CHAPTER 3

Functions

3.1. Product of Sets

3.1.1 Definition Let A and B be sets. The *product* or *Cartesian product* of A and B is the set of all ordered pairs (a, b) where $a \in A$ and $b \in B$. This set is denoted $A \times B$:

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

If A has m elements and B has n elements then $A \times B$ has mn elements:

$$|A \times B| = |A| \cdot |B|.$$

3.1.2 Example

- ★ $\{1, 2, 3\} \times \{a, b, c\} = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c), (3, a), (3, b), (3, c)\}$.
- ★ $\mathbb{R} \times \mathbb{R}$ consists of all pairs (x, y) with $x \in \mathbb{R}$ and $y \in \mathbb{R}$. That is, $\mathbb{R} \times \mathbb{R}$ is the xy plane. This set is often denoted $\mathbb{R}^2$ in calculus.
- ★ $[0, 1] \times [0, 1]$. This is the unit square in $\mathbb{R}^2$.

Similarly we define the product $A_1 \times A_2 \times \cdots A_n$ of n sets to be the set of all n-tuples $(a_1, a_2, \ldots, a_n)$ with each $a_i \in A_i$.

3.2. Definition of a Function

One of the most fundamental concepts in mathematics is that of a *function.* Intuitively a function f is a *rule* (machine, recipe) that gives for each input a single output.

Often we encounter functions given by a formula, such as

$$f(x) = x^2 - 1.$$

Here the rule is given by describing what it does to an arbitrary input x (square and subtract one in this case).

If we collect all the possible inputs x of our function, and for each x list the corresponding output $f(x)$ then we can write down a list of pairs $(x, f(x))$ that completely describe the function. (If we plot all these pairs we then get the graph of f.) The defining property of a function is that for each x we have a unique $f(x)$. (Graphically this is the “vertical line test”.)

The set of possible inputs for the function is called the *domain.* The outputs may belong to a different set called the *codomain.* A function f with domain A and codomain B is denoted $f \colon A \to B$.

Here is the formal definition:

3.2.1 Definition Let A and B be sets. A *function* $f \colon A \to B$ is a subset of $A \times B$ satisfying:

(**) For every $a \in A$ there exists a unique pair $(a, b) \in f$.

We write $f \colon A \to B$ and call A the *domain* and B the *codomain* of f. A function is also called a *map* .

Strictly speaking, a function is thus a set of ordered pairs (a, b) with $a \in A$ and $b \in B$. Usually however we think of it as a rule. Instead of writing $(a, b) \in f$ we write $f(a) = b$, and think of f as sending a to b. This makes sense because of (**) above: for each a there is exactly one pair $(a, b) \in f$, so b is uniquely defined once a is known.

3.2.2 Example

- ★ If A is any set, let $1_A \colon A \to A$ be the function consisting of all pairs (a, a) with $a \in A$. That is, $1_A(a) = a$ for every $a \in A$. This is called the *identity function* on A.
- ★ More generally, if A is a subset of C there is an *inclusion function* $i \colon A \to C$, defined by $i(a) = a$ for every $a \in A$.
- ★ The function $f \colon \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = 1/x$. Literally $f = \{\dots (2, 1/2), \dots, (3, 1/3), \dots, (\pi, 1/\pi), \dots\}$.
- ★ Let $M_n(\mathbb{R})$ be the set of all $n \times n$ real matrices. The determinant function $\det \colon M_n(\mathbb{R}) \to \mathbb{R}$.
- ★ Let $\mathscr{F}$ be the set of all functions $\mathbb{R} \to \mathbb{R}$, and let $\mathscr{D}$ be the subset of everywhere differentiable functions $\mathbb{R} \to \mathbb{R}$. Define a function $D \colon \mathscr{D} \to \mathscr{F}$ by $D(f) = f'$ the derivative of f. This function is usually called the differentiation operator.
- ★ If A and B are sets, there are two *projection maps* $\pi_1 \colon A \times B \to A$ defined[1] by $\pi_1(a, b) = a$ and $\pi_2 \colon A \times B \to B$ defined by $\pi_2(a, b) = b$.

In calculus most functions have domain and codomain some subset of $\mathbb{R}$. Often the domain is taken by convention to be the largest subset of $\mathbb{R}$ where the rule defining the function "makes sense". In algebra we need to be more careful. Thus to define a function f, we must specify the domain, the codomain and describe what f does to every element of its domain.

