terminology (compare subsets, vector subspaces and topological subspaces):

Definition 1.11.1. Let $(S, *)$ be a semigroup. A subset T of S is said to be a **subsemigroup** of $(S, *)$ if it is closed under the operation $*$, that is, if all $a, b \in T$ satisfy $a * b \in T$.

If this is the case, then we write $T \leq_{\text{semi}} (S, *)$, or just $T \leq_{\text{semi}} S$ when the operation $*$ is clear from the context.

Definition 1.11.2. Let $(S, *)$ be a monoid. A subset T of S is said to be a **submonoid** of $(S, *)$ if it contains the neutral element of $(S, *)$ and is closed under the operation $*$.

If this is the case, then we write $T \leq_{\text{monoid}} (S, *)$, or just $T \leq_{\text{monoid}} S$ when the operation $*$ is clear from the context.

Thus, a submonoid of a monoid $(S, *)$ is the same as a subsemigroup of $(S, *)$ that contains the neutral element of $(S, *)$.

Definition 1.11.3. Let $(G, *)$ be a group. A subset H of G is said to be a **subgroup** of $(G, *)$ if it satisfies the following three axioms:

1. it is closed under $*$, that is, we have $a * b \in H$ for all $a, b \in H$;
2. it contains the neutral element of $(G, *)$;
3. it is closed under taking inverses, that is, we have $a^{-1} \in H$ for all $a \in H$.

If this is the case, then we write $H \leq (G, *)$, or just $H \leq G$.

So a subgroup of a group $(G, *)$ is the same as a submonoid of $(G, *)$ that is closed under taking inverses (i.e., contains the inverse of each of its elements).

All three definitions we just gave follow the same pattern: A sub- X of a given X (where X is one of “semigroup”, “monoid” and “group”) is a subset of this X that contains (parts of) all the defining features of this X . Here, we regard the neutral element as a defining feature of monoids, and we regard the neutral element and inverses as defining features of groups.

Here are some examples:

- We have $\mathbb{N} \leq_{\text{monoid}} (\mathbb{Z}, +)$, because $\mathbb{N}$ is a subset of $\mathbb{Z}$ that contains the neutral element 0 of $(\mathbb{Z}, +)$ and is closed under $+$ (any sum of two nonnegative integers is a nonnegative integer).
- We have $\mathbb{Z}_{>0} \leq_{\text{semi}} (\mathbb{Z}, +)$ (but not $\leq_{\text{monoid}}$ because $\mathbb{Z}_{>0}$ does not contain the neutral element 0 of $(\mathbb{Z}, +)$).
- We have $\mathbb{Z} \leq \mathbb{Q}$ and $\mathbb{Q} \leq \mathbb{R}$ and $\mathbb{R} \leq \mathbb{C}$, where all the groups are understood to be additive (i.e., read $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ as $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ and $(\mathbb{R}, +)$).
- Let $\text{GL}_n^{\text{diag}}(\mathbb{R})$ be the set of all diagonal $n \times n$ -matrices with real entries such that all their diagonal entries are nonzero. Note that the product of two diagonal $n \times n$ -matrices is again diagonal and is given by the formula

$$\text{diag}(a_1, a_2, \dots, a_n) \cdot \text{diag}(b_1, b_2, \dots, b_n) = \text{diag}(a_1 b_1, a_2 b_2, \dots, a_n b_n);$$

thus, all such matrices are invertible, with inverses given by

$$(\text{diag}(a_1, a_2, \dots, a_n))^{-1} = \text{diag}(a_1^{-1}, a_2^{-1}, \dots, a_n^{-1}).$$

Thus, the set $\text{GL}_n^{\text{diag}}(\mathbb{R})$ is a subset of $\text{GL}_n(\mathbb{R})$ that is closed under multiplication, contains the identity matrix, and is closed under taking inverses. Hence, $\text{GL}_n^{\text{diag}}(\mathbb{R}) \leq \text{GL}_n(\mathbb{R})$ (where $\text{GL}_n(\mathbb{R})$ means $(\text{GL}_n(\mathbb{R}), \cdot)$ as usual).

- Let S be a semigroup (written multiplicatively), and let $a \in S$ be arbitrary. Then,

$$\{a^1, a^2, a^3, \dots\} \leq_{\text{semi}} S, \quad (2)$$

because the set $\{a^1, a^2, a^3, \dots\}$ is closed under multiplication (since $a^n a^m = a^{n+m}$ for all $n, m > 0$).

