Algebraic numbers, transcendental numbers (Optional !)

Definition:

(1) We use the notation $\mathbb{Z}[x]$ for the set of all polynomials with integer coefficients. So a typical element of $\mathbb{Z}[x]$ is of the form

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

where each $a_i \in \mathbb{Z}$. If $a_n \neq 0$ then a_n is called the leading coefficient of $p(x)$ and n is called the degree of $p(x)$. If each $a_i = 0$ (i.e., $p(x) = 0$) then $p(x) = 0$ is called the zero polynomial and its degree is undefined.

(2) A real number is called an algebraic number if it is a root of a nonzero polynomial in $\mathbb{Z}[x]$.

(3) Any real number which is not algebraic is called a transcendental number.

Ex: (1) $\frac{7}{8}$ is an algebraic number because it is a root of $0 \neq 8x - 7 \in \mathbb{Z}[x]$.

(2) Any rational number is algebraic because, for any $a, b \in \mathbb{Z}$ with $b \neq 0$, $\frac{a}{b}$ is a root of $0 \neq bx - a \in \mathbb{Z}[x]$.

(3) $\sqrt[3]{5}$ is algebraic because it is a root of $0 \neq x^3 - 5 \in \mathbb{Z}[x]$

(4) $1 + \sqrt{2}$ is algebraic because it is a root of $0 \neq x^2 - 2x - 1 \in \mathbb{Z}[x]$

Let $\mathbb{A} = \{\alpha \in \mathbb{R} \mid \alpha \text{ is algebraic}\}$. By the previous example $\mathbb{Q} \subsetneq \mathbb{A} \subseteq \mathbb{R}$
Now $\mathbb{R} - \mathbb{A}$ is the set of transcendental numbers. Is it nonempty? That is, is there a transcendental number? The answer had not known until 1820, when Cantor proved that there are uncountably many transcendental numbers. However, Cantor proved only existence and did not give an example of a transcendental number. The first example of a transcendental number was found in 1844, which is $\sum_{k=1}^{\infty} 10^{-k!}$

Here we will prove that $\mathbb{A}$ is countable. As $\mathbb{R}$ is uncountable, it follows that $\mathbb{R} - \mathbb{A}$ is uncountable (and $\mathbb{R} - \mathbb{A} \sim \mathbb{R}$).

Remark: For any $d \in \mathbb{N}$ let $Z_d = \{p(x) \in \mathbb{Z}[x] \mid \deg(p(x)) = d\}$ be the set of all polynomials $p(x) \in \mathbb{Z}[x]$ of degree d.

(1) The map $Z_d \to \mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}$ ($d+1$ times) is injective.

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d \mapsto (a_0, a_1, a_2, \ldots, a_d)$$

(2) Z_d is countable for all $d \in \mathbb{N}$.

(3) $\mathbb{Z}[x] - \{0\} = \bigcup_{d=0}^{\infty} Z_d$

(4) $\mathbb{Z}[x] - \{0\}$ is countable

(5) For any $p(x) \in \mathbb{Z}[x] - \{0\}$ the set $\{r \in \mathbb{R} \mid p(r) = 0\}$ of all the real roots of $p(x)$ is finite.

(6) $\mathbb{A} = \bigcup_{p(x)\in\mathbb{Z}[x]-\{0\}} \{r\in\mathbb{R} \mid p(r)=0\}$

(7) The set $\mathbb{A}$ of algebraic numbers is countable.
The set $\mathbb{R}-\mathbb{A}$ of transcendental numbers is uncountable.

Proof:

(1): This is because two polynomials are equal iff their coefficients of all like powers of x are equal.

(2): As the cartesian product of finitely many countable sets is countable and as $\mathbb{Z}$ is countable, $\mathbb{Z}\times\mathbb{Z}\times\cdots\times\mathbb{Z}$ ($d+1$ times) is countable. Since there is an injective function from $\mathbb{Z}_d$ to a countable set, it follows that $\mathbb{Z}_d$ is countable.

(3): Any nonzero polynomial in $\mathbb{Z}[x]$ must have a degree and so must be in some $\mathbb{Z}_d$. This implies the result.

(4): From (3) and (2) we know that $\mathbb{Z}[x]-\{0\}$ is the union of countably many countable sets $(\mathbb{Z}_0, \mathbb{Z}_1, \mathbb{Z}_2, \ldots, \mathbb{Z}_n, \ldots)$. As the union of countably many countable sets is countable, we see that $\mathbb{Z}[x]-\{0\}$ is countable.

(5): As a nonzero polynomial $q(x)$ can have at most $\deg(q(x))$ roots, the result follows.

