Chapter 5

Binary Operations

In the last lecture, we introduced the residue classes $\mathbb{Z}_n$ together with their addition and multiplication. We have also shown some properties that these two operations have. Today, we will study sets together with one operation in general and try to derive some properties that can be used regardless how an operation is defined and what elements a set has. We will define item-by-item groupids (the most general case), semigroups (groupoids that satisfy the associativity law), monoids (semigroups with a neutral element), and groups (monoids where every element is invertible).

5.1 Groupoids, Semigroups, Monoids

5.1.1 Groupoids. The most general notion of this section is the notion of a groupoid.

Definition. A *binary operation on a set* S is any mapping from the set of all pairs $S \times S$ into the set S.

A pair $(S, \circ)$ where S is a set and $\circ$ is a binary operation on S is called a *groupoid*. □

Note that the only condition for a binary operation on S is that **for every** pair of elements of S their result must be defined and must be an element in S.

A binary operation is usually denoted by $\cdot$, or $+$, $\circ$, $\star$ etc. (A binary operation $\circ$ assigns to elements x, y the element $x \circ y$.)

Examples of groupoids. The following are groupoids.

1) $(\mathbb{R}, +)$ where $+$ is addition on the set of all real numbers.
2) $(\mathbb{Z}, +)$ where $+$ is addition on the set of all integers.
3) $(\mathbb{N}, +)$ where $+$ is addition on the set of all natural numbers.
4) $(\mathbb{R}, \cdot)$ where $\cdot$ is multiplication on the set of all real numbers.
5) $(\mathbb{Z}, \cdot)$ where $\cdot$ is multiplication on the set of all integers.
6) $(M_n, \cdot)$ where M_n is the set of all square matrices of order n, and $\cdot$ is multiplication of matrices.
7) $(\mathbb{Z}_n, \oplus)$ for any $n > 1$.
8) $(\mathbb{Z}_n, \odot)$ for any $n > 1$.
9) $(\mathbb{Z}, -)$, where $-$ is subtraction on the set of all integers.

Examples which are not groupoids.

- $(\mathbb{N}, -)$ is not a groupoid because subtraction is not a binary operation on $\mathbb{N}$. Indeed, $3 - 4$ is not a natural number.
- $(\mathbb{Q}, :)$, where $:$ is the division, because $1 : 0$ is not defined.

5.1.2 Semigroups. General groupoids are structures where it is rather difficult to "calculate". Indeed, if we want to "multiply" four elements we must know in which order to do it. It means whether it is $a \circ ((b \circ c) \circ d)$, or $a \circ ((b \circ c) \circ d)$, or one of the other two possibilities. First, we will be interested in groupoids where we do not need to use brackets, these will be groupoids where the associative law holds.

Definition. Given a groupoid $(S, \circ)$. If for every $x, y, z \in S$ we have

$$x \circ (y \circ z) = (x \circ y) \circ z \tag{5.1}$$

$(S, \circ)$ is called a *semigroup.* □

The property 5.1 is called the *associative law.*

Examples of semigroups. The following groupoids are semigroups:

1) $(\mathbb{R}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{N}, +)$.
2) $(\mathbb{R}, \cdot)$, $(\mathbb{Z}, \cdot)$, $(\mathbb{N}, \cdot)$.
3) $(\mathbb{Z}_n, \oplus)$, $(\mathbb{Z}_n, \odot)$.
4) $(M_n, +)$, $(M_n, \cdot)$, where M_n is the set of square real matrices of order n and $+$ and $\cdot$ is addition and multiplication, respectively, of matrices.
5) $(A, \circ)$ where A is the set of all mappings $f\colon X \to X$ for a set X, and $\circ$ is the composition of mappings.

Examples of groupoids which are not semigroups.

- $(\mathbb{Z}, -)$, i.e. the set of all integers with subtraction. Indeed, $2 - (3 - 4) = 3$ but $(2 - 3) - 4 = -5$.
- $(\mathbb{R} \setminus \{0\}, :)$, i.e. the set of nonzero real numbers together with the division :. Indeed, $4 : (2 : 4) = 8$, but $(4 : 2) : 4 = \frac{1}{2}$.

5.1.3 Neutral (Identity) Element. A groupoid $(S, \circ)$ may or may not have an element that "does not change" anything if it is used. The precise definition is given bellow.