- Let S be a monoid (written multiplicatively), and let $a \in S$ be arbitrary. Then,

$$\{a^0, a^1, a^2, a^3, \dots\} \leq_{\text{monoid}} S, \quad (3)$$

because the set $\{a^0, a^1, a^2, a^3, \dots\}$ is closed under multiplication (since $a^n a^m = a^{n+m}$ for all $n, m \in \mathbb{N}$) and because $1 = a^0$.

- Let G be a group (written multiplicatively), and let $a \in G$ be arbitrary. Then,

$$\{a^n \mid n \in \mathbb{Z}\} \leq G, \quad (4)$$

because the set $\{a^n \mid n \in \mathbb{Z}\} = \{\dots, a^{-2}, a^{-1}, a^0, a^1, a^2, \dots\}$ is closed under multiplication (since $a^n a^m = a^{n+m}$ for all $n, m \in \mathbb{Z}$) and because $1 = a^0$ and $(a^n)^{-1} = a^{-n}$ (by Lemma 1.7.10 in Lecture 5).

In contrast, $\{a^0, a^1, a^2, a^3, \dots\}$ is only a submonoid of G in general, and $\{a^1, a^2, a^3, \dots\}$ is only a subsemigroup.

- Let A be a set. Let $\mathcal{P}_{\text{even}}(A)$ denote the set of all finite subsets of A whose size is even. This is a subgroup of $(\mathcal{P}(A), \Delta)$. Why? It is closed under Δ , because for every $S, T \in \mathcal{P}_{\text{even}}(A)$, the size

$$\begin{aligned} |S \Delta T| &= |(S \cup T) \setminus (S \cap T)| && (\text{since } S \Delta T = (S \cup T) \setminus (S \cap T)) \\ &= \underbrace{|S \cup T|}_{=|S|+|T|-|S \cap T|} - |S \cap T| && (\text{since } S \cap T \subseteq S \cup T) \\ &\quad \text{(by inclusion-exclusion:} \\ &\quad \text{e.g., look at the Venn diagram)} \\ &= |S| + |T| - |S \cap T| - |S \cap T| \\ &= \underbrace{|S|}_{\substack{\text{even} \\ (\text{since } S \in \mathcal{P}_{\text{even}}(A))}} + \underbrace{|T|}_{\substack{\text{even} \\ (\text{since } T \in \mathcal{P}_{\text{even}}(A))}} - \underbrace{2 \cdot |S \cap T|}_{\text{even}} && \text{is even,} \end{aligned}$$

so we have $S \Delta T \in \mathcal{P}_{\text{even}}(A)$. It contains the neutral element of $(\mathcal{P}(A), \Delta)$, since this neutral element is $\emptyset$ and has size $|\emptyset| = 0$. Finally, it is closed under taking inverses, since the inverse of any element $S \in \mathcal{P}(A)$ is S itself.

- The monoid $(\{0, 1, -1\}, \cdot)$ has a subsemigroup $\{0\}$ (since $0 \cdot 0 = 0$). But this subsemigroup $\{0\}$ is not a submonoid of $(\{0, 1, -1\}, \cdot)$, since it does not contain the neutral element 1 of the latter monoid.
- What are the subgroups of $(\mathbb{Z}, +)$? For any integer n , the set

$$n\mathbb{Z} := \{\text{all integer multiples of } n\}$$

is a subgroup of $(\mathbb{Z}, +)$, since it is closed under addition (because $na + nb = n(a + b)$ for all $a, b \in \mathbb{Z}$), contains the neutral element (because $0 = n0$) and is closed under taking inverses (because $-na = n(-a)$ for all $a \in \mathbb{Z}$).

Note that $0\mathbb{Z} = \{0\}$, so that too is a subgroup. And $1\mathbb{Z} = \mathbb{Z}$ is a subgroup as well.

Are there any others? We will see that the answer is “no”; we’ve found them all.

The most boring examples are ones that work in the greatest generality:

Proposition 1.11.4. (a) If M is any monoid (written multiplicatively), then $\{1\}$ and M itself are submonoids of M .

(b) If G is any group (written multiplicatively), then $\{1\}$ and G itself are subgroups of G .

(c) If S is a semigroup, then $\emptyset$ and S itself are subsemigroups of S . (Recall that the empty set is a semigroup, but not a monoid or a group.)