(6): It is clear that $\bigcup_{p(x)\in\mathbb{Z}[x]-\{0\}} \{r\in\mathbb{R} \mid p(r)=0\}$ is the set of all real numbers that are roots of some nonzero polynomials with integer coefficients

(7): We know from (6) and (5) that $/\!A$ is the union of finite sets $\{r \in \mathbb{R} \mid p(r)=0\}$ where $p(x)$ ranges in $\mathbb{Z}[x]-\{0\}$. The number of sets appearing in the union $/\!A = \bigcup_{p(x)\in\mathbb{Z}[x]-\{0\}} \{r \in \mathbb{R} \mid p(r)=0\}$ is $|\mathbb{Z}[x]-\{0\}|$ which is countable by (4). Thus $/\!A$ is the union of countably many finite (so countable) sets. As the union of countably many countable sets is countable, $/\!A$ must be countable.

For $\mathbb{R}-/\!A$, note that $\mathbb{R} = (\mathbb{R}-/\!A) \cup /\!A$. We have proved that $/\!A$ is countable. If $\mathbb{R}-/\!A$ were countable then, being the union of two countable sets, $\mathbb{R}$ would be countable. But we know (from a theorem of Cantor) that $\mathbb{R}$ is uncountable. Hence, $\mathbb{R}-/\!A$ must be uncountable. □

Time	Some History	Questions & Findings
(-1820)		Is there a transcendental number?
(1820)	There are uncountably many transcendental numbers	
(-1844)		What are some examples of transcendental numbers?
(1844)	First example of a transcendental number was found. ($\sum_{k=1}^{\infty} 10^{-k!}$ is transcendental)	
(1873)	It was proved that e is transcendental	
(1882)	It was proved that π is transcendental	

In general it is not known that whether the sum of two transcendental number is transcendental or algebraic. For instance, "is $e+\pi$ transcendental?". The answer is not known, even today.

Continuum Hypothesis, Aleph notation, an application of the Zorn's Lemma (Optional !)

Recall that $|\mathbb{N}^+| < |\mathcal{P}(\mathbb{N}^+)| = |\mathbb{R}|$ by some theorems of Cantor. Is there a set A such that $|\mathbb{N}^+| < |A| < |\mathbb{R}|$? Works of Gödel and Cohen indicated that no contradiction would arise if the answer to this question were Yes or No. This means that the answer to this question is independent of the axioms of the set theory we assumed. Cantor believed that the answer to this question is No.

Continuum Hypothesis (CH for short): There is no set A such that

$$|\mathbb{N}^+| < |A| < |\mathbb{R}|$$

Generalized Continuum Hypothesis (GCH for short): For any infinite set A, there is no set B such that $|A| < |B| < |\mathcal{P}(A)|$

Aleph numbers ($\aleph_0, \aleph_1, \aleph_2, \ldots$) are used to represent the cardinalities of infinite sets: $\aleph_0$ denotes the cardinality of an infinite set whose cardinality is the smallest among all infinite sets, so $\aleph_0 = |\mathbb{N}^+|$. $\aleph_1$ denotes the cardinality of an infinite set whose cardinality is the smallest among all infinite sets A such that $|A| > \aleph_0$. And so on. The cardinality of real numbers is sometimes called continuum and denoted by c. So $|\mathbb{R}| = c$.

If we assume the GCH, then we have

$$|\mathbb{N}^+| < |\mathcal{P}(\mathbb{N}^+)| < |\mathcal{P}(\mathcal{P}(\mathbb{N}^+))| < |\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N}^+)))| < \cdots$$

$$\| \qquad\quad \| \qquad\quad \| \qquad\quad \|$$

$$\aleph_0 \qquad |\mathbb{R}| \qquad |\mathcal{P}(\mathbb{R})| \qquad |\mathcal{P}(\mathcal{P}(\mathbb{R}))|$$

$$\| \qquad\quad \| \qquad\quad \|$$

$$\qquad\quad \aleph_1 \qquad\quad \aleph_2 \qquad\quad \aleph_3$$

$$\qquad\quad \|$$

$$\qquad\quad c$$

Recall that $\mathbb{N}^+ \times \mathbb{N}^+ \sim \mathbb{N}^+$. We may also see that the map $(0,1)\times(0,1) \longrightarrow (0,1)$ given by $(0.a_1a_2a_3\cdots, 0.b_1b_2b_3\cdots) \mapsto 0.a_1b_1a_2b_2a_3b_3\cdots$ is bijective where for any $x \in (0,1)$ the notation $0.x_1x_2x_3\cdots$ is the decimal expansion of x with the convention that we use the infinite decimal expansions for rationals such as $1/2 = 0.4999\cdots9\cdots$ not 0.5. Hence, $(0,1)\times(0,1) \sim (0,1)$; so $\mathbb{R}\times\mathbb{R} \sim \mathbb{R}$. In fact, this is true in general whose proof uses Zorn's Lemma (and so omitted here).