Definition. Given a groupoid $(S, \circ)$. An element $e \in S$ is called a *neutral* (also *identity*) element if

$$e \circ x = x = x \circ e \quad \text{for every } x \in S. \tag{5.2}$$

□

If the operation is denoted by $\cdot$ then we usually use the term "identity element" instead of a neutral element.

Examples of neutral elements.

1) For $(\mathbb{R}, +)$ the number 0 is its neutral element, the same holds for $(\mathbb{Z}, +)$.
2) For $(\mathbb{R}, \cdot)$ the number 1 is its neutral (identity) element, the same holds for $(\mathbb{Z}, \cdot)$, and $(\mathbb{N}, \cdot)$.
3) For $(M_n, \cdot)$ where $\cdot$ is the multiplication of square matrices of order n the identity matrix is its neutral (identity) element.
4) $(\mathbb{Z}_n, \oplus)$ has the class $[0]_n$ as its neutral element.
5) $(\mathbb{Z}_n, \odot)$ has the class $[1]_n$ as its neutral (identity) element.

Example of a groupoid that does not have a neutral element. The groupoid $(\mathbb{N} \setminus \{0\}, +)$ does not have a neutral element. Indeed, there is not a positive number e for which $n + e = n = e + n$ for every positive $n \in \mathbb{N}$

5.1.4 Uniqueness of the Neutral Element. The following proposition shows that if a groupoid $(S, \circ)$ has its neutral element then it is unique.

Proposition. Given a groupoid $(S, \circ)$. If there exist elements e and f such that for every $x \in S$ we have $e \circ x = x$ and $x \circ f = x$, then $e = f$ is the neutral element of $(S, \circ)$. □

Justification. Consider the product $e \circ f$. From the property of e we have $e \circ f = f$ (indeed, take $x = f$); from the property of f we have $e \circ f = e$ (indeed, take $x = e$). Hence $e = f$, and in this case e is the neutral element. □

5.1.5 Monoids. We will be mainly interested in semigroups which have the neutral element; they will be called monoids.

Definition. If in a semigroup $(S, \circ)$ there exists a neutral element then we call $(S, \circ)$ a *monoid*. □

In the paragraph above, we gave couple of examples of monoids and also an example of a semigroup which is not a monoid.

Convention. In the following text, the fact that $(S, \circ)$ is a monoid with the neutral element e will be shortened to $(S, \circ, e)$.

5.1.6 Powers in a Monoid. Similarly as powers are defined in $(\mathbb{R}, \circ, 1)$ we can introduce powers in an arbitrary monoid.

Definition. Given a monoid $(S, \circ, e)$ and its element $a \in S$. The *powers* of a are defined by:

$$a^0 = e, \;\; a^{i+1} = a^i \circ a \;\text{ for every } i \geq 0.$$

□

Note that if the operation is $+$ with neutral element 0 then we write $0\,a = 0$ instead of a^0 and $k\,a$ instead of a^k.

5.1.7 Invertible Elements. In many examples given above, we can somehow "reverse" the operation. For instance, in $(\mathbb{R}, +, 0)$ we can subtract; in $(\mathbb{R}, \cdot, 1)$ we can divide by any nonzero number; in $(M_n, \cdot, E)$ where M_n is the set of all square matrices of order n, and E is the identity matrix, we can cancel all the regular matrices (this means multiplying by the inverse matrix to a given regular one). In this paragraph, roughly speaking, we characterize those elements of a monoid that not only permit "cancellation" but "help solving equations". More precisely:

Definition. Given a monoid $(S, \circ, e)$. We say that an element $a \in S$ is *invertible* if there exists an element $y \in S$ such that

$$a \circ y = e = y \circ a. \tag{5.3}$$

□

Let us show that if y from 5.3 exists then it is unique.