Bad habit: do not write "the function f(x)". The function is called f, while $f(x)$ means the value of the function f at the argument x.

3.2.3 Example Recall that $\mathbb{Q}$ denotes the set of rational numbers (fractions). Define $f \colon \mathbb{Q} \to \mathbb{Q}$ by

$$f(a/b) = a.$$

That is, f is the "numerator function". *This function is not well defined.* The problem is that $1/2 = 2/4 = 3/6 = \cdots$ but $f(1/2) = 1$, $f(2/4) = 2$, $f(3/6) = 3$ etc. The same input has more than one output.

34 Exercise How should the numerator function be defined?

There is a general principle here: Suppose we define a function $f \colon A \to B$ by giving a rule $f(a) =$ [something]. If identical elements of A can have more than one name, we must ensure that our rule does not depend on the particular name chosen. That is, if $a = b$ (so a and b are different names for the same thing) we must check that $f(a) = f(b)$. This is called showing that the function is *well defined.*

3.2.4 Example Let $f \colon \mathbb{Q} \to \mathbb{Q}$ be defined by $f(x) = x + 1$. This is well defined. Eg $f(1/2) = 1.5$, $f(2/4) = 1.5$ etc and we always get the same output for the same input.

Proof that f is well defined: if $x = x'$ then $x + 1 = x' + 1$, so $f(x) = f(x')$, so the particular representation of the input chosen (x or x') does not affect the output.

We shall see further examples later in the course.

Since functions are mathematical objects (in fact they are just sets of pairs) it makes sense to examine when two functions are equal.

[1] Strictly speaking the input to π_1 is an ordered pair (a, b), but we write $\pi_1(a, b)$ instead of $\pi_1((a, b))$.

3.2.5 Definition Two functions f and g are *equal* if they have the same domain, the same codomain and $f(x) = g(x)$ for every x in their common domain.

Thus to prove that two functions are equal, we must check that they have the same domain and codomain, and then evaluate each at an arbitrary element and check that the output is the same.

3.2.6 Example

- ★ Let $f\colon \mathbb{Z} \to \mathbb{Z}$ be given by $f(x) = x + 1$, and $g\colon \mathbb{R} \to \mathbb{R}$ also be given by $g(x) = x + 1$. Then f and g are not equal since they have different domains.
- ★ Let $F\colon \mathbb{R} \to \mathbb{R}$ and $G\colon \mathbb{R} \to \mathbb{R}$ be given by $F(x) = x^2 - 1$ and $G(x) = (x - 1)(x + 1)$. Then $F = G$, since the have the same domain and codomain and $F(x) = G(x)$ for every x in their domain.

3.2.7 Definition Let $f\colon A \to B$ be a function, and suppose $D \subseteq A$. The *restriction* of f to D denoted $f\mid_D$ is the function $f\mid_D\colon D \to B$ given by

$$f\mid_D(d) = f(d) \qquad \text{for every } d \in D.$$

That is, we $f\mid_D$ is the same as f, except that we only allow inputs from D. The functions f and $f\mid_D$ are not the same, because they do not have the same domain.

3.2.8 Definition Let $f\colon A \to B$ be a function. The *image of* f or *range of* f is the following subset of B:

$$f(A) = \{f(a) \mid a \in A\}.$$

This set is also denoted $\mathrm{Im} f$.

The range or image is always a subset of the codomain. (Recall that the codomain is the set given in the definition of the function that accepts all the outputs of the function.)

3.2.9 Example

- ★ The function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$. The range of f is $[0, \infty)$, the set of non-negative reals.
- ★ The range of the sine and cosine functions are $[-1, 1]$.

If $f\colon A \to B$ and $g\colon B \to C$ then we can compose f and g to form a new function $A \to C$.

3.2.10 Definition Suppose $f\colon A \to B$ and $g\colon B \to C$. The *composition* $g \circ f$ is the function $A \to C$ given by

$$(g \circ f)(x) = g\Big(f(x)\Big) \qquad \text{for all } x \in A$$

In the displayed equation we are defining a function $g \circ f$ (with a long name). To define it, we must specify what it does to each input. The rule is: input x goes to output $g(f(x))$. This makes sense because $f(x) \in B$ so g can act on $f(x)$.