<u>Theorem:</u> $A \times A \sim A$ for any infinite set A.

—— o ——

15B. The Pigeonhole Principle

If a mail carrier has m letters to distribute among n mailboxes (or "pigeonholes"), and $m > n$, then it seems clear that at least one of the mailboxes will have to get more than one letter. This important observation is known as the "Pigeonhole Principle." See Exercise 16.18 for its proof.

PROPOSITION 15.20 (Pigeonhole Principle). *Let B and $A_1, A_2, \ldots, A_n$ be finite sets. If*

$$B \subset A_1 \cup A_2 \cup \ldots \cup A_n,$$

and $n < \#B$, then $\#A_i \geq 2$, for some i.

Here are a few of the many applications of the Pigeonhole Principle. In these real-world examples, our explanations will be a bit informal.

EXAMPLE 15.21. Bob's sock drawer has many, many socks in it, and they come in 4 colours. Unfortunately, the light in his room has burned out, so he cannot see anything. How many socks should he grab from the drawer, so that he can be sure at least two of them are of the same colour.

SOLUTION. Bob should grab 5 (or more) socks.

To see this, note, first, that taking 4 socks may not be enough: If Bob grabs only 4 socks, it is possible that he has one sock of each of the 4 different colours. Then he would not have two socks of the same colour.

Now suppose Bob grabs (at least) 5 socks. He can sort them into 4 piles, by colour. Since $5 > 4$, one of the piles must have more than one sock. So there are (at least) 2 socks of the same colour. □

EXAMPLE 15.22. If you pick 50 numbers from 1 to 98, then it is guaranteed that two of them will add up to exactly 99.

SOLUTION. The numbers from 1 to 98 can be divided into 49 pigeonholes:

$$\{1, 98\},\ \{2, 97\},\ \{3, 96\},\ \ldots,\ \{49, 50\}.$$

(So two different numbers x and y are in the same pigeonhole iff $x+y = 99$.) Since $50 > 49$, two of the numbers we chose must be in the same pigeonhole. Then the sum of these two numbers is exactly 99. □

EXAMPLE 15.23. If you pick 5 points on the surface of a (spherical) orange, then there is always a way to cut the orange exactly in half, such that at least 4 of your points are in the same half. (We assume any point that is exactly on the cut is considered to belong to both halves.)

SOLUTION. Any two of the points will lie on a great circle of the sphere, so we can cut the orange so that 2 of the points are exactly on the cut. The other 3 points are distributed in some way among the two halves of the orange. By the Pigeonhole Principle, at least two of those three points are in the same half. Then that half contains the 2 points on the cut, plus these additional 2 points, for a total of (at least) 4 of the points you picked. □

EXERCISES 15.24.

1) It is known that:
   - No one has more than 300,000 hairs on their head.
   - More than a million people live in Calgary.

   Show that there are two people in Calgary who have exactly the same number of hairs on their heads.

2) Show that if you put 5 points into an equilateral triangle of side length 2 cm, then there are two of the points that are no more than 1 cm apart.
   [*Hint:* Divide the triangle into 4 equilateral triangle of side length 1 cm.]

In addition to the above real-world examples, the Pigeonhole Principle has important applications in theoretical mathematics.

COROLLARY 15.25. *Suppose A and B are finite sets, with $\#A = m$ and $\#B = n$.*

1) If there exists a one-to-one function $f\colon A \to B$, then $m \leq n$.

2) If there exists an onto function $f\colon A \to B$, then $m \geq n$.

PROOF. (1) Suppose $f\colon A \to B$ is onto, and $m > n$. There is no harm in assuming $B = \{1, 2, \ldots, n\}$, and then we may let

$$A_i = f^{-1}(i)$$

for $i = 1, 2, \ldots, n$. For any $a \in A$, we have $a \in f^{-1}\big(f(a)\big) = A_{f(a)}$, so $a \in A_1 \cup A_2 \cup \cdots \cup A_n$. Since a is an arbitrary element of A, this implies $A \subset A_1 \cup A_2 \cup \ldots \cup A_n$. Because $\#A = m > n$, we conclude that $\#A_i \geq 2$ for some i. This means $\#f^{-1}(i) > 1$, which contradicts the fact that f is one-to-one.