Proposition. Given a monoid $(S, \circ, e)$. Assume that there are elements $a, x, y \in S$ such that

$$x \circ a = e \;\text{ and }\; a \circ y = e,$$

then $x = y$. □

Justification. Consider the product $x \circ a \circ y$. Since we are in a semigroup it holds that

$$y = e \circ y = (x \circ a) \circ y = x \circ (a \circ y) = x \circ e = x.$$

□

5.1.8 The Inverse Element. Since y from 5.3 is unique we can define:

Definition. Let $(S, \circ, e)$ be a monoid, and $a \in S$ an invertible element. Let $y \in S$ satisfy

$$a \circ y = e = y \circ a.$$

Then y is called the *inverse element to* a and is denoted by a^{-1}. □

Remark. If a binary operation is denoted by $+$ we speak about the *opposite* element (instead of the inverse element) and denote it by $-a$ (instead of a^{-1}). The reason is that we sometimes have two different binary operations defined on the same set, (indeed, on the set $\mathbb{R}$ we have both $+$ and $\cdot$), hence it is convenient to distinguish between "inverses" with respect to $+$ and with respect to $\cdot$.

5.1.9 We know that not every element of a general monoid is invertible. Indeed, consider for example the set of all square matrices together with multiplication and the identity matrix. Then only regular matrices are invertible, and moreover for any regular matrix A it holds that $(A^{-1})^{-1} = A$. The next proposition shows that properties of invertible elements and their inverses are the same in any monoid.

Proposition. Let $(S, \circ, e)$ be a monoid. Then

1. e is invertible and $e^{-1} = e$.
2. If a is invertible then so is a^{-1}, and we have $(a^{-1})^{-1} = a$.
3. If a and b are invertible elements then so is $a \circ b$, and we have $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$.

□

Justification.
1. It suffices to notice that $e \circ e = e$, this immediately means that $e^{-1} = e$.

2. Assume that a is invertible. Then we have $a \circ a^{-1} = e = a^{-1} \circ a$. If we look at the last identities we see that a is the element such that if we multiply by it the element a^{-1} we get e. Hence $a = (a^{-1})^{-1}$.

3. Assume that a^{-1} and b^{-1} exist. Then

$$(a \circ b) \circ (b^{-1} \circ a^{-1}) = a \circ (b \circ b^{-1}) \circ a^{-1} = a \circ e \circ a^{-1} = a \circ a^{-1} = e.$$

Similarly, we get that $(b^{-1} \circ a^{-1}) \circ (a \circ b) = e$. We have shown that $(a \circ b)^{-1} = b^{-1} \circ a^{-1}$. □

Remark. Note that **it is not** always the case that $(a \circ b)^{-1} = a^{-1} \circ b^{-1}$. This holds when the operation $\circ$ is commutative, i.e. $x \circ y = y \circ x$ for every x and y.

5.1.10 An Invertible Element Can Be Canceled.

Proposition. Let $(S, \circ, e)$ be a monoid, and let $a \in S$ is its invertible element. Then

$$a \circ b = a \circ c, \text{ or } b \circ a = c \circ a \quad \text{implies} \quad b = c.$$

□

Justification. Assume that a^{-1} exists and

$$a \circ b = a \circ c. \tag{5.4}$$

Multiply 5.4 by a^{-1} form the left. We get

$$a^{-1} \circ (a \circ b) = a^{-1} \circ (a \circ c), \text{ which gives } (a^{-1} \circ a) \circ b = (a^{-1} \circ a) \circ c \text{ and } b = c.$$

Similarly for $b \circ a = c \circ a$. The only difference is that here we multiply by a^{-1} from the right. (Notice the similarity with matrix operations.) □

5.1.11 Groups. In couple of examples above, every element was invertible; indeed, it holds for $(\mathbb{Z}, +, 0)$, $(\mathbb{R} \setminus \{0\}, \cdot, 1)$, and $(\mathbb{Z}_n, +, 0)$. Such monoids are of great importance and they are called groups.

Definition. A monoid $(S, \circ, e)$ in which every element is invertible is called a *group*. □

Examples of groups. The following monoids are groups:

1) The monoid $(\mathbb{R}, +, 0)$. Indeed, for every $x \in \mathbb{R}$ there exists $-x$ for which $x + (-x) = 0 = (-x) + x$.
2) The monoid $(\mathbb{Z}, +, 0)$. Indeed, for each integer x there exists an integer $-x$ for which $x + (-x) = 0 = (-x) + x$.
3) The monoid $(\mathbb{R}^+, \cdot, 1)$, where $\mathbb{R}^+$ is the set of all positive real numbers. Indeed, for every positive real number x there exists a positive real number $\frac{1}{x}$ for which $x \cdot \frac{1}{x} = 1 = \frac{1}{x} \cdot x$.