Note that in $g \circ f$ it is f that is applied to the input first. The reason for the awkward “backwards” notation is that we write function application from left to right: $f(x)$ not $(x)f$. In retrospect, the latter may have been preferable …

3.2.11 Example Let $f\colon A \to B$ be a function, and suppose $D \subseteq A$. Let $i\colon D \to A$ be the inclusion function (example 3.2.2). Then

$$f \circ i = f\mid_D .$$

Proof Both $f \circ i$ and $f\mid_D$ map from $D \to B$. For any $d \in D$, $f\mid_D(d) = f(d)$ and $f \circ i(d) = f(i(d)) = f(d)$ also. So the two functions are equal. □

3.2.12 Example If $f\colon A \to B$ then $f \circ 1_A = f$ and $1_B \circ f = f$.

35 Exercise Prove this statement.

3.2.13 Theorem Suppose $h\colon A \to B$, $g\colon B \to C$, $f\colon C \to D$ are functions. Then

$$(f \circ g) \circ h = f \circ (g \circ h).$$

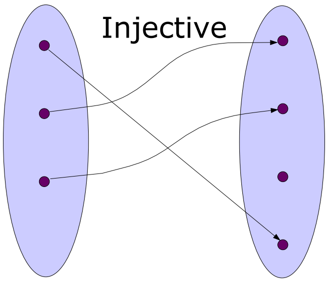
Proof The LHS and RHS both are functions from A to D. We must show they agree at all $x \in A$. But

$$\begin{aligned}
\Big((f \circ g) \circ h\Big)(x) &= \Big(f \circ g\Big)(h(x)) && \text{Definition of } ? \circ h \\
&= f\Big(g(h(x)\Big) && \text{Definition of } f \circ g \\
&= f\Big((g \circ h)(x)\Big) && \text{Definition of } g \circ h \\
&= \Big(f \circ (g \circ h)\Big)(x) && \text{Definition of } f \circ ? \quad \square
\end{aligned}$$

3.3. Injective, Surjective, Bijective

For each input a function must produce a single output. However different inputs may have the same output.

3.3.1 Definition A function $f\colon A \to B$ is *injective* or *one-to-one* if $f(a) = f(a')$ implies $a = a'$.



In other words,[2] f is injective if $a \neq a'$ implies $f(a) \neq f(a')$. Thus a function is injective if different inputs go to different outputs. (In calculus this is the "horizontal line test".)

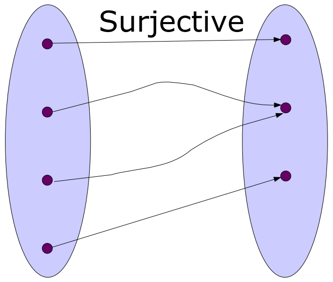
To prove a function is injective, we let a, $a' \in A$, assume $f(a) = f(a')$ and prove that $a = a'$. To prove that it is not injective, it is enough to find a single example with $a \neq a'$ but $f(a) = f(a')$.

3.3.2 Example

- ★ The function $f(x)\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^n$ is injective if n is odd, but not injective if n is even.
- ★ The sine function $\sin\colon \mathbb{R} \to \mathbb{R}$ is not injective. However $\sin|_{[-\pi/2,\pi/2]}$ is injective.
- ★ The differentiation function $D\colon \mathscr{D} \to \mathscr{F}$ (example 3.2.2) is not injective. For example let $f\colon \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$ and $f\colon \mathbb{R} \to \mathbb{R}$ be given by $g(x) = x^2 + 1$. Then $f, g \in \mathscr{D}$, $f \neq g$, but $D(f) = D(g)$, since $f' = g'$.
- ★ The determinant function $\det\colon M_n(\mathbb{R}) \to \mathbb{R}$ is not injective (provided $n \geq 2$).

3.3.3 Definition A function $f\colon A \to B$ is *surjective* or *onto* if $f(A) = B$.

[2] The contrapositive

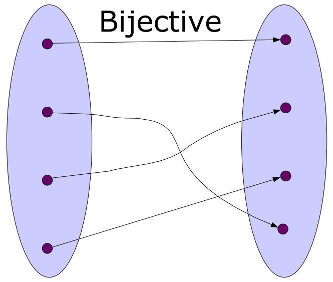


Thus a function is surjective if for every $b \in B$ there exists $a \in A$ with $f(a) = b$. To prove that $f\colon A \to B$ is surjective, let $b \in B$ and show that there is an element $a \in A$ with $f(a) = b$. To prove it is not surjective, find a single example of an element $b \in B$ not in the image of f.