(2) Suppose $f\colon A \to B$ is onto, and $m < n$. There is no harm in assuming $A = \{1, 2, \ldots, m\}$, and then we may let

$$B_i = \{f(i)\}$$

for $i = 1, 2, \ldots, m$. Since f is onto, we know, for any $b \in B$, there is some $i \in A$, such that $f(i) = b$. This means $b \in B_i$; hence, $b \in B_1 \cup B_2 \cup \cdots \cup B_m$. Since b is an arbitrary element of B, this implies $B \subset B_1 \cup B_2 \cup \ldots \cup B_m$. Because $\#B = n > m$, we conclude that $\#B_i \geq 2$ for some i. This contradicts the fact that $\#B_i = 1$ (because $B_i = \{f(i)\}$ has only one element. □

EXAMPLE 15.33. Suppose a hotel has n rooms, numbered $1, 2, 3, \ldots, n$.

1) If A is a tour group of n people $a_1, a_2, \ldots, a_n$, then the hotel clerk will obviously have no trouble assigning each of them a room: a_i can be put in room i. There will be no empty rooms left.

2) On the other hand, if, in addition to this tour group, there is another person b who wants a room, then the situation is hopeless. There is no way to give each of these $n + 1$ people a room, without making two of them share a room. In general:

> If there are more guests than hotel rooms, then not everyone can have a room.

This is a restatement of the Pigeonhole Principle (15.20).

EXAMPLE 15.34 (Hotel Infinity). Now consider a hotel with a countably infinite number of rooms, numbered $1, 2, 3, \ldots$. (There is one room for each $i \in \mathbb{N}^+$.)

1) If A is a tour group of n people $a_1, a_2, \ldots, a_n$, then the hotel clerk will obviously have no trouble giving each of them a room: a_i can be put in room i. There will be lots of empty rooms left over.

2) Even if A is countably infinite, rather than finite, with people $a_1, a_2, \ldots$, the hotel clerk can accommodate all of them, by putting a_i in room i. There will be no empty rooms left.

3) Now suppose that, in addition to this tour group, there is another person b who also wants a room. The hotel clerk can handle this situation quite easily, by putting

   - b into room 1, and
   - a_i in room $i + 1$.

   Everyone will have his or her own room.

4) The same idea works, even if, instead of just one person, there is a whole group B of n people $b_1, b_2, \ldots, b_n$ that want rooms. The clerk can put

   - b_i in room i, and
   - a_i in room $i + n$.

5) It may seem that there would be a problem if the second group B consists of infinitely many people $b_1, b_2, b_3, \ldots$, but a clever hotel clerk can accommodate even this situation. Note that there are infinitely many odd-numbered rooms, so all of A can be put in those rooms, and there are also infinitely many even-numbered rooms, so all of B can be put in there. More precisely, the clerk can put

   - a_i in room $2i - 1$, and

- b_i in room $2i$.

6) Even if there are several of these countably infinite tour groups, not just 2 of them, they can all be accommodated. Namely, if there are n tour groups $A_1, A_2, A_3, \ldots, A_n$, then note that there are infinitely many numbers that are congruent to k modulo n, so all of A_k can be put in those rooms. More precisely, let $a_{k,1}, a_{k,2}, a_{k,3}, \ldots$ be a list of the people in A_k. Then the clerk can put

   - $a_{k,i}$ in room $k + in$.

   EXERCISES 15.35.

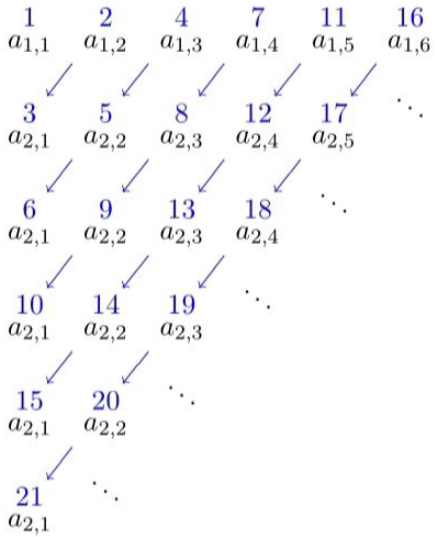
   (a) Show that the clerk has not put two guests in the same room. That is, show, for all $k_1, k_2 \in \{1, 2, \ldots, n\}$ and $i_1, i_2 \in \mathbb{N}^+$, such that $k_1 + i_1 n = k_2 + i_2 n$, we have $k_1 = k_2$ and $i_1 = i_2$.

   (b) The clerk has left some rooms empty, which is wasteful. (For example, it is not difficult to see that no guest has been put in room 1.) Modify the clerk's formula to obtain a method that does not leave any empty rooms (and, of course, does not put two guests in the same room.)