4) The monoid $(\mathbb{Z}_n, \oplus, [0]_n)$. Indeed, for a class $[i]_n$ there exists a class $[n-i]_n$ for which $[i]_n \oplus [n-i]_n = [0]_n = [n-i]_n \oplus [i]_n$.
5) Let A be the set of all permutation of the set $\{1, 2, \ldots, n\}$, and let $\circ$ be the composition of permutations. Then $(A, \circ)$ is a monoid with the neutral element the identity permutation id. Moreover, for every permutation ϕ there exists its inverse permutation ϕ^{-1} for which $\phi \circ \phi^{-1} = id = \phi^{-1} \circ \phi$.

Examples of monoids that are not groups.

1) The monoid $(\mathbb{Z}, \cdot, 1)$. Indeed, for example 2 is not invertible because there is no **integer** k such that $2 \cdot k = 1$.
2) The monoid $(\mathbb{Z}_n, \odot, [1]_n)$. Indeed, the class $[0]_n$ is not invertible because for any $[i]_n$ we have $[0]_n \odot [i]_n = [0]_n \neq [1]_n$.
3) Let B be the set of all mappings from the set $\{1, 2, \ldots, n\}$ into itself, where $n > 1$. Let $\circ$ be the composition of mappings. Then $(B, \circ, id)$ is a monoid where id is the identity mapping. Any mapping that is not one-to-one is not invertible.

5.1.12 Groups can be characterized as those semigroups $(S, \circ)$ where every equation $a \circ x = b$ and $y \circ a = b$ has a solution. In that case, the solution is unique. From this it immediately follows that

1. If $(S, \circ)$ is not a group, then there is an equation which does not have a solution.
2. Given a semigroup $(S, \circ)$. If there exists an equation with two distinct solutions, then $(S, \circ)$ is not a group, and moreover there is an equation that does not have a solution.

The following two paragraphs prove it.

5.1.13 Proposition. Given a group $(S, \circ)$ with its neutral element e. Then for every two elements $a, b \in S$ there exist unique $x, y \in S$ such that

$$a \circ x = b, \qquad y \circ a = b.$$

□

Justification. Since $(S, \circ, e)$ is a group and $a \in S$, there exists its inverse a^{-1}. If we multiply the equation $a \circ x = b$ by a^{-1} from the left we obtain

$$x = (a^{-1} \circ a) \circ x = a^{-1} \circ (a \circ x) = a^{-1} \circ b.$$

Similarly we obtain $y = b \circ a^{-1}$ from the second equation; indeed, we multiply the second equation by a^{-1} from the right and get the desired solution.

Let us show the uniqueness. Assume that $a \circ x_1 = b$ and $a \circ x_2 = b$. Then $a \circ x_1 = a \circ x_2$. Now, the proposition 5.1.10 completes the argument because it states that $x_1 = x_2$. Similarly from $y_1 \circ a = b$ and $y_2 \circ a = b$ we get $y_1 = y_2$.

5.1.14 Theorem. A semigroup $(S, \circ)$ is a group if and only if every equation of the form $a \circ x = b$ and every equation of the form $y \circ a = b$ has at least one solution.

More precisely: A semigroup $(S, \circ)$ is a group if and only if for every two elements $a, b \in S$ there exist $x, y \in S$ such that $a \circ x = b$ and $y \circ a = b$. □

Justification. First we show that if a semigroup $(S, \circ)$ satisfies the above conditions then it has got a neutral element.

Choose any $a \in S$. There exists $e_a \in S$ such that $e_a \circ a = a$; indeed, it is a solution of $y \circ a = a$. Now, take an arbitrary $b \in S$. We know that $b = a \circ x$ for some $x \in S$, hence

$$e_a \circ b = e_a \circ (a \circ x) = (e_a \circ a) \circ x = a \circ x = b.$$

Similarly, it can be shown that the element f_a for which $a \circ f_a = a$ satisfies $b \circ f_a = b$ for any $b \in S$.

Therefore, from 5.1.4 we get that $e_a = f_a$ is the neutral element of $(S, \cdot)$.