3.3.4 Example

- ★ The function $f(x)\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^n$ is surjective if n is odd, but not surjective if n is even.
- ★ The projection maps $\pi_1\colon A \times B \to A$ and $\pi_2\colon A \times B \to B$ are both surjective. (Exercise)
- ★ The determinant function $\det\colon M_n(\mathbb{R}) \to \mathbb{R}$ is surjective. *Proof* Let $a \in \mathbb{R}$. We must find an $n \times n$ matrix with determinant a. This is easy: choose a diagonal matrix with a in the $(1,1)$ position, and 1's elsewhere on the diagonal.

3.3.5 Definition A *bijection* is a function that is injective and surjective.



A bijection $f\colon A \to B$ is a perfect matching between the elements of A and the elements of B. Each $a \in A$ is sent to a unique element of B (f is a function), different a are sent to different b (f is injective), and every b in B is hit in this way (f is surjective).

3.3.6 Theorem Suppose A is a finite set and $f\colon A \to B$ is a bijection. Then B is finite and A and B have the same number of elements.

Proof Let A have n elements, $\{a_1, \dots, a_n\}$. Since f is surjective, every element of B is of the form $f(a)$ for some $a \in A$. Thus every element of B must be in $\{f(a_1), \dots, f(a_n)\}$. Thus B has at most n elements. On the other hand, f is injective, so the elements $f(a_1), \dots, f(a_n)$, are all different. Thus the elements of B are precisely the n elements $\{f(a_1), \dots, f(a_n)\}$. □

Along the same lines:

3.3.7 Theorem Suppose A and B are finite sets, each with n elements. If $f\colon A \to B$ is injective or surjective then it is bijective.

Proof Suppose f is injective. Let $A = \{a_1, \dots, a_n\}$. Then the image of f is $\{f(a_1), \dots, f(a_n)\}$, and this consists of n distinct elements because f is injective. Therefore the image of f is all of B, so f is surjective. The case f is surjective is an exercise. □

36 Exercise Prove the other half of this theorem.

37 Exercise Let $f\colon A \to B$, $g\colon B \to C$ be functions. Prove or disprove the following:

(a) If f and g are injective so is $g \circ f$.
(b) If $g \circ f$ is injective, so is g.
(c) If $g \circ f$ is injective, so is f.
(d) If f and g are surjective so is $g \circ f$.
(e) If $g \circ f$ is surjective, so is g.
(f) If $g \circ f$ is surjective, so is f.

3.4. Inverse Functions

3.4.1 Definition Let $f\colon A \to B$ be a function. We say that f is *invertible* if there exists a function $g\colon B \to A$ with

$$f \circ g = 1_B \qquad \text{and} \qquad g \circ f = 1_A.$$

In this case g is called an *inverse* of f.

3.4.2 Theorem If f is invertible, it has a unique inverse.

Proof The proof is structurally identical to the result for inverses mod n, theorem 2.6.7.

Suppose g and $h\colon B \to A$ are inverses for f. Then

$$\begin{aligned} g &= g \circ 1_B && \text{Definition of } 1_B \text{ (example 3.2.2)} \\ &= g \circ (f \circ h) && h \text{ is an inverse of } f \\ &= (g \circ f) \circ h && \text{theorem 3.2.13} \\ &= 1_A \circ h && g \text{ is an inverse of } f \\ &= h && \text{Definition of } 1_A \text{ (example 3.2.2)} \quad \square \end{aligned}$$

3.4.3 Definition If f is invertible, we write f^{-1} to denote the unique inverse of f.

Note that f^{-1} is not the function $1/f$.

3.4.4 Example

★ Let $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^3$. Then f is invertible.

Proof Let $g\colon \mathbb{R} \to \mathbb{R}$ be the cube root function $g(x) = \sqrt[3]{x}$. Then $f \circ g(x) = (\sqrt[3]{x})^3 = x$, and $g \circ f(x) = \sqrt[3]{x^3} = x$. So $f \circ g$ and $g \circ f$ are both the identity function $1_{\mathbb{R}}$.