7) Going further, even if there were infinitely many infinite tour groups $A_1, A_2, \ldots$, they could all be accommodated. (We assume that each tour group is countably infinite, and that the number of groups is countably infinite.) To see this, start by considering an infinitely large table (or matrix) that lists the elements of A_k in the kth row:

$$\begin{array}{lccccccc}
A_1: & a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & a_{1,5} & \cdots \\
A_2: & a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & a_{2,5} & \cdots \\
A_3: & a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & a_{3,5} & \cdots \\
A_4: & a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} & \cdots \\
A_5: & a_{5,1} & a_{5,2} & a_{5,3} & a_{5,4} & a_{5,5} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}$$

We can assign rooms $1, 2, 3, \ldots$ to the entries of this table as indicated in the following figure:



The numbering:

- Begins with 1 in the top left corner.
- Then places 2 at the top of the second column and moves diagonally (down and to the left) to place 3.
- Then places 4 at the top of the third column, and moves diagonally (down and to the left) to place 5 and 6.
- Then places the next number (namely, 7) at the first open spot in the top row (namely, at the top of the fourth column), and moves diagonally (down and to the left) to place the following numbers (namely, 8, 9, and 10), until a number (namely, 10) is placed in the first column.
- The procedure then moves to the first open spot in the top row, and repeats infinitely.

No entries of the table are omitted from the numbering, and no room numbers are repeated, so each guest has his or her own room.

Remark 15.36. In Eg. 15.34(7), It can be shown that guest $a_{k,i}$ is given room

$$\frac{(k+i-1)(k+i-2)}{2}+k,$$

but we have no need for this formula.

It might seem that Hotel Infinity could accommodate every set of tourists, but that is not the case. For example, we will see later in the chapter that if all of the real numbers want rooms at the hotel, then some of them will have to share. In other words, the set $\mathbb{R}$ of real numbers is *uncountable.*

EXERCISES 15.47.

1) Suppose A is countably infinite, and $b \notin A$. Show, directly from the definition, that $A \cup \{b\}$ is countably infinite.

2) Suppose A is countably infinite, and $a \in A$. Show, directly from the definition, that $A \setminus \{a\}$ is countably infinite.

3) Suppose A and B are countably infinite and disjoint. Show, directly from the definition, that $A \cup B$ is countably infinite.

4) Suppose

- A_1 is disjoint from B_1,
- A_2 is disjoint from B_2,
- $A_1 \approx A_2$, and
- $B_1 \approx B_2$.

Show that $(A_1 \cup B_1) \approx (A_2 \cup B_2)$.

5) Suppose A is infinite. Show there is a *proper* subset B of A, such that $A \approx B$.
[*Hint:* Combine Thm. 15.40(1) with items 2 and 4.]

EXERCISES 15.50.

1) Show that the interval $(0, 1)$ is uncountable.
[*Hint:* $[0, 1) = (0, 1) \cup \{0\}$ is uncountable.]

2) Suppose $a, b \in \mathbb{R}$. Show that if $a < b$, then the interval (a, b) has the same cardinality as $(0, 1)$.
[*Hint:* Define $f : (0, 1) \to (a, b)$ by $f(x) = a + (b - a)x$.]

3) Suppose $a \in \mathbb{R}$. Show that the interval (a, ∞) has the same cardinality as $(0, 1)$.
[*Hint:* Define $f : (0, 1) \to (a, \infty)$ by $f(x) = (1/x) + a - 1$.]

4) Decide which of the following sets are countable, and which are uncountable. (*You do not need to justify your answers.*)

(a) $[3, 3.1]$.

(b) $\{1, 2, 3, \ldots, 1000\}$.

(c) $\mathbb{Z} \times \mathbb{Z}$.

(d) $\mathbb{Z} \times \mathbb{Q}$.

(e) $\mathbb{Z} \times \mathbb{R}$.

(f) $\mathbb{R} \setminus \mathbb{Q}$.

(g) $\{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \}$.

(h) $\{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = -1 \}$.

Semigroups

Definition: Let S be a nonempty set. By a binary operation $*$ on S we mean a function $*$ from $S \times S$ to S. For any $(a,b) \in S \times S$ we use the notation $a * b$ instead of $*((a,b))$ to denote the image of (a,b) under the function $*: S \times S \to S$.