To show that every element $a \in S$ is invertible, it suffices to use the proposition from 5.1.7. Indeed, from the fact that there exist $x, y \in S$ with $a \circ x = e$ and $y \circ a = e$ we know that $x = y$, and $x = a^{-1}$. So, a is invertible. Since a was an arbitrary element of S, $(S, \circ, e)$ is a group. □

5.1.15 Commutative Semigroups, Monoids, Groups. In many examples above (but not in all) it does not matter whether we calculate $a \circ b$ or $b \circ a$, we get the same results.

Definition. A semigroup $(S, \circ)$ (monoid, group) is called *commutative* if it satisfies the *commutative law*, i.e. for every two elements $x, y \in S$

$$x \circ y = y \circ x.$$

□

5.1.16 Subsemigroups. Given a semigroup $(S, \circ)$ and a set $T \subseteq S$. It may happen (but does not need to) that T together with the same operation $\circ$ is again a semigroup. In that case, we will call $(T, \circ)$ a subsemigroup of $(S, \circ)$.

Definition. Given a semigroup $(S, \circ)$. A subset $T \subseteq S$ together with an operation $\circ$ forms a *subsemigroup* of the semigroup $(S, \circ)$, if for every two elements $x, y \in T$ we have $x \circ y \in T$. (In this case $(T, \circ)$ is also a semigroup.) □

Remark. Next, we will say less exactly "T is a subsemigroup" instead of "T forms a subsemigroup". It will be mainly in the situation where the operation is clear from the context.

Examples of subsemigroups. The following are examples of subsemigroups:

1) $\mathbb{N}$ together with addition forms a subsemigroup of $(\mathbb{Z}, +)$.
2) The set of all regular matrices together with multiplication of matrices forms a subsemigroup of $(M_n, \cdot)$, where M_n is the set of all square matrices of order n.
3) The set of all positive real numbers together with multiplication forms a subsemigroup of $(\mathbb{R}, \cdot)$.

Example of a subset that does not form a subsemigroup. The set of all regular square matrices of order n together with addition of matrices does not form a subsemigroup of $(M_n, +)$. Indeed, it does not hold that sum of two regular matrices is a regular matrix, e.g. coincide the identity matrix E. Then E and $-E$ are regular matrices but $E + (-E)$ is the zero matrix which is not regular.

5.1.17 Submonoids.

Definition. Given a monoid $(S, \circ, e)$. A subset $T \subseteq S$ forms a submonoid if it forms a subsemigroup and moreover $e \in T$. (In this case $(T, \circ, e)$ is also a monoid.) □

Examples of submonoids.

1) The set of all natural numbers $\mathbb{N}$ together with addition is a submonoid of $(\mathbb{Z}, +, 0)$, since $0 \in \mathbb{N}$.
2) The set of all regular square matrices of order n together with multiplication of matrices forms a submonoid of $(M_n, \cdot, E)$, since the identity matrix E is regular.
3) Denote by T_X the set of all mappings from a set X into itself. Consider the operation composition of mappings $\circ$. Then $(T_X, \circ, id)$ where id is the identity mapping (defined by $id(x) = x$ for all $x \in X$) is a monoid. The set of all bijections from T_X forms a submonoid of $(T_X, \circ)$, indeed, a composition of two bijections is a bijection, and the identity mapping is a bijection.

5.1.18 Remark. Notice that a subsemigroup $(T, \circ)$ of $(S, \circ, e)$ may contain a neutral element which is different from the neutral element e (but in this case $e \notin T$). If this is the case $(T, \circ)$ is a subsemigroup of $(S, \circ)$ but not a submonoid of $(S, \circ, e)$. Next, there is an example of such a situation.

Example. Let $X = \{1, 2, 3\}$. Denote by S the set of all mappings from X to X. Then $(S, \circ, id)$ is a monoid ($\circ$ is the composition of mappings, id is the identity mapping).

Consider the mapping $f\colon X \to X$ defined by $f(1) = 2$, $f(2) = 3$, $f(3) = 4$, and $f(4) = 2$. Then $f^4 = f$ and $T = \{f, f^2, f^3\}$ forms a subsemigroup of $(S, \circ, id)$. T does not form a submonoid, since $id \notin T$. On the other hand, f^3 is the neutral element of $(T, \circ)$ and $(T, \circ, f^3)$ is in fact a group. Indeed, $f \circ f^3 = f = f^3 \circ f$, $f^2 \circ f^3 = f^2 = f^3 \circ f^2$, and $f^3 \circ f^3 = f^3$.