★ Let $f\colon \mathbb{R} \to \mathbb{R}^+$ be the exponential function. This is invertible, with inverse the (natural[3]) log function $f^{-1} = \log\colon \mathbb{R}^+ \to \mathbb{R}$.

Proof : $f \circ f^{-1}(x) = e^{\log(x)} = x$ and $f^{-1} \circ f(x) = \log(e^x) = x$. So $f \circ f^{-1} = 1_{\mathbb{R}\setminus\{0\}}$, $f^{-1} \circ f = 1_{\mathbb{R}}$.

Intuitively, the inverse of f is the rule that “undoes” whatever f does. If we cube x, then to undo this operation we need to take cube roots. If we think of f represented by arrows $a \mapsto b$ for each $a \in A$, then f^{-1} is obtained by reversing all the arrows: $b \mapsto a$. This process will not always produce a function, so not every function is invertible. Indeed, to get a function, for each $b \in B$ there must be exactly one incoming arrow from a unique $a \in A$. (If there is no arrow, then $f^{-1}(b)$ will not be defined. If there are two or more arrows, again $f^{-1}(b)$ will not be defined.) This exactly means that f must be bijective.

3.4.5 Theorem A function $f\colon A \to B$ is invertible iff it is bijective.

[3] Throughout these notes, log means natural log.

Proof $\Longrightarrow$ Suppose f is invertible, with inverse f^{-1}. We show that f is a bijection.

Surjective: Let $b \in B$. We find an $a \in A$ with $f(a) = b$. To do so, take a to be $f^{-1}(b)$. Then $f(a) = f(f^{-1}(b)) = b$. Since $b \in B$ was arbitrary, f is surjective.

Injective: Now suppose $f(a) = f(a')$. Applying f^{-1} to both sides, $f^{-1}(f(a)) = f^{-1}(f(a'))$ so $a = a'$. Thus f is injective.

$\Longleftarrow$ Suppose f is a bijection. For each $b \in B$ there exists $a \in A$ with $f(a) = b$ (because f is surjective) and this a is unique (because f is injective). Thus we can define a function $g\colon B \to A$ by $g(b) =$ the unique a with $f(a) = b$. Let $a \in A$ be arbitrary and suppose $f(a) = b$. Then $g(f(a)) = g(b) = a$ by definition of g, so $g \circ f = 1_A$. Now let $b' \in B$ be arbitrary, and let $g(b') = a'$. This means $f(a') = b'$. Thus $f(g(b')) = f(a') = b'$. So $f \circ g = 1_B$. □

Note: in calculus it is often said that a function $f\colon \mathbb{R} \to \mathbb{R}$ is invertible iff it passes the horizontal line test: that is, iff it is injective. Why is the surjective condition not stated? The reason is that in calculus it is common to consider f not as a function from $\mathbb{R} \to \mathbb{R}$, but as a function from $\mathbb{R}$ onto its image (range), making it automatically surjective.

For example, the exponential function is not invertible if it is considered as a function $\mathbb{R} \to \mathbb{R}$ (what should the inverse image of 0 be?), but is invertible if it considered as a function $\mathbb{R} \to \mathbb{R}^+$.

It is also common in calculus to restrict the domain of non-injective functions to make them into bijections. For example, $\sin\colon \mathbb{R} \to [-1, 1]$ is not injective, but

$$\sin|_{[-\frac{\pi}{2},\frac{\pi}{2}]}\colon [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$$

is a bijection. Its inverse is the arcsin function.

38 Exercise Let $\sigma\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ be defined as the swap function: $\sigma(x, y) = (y, x)$. Prove that σ is a bijection.

39 Exercise Define a function $f\colon \mathbb{Z} \to \mathbb{N}$ by:

$$f(n) = \begin{cases} 2n, & \text{if } n \geq 0 \\ -1 - 2n, & \text{if } n < 0 \end{cases}$$

Prove that f is a bijection.

40 Exercise Let A and B be sets and let $g\colon A \to B$ be a function. A function $f\colon B \to A$ is a *left inverse* for g if $f \circ g = 1_A$. A function $h\colon B \to A$ is a *right inverse* for g if $g \circ h = 1_B$.