Ex: The usual addition $+$ of integers is a binary operation on $\mathbb{Z}$. Indeed, $+: \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ is the function defined by

$$+((m,n)) = m+n$$

Remark: If $*$ is a binary operation on a nonempty set S, then

"$a * b \in S$ for all $a, b \in S$"

(This is called the closedness, and we say that S is closed under $*$)

Proof: As $*: S \times S \to S$ is a function, every element (a,b) in the domain must have (a unique) image $a * b$ in the codomain S. □

Definition: Let S be a nonempty set and $*$ be a binary operation on S. Consider the following properties of $*$ on S:

(0) $a * b \in S$ for all $a, b \in S$. (Closedness / S is closed under $*$)

(1) $(a * b) * c = a * (b * c)$ for all $a, b, c \in S$. (Associativity / $*$ is associative on S)

(2) There is an $e \in S$ such that $e * a = a = a * e$ for all $a \in S$ (Such an element e is called an identity of S with respect to $*$.)

(3) (Assuming that S has an identity) For any $a \in S$, there is an $a' \in S$ such that $a * a' = e = a' * a$ where e is an identity of S. (Such an element a' is called an inverse of a with respect to $*$)

(4) There is a $f \in S$ such that $a * f = f = f * a$ for all $a \in S$.
(Such an element f is called a zero element of S with respect to $*$)

(5) $a * b = b * a$ for all $a, b \in S$. (Commutativity / S is commutative with respect to $*$)

Then we say that:

a) $(S, *)$ is a semigroup (or S is a semigroup wrt $*$ (with respect to)) if (0) and (1) are satisfied.

b) $(S, *)$ is a monoid (or S is a monoid wrt $*$) if (0) and (1) and (2) are satisfied.

c) $(S, *)$ is a group (or S is a group wrt $*$) if (0) and (1) and (2) and (3) are satisfied

Moreover, if $(S, *)$ is a $\left\{\begin{matrix}\text{semigroup}\\ \text{monoid}\\ \text{group}\end{matrix}\right\}$ satisfying (5), then we say that $(S, *)$ is a commutative $\left\{\begin{matrix}\text{semigroup}\\ \text{monoid}\\ \text{group}\end{matrix}\right\}$.

Furthermore, if $(S, *)$ is a $\left\{\begin{matrix}\text{semigroup}\\ \text{monoid}\end{matrix}\right\}$ satisfying (4), then we say that $(S, *)$ is a $\left\{\begin{matrix}\text{semigroup}\\ \text{monoid}\end{matrix}\right\}$ with zero.

Remark: (1) $\{\text{groups}\} \subsetneq \{\text{monoids}\} \subsetneq \{\text{semigroups}\}$

(2) Monoid is a semigroup with identity.

(3) In a semigroup there is at most one identity element and there

is at most one zero element. So in a monoid/group there is a unique identity element (which is called the identity of the monoid/group, and denoted by e or 1)

(4) In a group each element a has a unique inverse (which is called the inverse of a, and denoted by a^{-1})

Proof: (1) and (2) are clear (exercise)

(3): Suppose for a contradiction that there is a semigroup $(S,*)$ having two distinct identity elements e_1 and e_2. Then,

$$e_1 * e_2 = e_1 \quad \text{because } e_2 \text{ is an identity}$$
$$e_1 * e_2 = e_2 \quad \text{because } e_1 \text{ is an identity}$$

So $e_1 = e_2$, which is a contradiction.

A similar proof for zero elements can be given (exercise).

(4): Suppose for a contradiction that there is an element a in a group $(S,*)$ having two distinct inverses a' and a''. Then,

$$a' = a' * e = a' * (a * a'') = (a' * a) * a'' = e * a'' = a''$$

- $a' = a' * e$: e is the identity
- $a' * e = a' * (a * a'')$: $e = a * a''$ because a'' is an inverse of a
- $(a' * a) * a'' = e * a''$: associativity; $a' * a = e$ because a' is an inverse of a
- $e * a'' = a''$: e is the identity

So $a' = a''$, a contradiction. □

More general than part (4) of the previous result we have

Definition/Remark: Let $(S,*)$ be a monoid and 1 be the identity of it. (Let $a \in S$. We say that a has an inverse (or a is invertible) if there is a $b \in S$ such that $b * a = 1 = a * b$.

Any such b is called an (actually the!) inverse of a)

(1) Any invertible element in $(S,*)$ has a unique inverse.

(2) The set of all invertible elements of S is a group wrt $*$. (That is invertible elements of a monoid form a group).

Proof (1): Imitate the proof of part (4) of the previous result. (Exercise)

(2): Let $G=\{a\in S \mid \text{there is a } b\in S \text{ such that } a*b=1=b*a\}$ be the set of all invertible elements of the monoid $(S,*)$. We want to show that G is nonempty and $(G,*)$ satisfies the axioms (0),(1),(2),(3) of the definition.