Proof. LTTR. (The subset $\{1\}$ is closed under taking inverses since 1 is its own inverse. Everything else is entirely trivial.) $\square$

One important feature of subsemigroups is the following: Any subsemigroup T of a semigroup $(S, *)$ becomes a semigroup in its own right, simply by inheriting the operation $*$ from S (that is, for any $a, b \in T$, we define the “product” $a * b$ in T to be their original “product” $a * b$ in S ; the definition of a subsemigroup ensures that this “product” does belong to T). Let us codify this:

Theorem 1.11.5. Let $(S, *)$ be a semigroup, and T a subsemigroup of $(S, *)$. Define a new operation $*_T : T \times T \rightarrow T$ by setting

$$a *_T b = a * b \quad \text{for all } a, b \in T.$$

Then, $(T, *_T)$ is again a semigroup.

Proof. Since T is closed under $*$, the operation $*_T$ takes its values in T . Thus, $*_T$ is a well-defined binary operation on T . The associativity of $*_T$ follows directly from associativity of $*$, since for any $a, b, c \in T$ we have

$$\begin{aligned} a *_T (b *_T c) &= a * (b * c) & \text{and} \\ (a *_T b) *_T c &= (a * b) * c. \end{aligned}$$

This shows that $(T, *_T)$ is a semigroup. $\square$

The same construction, applied to submonoids and subgroups, yields monoids and groups again:

Theorem 1.11.6. Let $(M, *)$ be a monoid, and N a submonoid of $(M, *)$. Then, the semigroup $(N, *_N)$ (defined as in Theorem 1.11.5 for $S = M$ and $T = N$) is a monoid.

Proof. The submonoid N is a subsemigroup of M ; thus, by Theorem 1.11.5, we know that $(N, *_N)$ is a semigroup. The neutral element of M belongs to N by the definition of a submonoid, and is still neutral in $(N, *_N)$ (because $*_N$ is just a restriction of $*$). Thus, $(N, *_N)$ is a monoid. $\square$

Theorem 1.11.7. Let $(G, *)$ be a group, and H a subgroup of $(G, *)$. Then, the monoid $(H, *_H)$ (defined as in Theorem 1.11.6 for $M = G$ and $N = H$) is a group. Moreover, the inverse a^{-1} of any element $a \in H$ computed in this group $(H, *_H)$ is precisely the inverse of a computed in $(G, *)$.

Proof. The subgroup H is a submonoid of G , so Theorem 1.11.6 shows that $(H, *_H)$ is a monoid with the same neutral element as G . For every $a \in H$, its inverse a^{-1} in G belongs to H by the definition of a subgroup. Moreover, it still plays the role of inverse of a in H , because the operation $*_H$ of H is just a restriction of $*$ (so we have $a *_H a^{-1} = a * a^{-1} = e$ and likewise $a^{-1} *_H a = e$). Thus, each $a \in H$ has an inverse in H . Therefore, the monoid $(H, *_H)$ is a group. $\square$

Definition 1.11.8. Theorem 1.11.5, Theorem 1.11.6 and Theorem 1.11.7 will often be used automatically, without explicit mention. Thus, when we say that H is a subgroup of a group $(G, *)$, we will take it as self-evident that this makes $(H, *_H)$ into a group (by Theorem 1.11.7), and we shall simply refer to this group $(H, *_H)$ as H without mentioning $*_H$. When we do want to mention its operation, we will often write it as $*$ instead of $*_H$, since it is just a restriction of $*$; this is a bit sloppy but should cause no trouble. (After all, we are already using the same symbol $+$ for the operations of the groups $(\mathbb{Z}, +)$ and $(\mathbb{Q}, +)$.) Conversely, when we say that a group $(H, \odot)$ is a subgroup of a group $(G, *)$, this will always be understood to mean that H is a subgroup of $(G, *)$ and that $\odot$ is the operation $*_H$ obtained by restricting $*$ to $H \times H$.

All this applies similarly to submonoids and subsemigroups.

Thus, for example, given any set A , the subgroup $\mathcal{P}_{\text{even}}(A)$ of $(\mathcal{P}(A), \Delta)$ becomes itself a group $(\mathcal{P}_{\text{even}}(A), \Delta)$.

Warning 1.11.9. A subsemigroup N of a monoid $(M, *)$ can be a monoid itself (i.e., the semigroup $(N, *)$ can be a monoid) without being a submonoid of $(M, *)$! For example, the monoid $(\{0\}, \cdot)$ is a subsemigroup of the monoid $(\{0, 1, -1\}, \cdot)$, but not a submonoid of $(\{0, 1, -1\}, \cdot)$, because its neutral element is 0, not 1. A submonoid has to contain the neutral element of the whole monoid, not its own different neutral element!