- As $1*1=1$, $1\in G$. So $G\neq\emptyset$.
- Suppose that $x,y\in G$. There are $u,v\in S$ such that
$$x*u=1=u*x \quad \text{and} \quad y*v=1=v*y.$$

Now, $v*u\in S$ and $(x*y)*(v*u)=((x*y)*v)*u$

(associativity)

$$=(x*(y*v))*u$$
$$=(x*1)*u$$
$$=x*u$$
$$=1$$

Similarly $(v*u)*(x*y)=1$. Hence, $x*y\in G$. So $(G,*)$ satisfies the axiom (0).

- Elements of S satisfy the associativity axiom (1). As $G\subseteq S$,

elements of G satisfy (1) too.

— As $1 \in G$ and $1 * a = a = a * 1$ for all $a \in G$ (because 1 is the identity of $(S,*)$), $(G,*)$ satisfies the axiom (2).

— Let $a \in G$. By the definition of G there is a $b \in S$ such that $a * b = 1 = b * a$. Note that this implies $b \in G$. So $(G,*)$ satisfies the axiom (3). □

Remark: In a semigroup $(S,*)$ it follows from the associativity axiom that we may define the operation of finitely many elements without putting any paranthesis (because the associativity axiom says that without changing the order of elements we may put paranthesis any where we want). For instance, as

$$((a_1 * a_2) * a_3) * a_4) * a_5 = a_1 * ((a_2 * a_3) * (a_4 * a_5)) = \cdots$$

we may simply write $a_1 * a_2 * a_3 * a_4 * a_5$ without any paranthesis.

Ex: (1) Consider $(\mathbb{N}, +)$ the usual addition

It is a monoid, its identity is 0, it has no zero element
It is commutative. The only invertible element is 0.
It is not a group.

(2) Consider $(\mathbb{N}, \cdot)$ the usual multiplication.

It is a monoid, its identity is 1. It has the zero element, 0.
It is commutative. The only invertible element is 1. It is not a group

(3) $\{n \in \mathbb{N} \mid n \geq 7\}$ is a semigroup / not a monoid wrt usual $+$ & $\times$

(4) Let X be any nonempty set and $\mathcal{F}_X = \{f \mid f: X \to X \text{ is a function}\}$ be the set of all functions from X to X. Then $\mathcal{F}_X$ is a monoid wrt the function composition. Its identity element is the identity function 1_X on X. An element $f \in \mathcal{F}_X$ is invertible iff f is bijective. So the set $S_X = \{f \mid f: X \to X \text{ is a bijection}\}$ is a group wrt the function composition. S_X is called the symmetric group on X. Note that if $|X| = n$ then $|\mathcal{F}_X| = n^n$ and $|S_X| = n!$.

(5) $(\mathbb{R}, +)$ is a commutative group.
usual addition

$(\mathbb{R}, \cdot)$ is a monoid with the identity 1. It has the zero element 0.
usual multip.
All of its elements except 0 are invertible. So $(\mathbb{R} - \{0\}, \cdot)$ is a group.

(6) Let $n \in \mathbb{N}^+$. Consider the integers modulo n and the addition and multiplication modulo n on it.

$$\mathbb{Z}_n = \{[k]_n \mid k \in \mathbb{Z}\} = \{[0]_n, [1]_n, [2]_n, \dots, [n-1]_n\},$$

$$[k] = \{k + nt \mid t \in \mathbb{Z}\}, \quad [k]_n = [l]_n \Leftrightarrow k \equiv l \bmod n,$$

$$[a]_n +_n [b]_n = [a+b]_n, \quad [a]_n \cdot_n [b]_n = [a \cdot b]_n$$

$(\mathbb{Z}_n, +_n)$ is a group, its identity is $[0]_n$, the inverse of $[a]_n$ is $[-a]_n$.

$(\mathbb{Z}_n, \cdot_n)$ is a monoid with identity $[1]_n$, it is not a group if $n > 1$

Indeed, for any element $[a]_n \in (\mathbb{Z}_n, \cdot_n)$,

$[a]_n$ is invertible iff $[a]_n \cdot_n [b]_n = [1]_n$ for some $[b]_n \in \mathbb{Z}_n$

iff $[ab]_n = [1]_n$

iff $ab \equiv 1 \bmod n$

iff $ab = 1 + kn$ $\quad \exists\, b, k \in \mathbb{Z}$

iff $\gcd(a,n) = 1$.

So $\{[a]_n \in \mathbb{Z}_n \mid \gcd(a,n)=1\}$ is a group wrt $\cdot_n$, and it is sometimes called the <u>unit group</u> of $\mathbb{Z}_n$ and denoted by $U(\mathbb{Z}_n)$ or $\mathbb{Z}_n^\times$

Hence $U(\mathbb{Z}_n) = \{[a]_n \mid a \in \mathbb{Z} \text{ and } \gcd(a,n)=1\}$ is a group wrt $\cdot_n$.

For instance, $(U(\mathbb{Z}_9), \cdot_9) = \{[1]_9, [2]_9, [4]_9, [5]_9, [7]_9, [8]_9\}$. Its multiplication table is

	$[1]_9$	$[2]_9$	$[4]_9$	$[5]_9$	$[7]_9$	$[8]_9$
$[1]_9$						
$[2]_9$			$[8]_9$			$[7]_9$
$[4]_9$				$[2]_9$	$[1]_9$	
$[5]_9$		$[1]_9$				
$[7]_9$						
$[8]_9$						

Multiplication table of any finite semigroup $(S, *)$ is drawn similarly:

$*$		t
s		$s*t$

The order of rows and columns are the same. If there is an identity, list it first.

(7) Let $M_{m\times n}(\mathbb{R})$ be the set of all $m\times n$ matrices with real entries where m and n are positive integers.

$M_{m\times n}(\mathbb{R})$ is a group wrt the matrix addition. Its identity is the $m\times n$ zero matrix, and the inverse of any matrix A is $-A$.

Consider now $M_{m\times n}(\mathbb{R})$ with the matrix multiplication. To multiply matrices we must assume that $m=n$. Note that $M_{n\times n}(\mathbb{R})$ is a monoid wrt the matrix multiplication and its identity is the identity matrix I. Note that a matrix A has an inverse in the monoid $M_{n\times n}(\mathbb{R})$ iff A is invertible as a matrix. So $M_{n\times n}(\mathbb{R})$ is not a group wrt the matrix multiplication (because there are noninvertible matrices) but its invertible elements forms a group, which is called the general linear group of degree n on $\mathbb{R}$ and denoted by $GL_n(\mathbb{R})$. Thus,

$$GL_n(\mathbb{R}) = \{ A \in M_{n\times n}(\mathbb{R}) \mid \det(A) \neq 0 \}$$

It is possible to change the entries of matrices to obtain other monoids and groups. For instance, we have the group

$$GL_n(\mathbb{Z}_m) = \{ A \in M_{n\times n}(\mathbb{Z}_m) \mid \det(A) \in U(\mathbb{Z}_m) \}$$

---

(8) $(\mathbb{Z}, *)$ is a semigroup where $a * b = \max(a, b)$

$(\mathbb{N}, *)$ is a monoid where $a * b = \max(a, b)$ (Exercise)

---

(9) Let A and B be nonempty sets. Then $A \times B$ becomes a semigroup wrt $*$ where, for any $(a_1, b_1), (a_2, b_2) \in A \times B$,

$$(a_1, b_1) * (a_2, b_2) = (a_1, b_2)$$

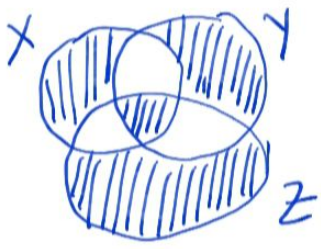
It has no identity unless $|A| = |B| = 1$.

(10) Let A be a nonempty set. Then

(i) $(\mathcal{P}(A), \cap)$ is a monoid with the identity A. The only invertible element is A.

(ii) $(\mathcal{P}(A), \cup)$ is a monoid with the identity ϕ. The only invertible element is ϕ.

(iii) $(\mathcal{P}(A), \Delta)$ (Δ: the symmetric difference) is a group with the identity ϕ. The inverse of any $B \in \mathcal{P}(A)$ is B itself. Proving the associativity of Δ on $\mathcal{P}(A)$ is tedious. With the help of Venn Diagrams, the shaded region



is equal to the both of $(X \Delta Y) \Delta Z$ and $X \Delta (Y \Delta Z)$.

(11) $(\mathbb{N}^+, *)$ with $a * b = a^b$ is not a semigroup because $*$ is not associative on $\mathbb{N}^+$. Indeed, $(a * b) * c = (a^b)^c = a^{bc}$ and $a * (b * c) = a^{(b^c)}$ which are not equal in general. For instance $(2 * 1) * 3 = 8$ but $2 * (1 * 3) = 2$.

(12) $(\mathbb{R} - \{0\}, *)$ with $a * b = 2ab$ is a group. Its identity is $1/2$ and the inverse of any $a \in \mathbb{R} - \{0\}$ is $1/4a$. (Exercise).