Remark 1.11.10. Let G be a group (written multiplicatively), and let H be a subgroup of G . Let $a \in H$. Then, any power a^n of a computed in the group H equals the corresponding power a^n computed in G . This is because a^n is defined entirely in terms of the defining features of a group (multiplication, neutral element and inverses), and all these features for H are restrictions of the corresponding features for G (for example, the inverses in H are the same as in G). The same applies to other features such as orders and commutativity. However, an element $a \in H$ can be central in H without being central

in G , because the group H may have fewer elements to “veto” its centrality than G does!

1.12. The center of a semigroup/monoid/group

When a semigroup (or monoid, or group) is not abelian, one may still study the extent to which commutativity holds for some of its elements. This leads to the following definition:

Definition 1.12.1. An element a of a semigroup S is said to be **central** if it commutes with every $b \in S$, that is, if $ab = ba$ for all $b \in S$ (where we write S multiplicatively).

Definition 1.12.2. The **center** $Z(S)$ of a semigroup S is the set of all central elements of S .

Theorem 1.12.3.

- (a) The center $Z(S)$ of a semigroup S is a subsemigroup of S .
- (b) The center $Z(M)$ of a monoid M is a submonoid of M .
- (c) The center $Z(G)$ of a group G is a subgroup of G .

Proof. We shall write S , M and G multiplicatively.

- (a) We must show that $Z(S)$ is closed under multiplication. That is, we must show that if $a_1, a_2 \in Z(S)$, then $a_1a_2 \in Z(S)$.
So let $a_1, a_2 \in Z(S)$. That is, a_1 and a_2 are central. Now, for each $b \in S$, we have

$$a_1 \underbrace{a_2 b}_{=ba_2} = \underbrace{a_1 b}_{=ba_1} a_2 = ba_1a_2.$$

(since a_2 is central) (since a_1 is central)

This shows that a_1a_2 is central. In other words, $a_1a_2 \in Z(S)$, just as we wished to prove. Thus, part (a) is proved.

- (b) Part (a) shows that $Z(M)$ is a subsemigroup of M . It remains to show that it also contains the neutral element 1 of M . In other words, we must show that 1 is central. But this is clear, since we have $1b = b = b1$ for each $b \in M$.
- (c) Part (b) shows that $Z(G)$ is a submonoid of G . It remains to prove that $Z(G)$ is closed under taking inverses. So we must show that each $a \in Z(G)$ satisfies $a^{-1} \in Z(G)$.

Let $a \in Z(G)$. We must show that $a^{-1} \in Z(G)$.

Let $b \in G$. Since a is central (because $a \in Z(G)$), we have $ab = ba$. Multiplying this equality by a^{-1} from the left, we obtain $a^{-1}ab = a^{-1}ba$.

In other words, $b = a^{-1}ba$ (since $\underbrace{a^{-1}a}_{=1}b = 1b = b$). Multiplying this latter equality by a^{-1} from the right, we obtain $ba^{-1} = a^{-1}b\underbrace{aa^{-1}}_{=1} = a^{-1}b1 = a^{-1}b$. In other words, $a^{-1}b = ba^{-1}$.

Thus, we have shown that $a^{-1}b = ba^{-1}$ for each $b \in G$. This means precisely that a^{-1} is central. That is, $a^{-1} \in Z(G)$. This completes our proof of part (c). $\square$

Example 1.12.4. If S is an abelian semigroup, then $Z(S) = S$ (since each $a \in S$ is central).

Example 1.12.5. Let $n \geq 3$. Consider the symmetric group S_n (that is, the group of all permutations of $\{1, 2, \dots, n\}$). It can be shown (see homework set #2) that $Z(S_n) = \{\text{id}\}$. That is, the **only** permutation $w \in S_n$ that commutes with all permutations in S_n is the identity id .

Example 1.12.6. Let $n \geq 1$. Consider the group $\text{GL}_n(\mathbb{R})$ of all invertible $n \times n$ -matrices with real entries (under multiplication). What is the center $Z(\text{GL}_n(\mathbb{R}))$?

A moment's thought suggests that each diagonal matrix with all diagonal entries equal (i.e., each matrix of the form $\lambda I_n = \text{diag}(\lambda, \lambda, \dots, \lambda)$ for $\lambda \in \mathbb{R} \setminus \{0\}$) is central in $\text{GL}_n(\mathbb{R})$. And this is indeed true, since every further $n \times n$ -matrix B satisfies $(\lambda I_n)B = \lambda B = B(\lambda I_n)$. Are there any other central matrices in $\text{GL}_n(\mathbb{R})$?