(a) Show that g has a left inverse iff it is injective.
(b) Show that g has a right inverse iff it is surjective.
(c) Hence show that g has an inverse iff it is bijective. (Hint: the $\Longrightarrow$ direction is easy, but in the $\Longleftarrow$ direction, you must show $f = h$.)

41 Exercise Let A, B and C be sets and let $g, h\colon A \to B$ and $f\colon B \to C$ be functions.

(a) Show that f is injective iff whenever $f \circ g = f \circ h$ we must have $g = h$.
(b) Formulate and prove a corresponding statement about surjective functions.

3.5. Proof That φ is Multiplicative

We can now prove theorem 2.7.4, that $\varphi(mn) = \varphi(m)\varphi(n)$ provided that $\gcd(m, n) = 1$. First note the following simple result.

3.5.1 Theorem $\gcd(a, mn) = 1$ iff $\gcd(a, m) = 1$ and $\gcd(a, n) = 1$.

Proof $\implies$ If $d \mid a$ and $d \mid m$ then $d \mid a$ and $d \mid mn$ so $d \mid \gcd(a, mn) = 1$ (by ($2'$) in theorem 1.7.7) hence $d = \pm 1$, so $\gcd(a, m) = 1$. Similarly $\gcd(a, n) = 1$.

$\impliedby$ (This is essentially theorem 1.8.4). Let p be any prime. If $p \mid a$ and $p \mid mn$ then by theorem 1.7.2, [$p \mid a$ and $p \mid m$] or [$p \mid a$ and $p \mid n$], but then $\gcd(a, m)$ or $\gcd(a, n) > 1$, contradiction. So no prime divides a and mn, so $\gcd(a, mn) = 1$. □

3.5.2 Theorem If $\gcd(m, n) = 1$ then $\varphi(mn) = \varphi(m)\varphi(n)$.

Proof Let $(\mathbb{Z}/n\mathbb{Z})^\times$ be the set of invertible elements mod x. We show that there is an invertible function F

$$(\mathbb{Z}/mn\mathbb{Z})^\times \to (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times.$$

An invertible function is a bijection (theorem 3.4.5), and if there is a bijection between two finite sets then they have the same number of elements (theorem 3.3.6). The cardinality of the LHS is $\varphi(mn)$, and the cardinality of the RHS is $\varphi(m)\varphi(n)$, so the existence of F will complete the proof.

The map F is defined as follows: given $a \in (\mathbb{Z}/mn\mathbb{Z})^\times$ we get a pair $F(a) = (b, c) \in (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$ by reducing a mod m and mod n so $a \equiv b \pmod{m}$ and $a \equiv c \pmod{n}$. The previous theorem implies that b and c are invertible, so this definition makes sense.

Now, given $(b, c) \in (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$ we may find an a with $a \equiv b \pmod{m}$ and $a \equiv c \pmod{n}$ by the CRT (theorem 2.8.3), and this a is unique mod mn. The previous theorem implies a is invertible mod mn, so $a \in (\mathbb{Z}/mn\mathbb{Z})^\times$. Thus we have a function $G\colon (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times \to (\mathbb{Z}/mn\mathbb{Z})^\times$. The maps $F\colon a \mapsto (b, c)$ and $G\colon (b, c) \mapsto a$ are inverses of each other, so F is invertible. □

3.6. Binary Operators

Very often in mathematics we study properties of operations such as addition (of numbers, integers mod n, vectors, matrices, functions ...), multiplication (of numbers, matrices ...), composition etc. These operations are all functions. For example, addition of real numbers, $+$, is a function $+\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ taking a pair of real numbers, and adding them together. We introduce a formalism to deal with all operations of this form.

3.6.1 Definition Let A be a set. A *binary operation* on A is a function $f\colon A \times A \to A$.

Thus a binary operation takes as input two elements of A (strictly speaking, an ordered pair $(a_1, a_2) \in A \times A$) and produces an output, again in A. Such functions are extremely common.

3.6.2 Example

- ★ Let $A = \mathbb{R}$. Let $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be given by $f(a_1, a_2) = a_1 + a_2$. That is, f is the function usually denoted $+$. Then f (or $+$) is a binary operation on $\mathbb{R}$.
- ★ Subtraction and multiplication are also binary operations on $\mathbb{R}$. Division is not, because it is not defined on all pairs in $\mathbb{R} \times \mathbb{R}$ since we cannot divide by 0.
- ★ Addition and multiplication mod n are binary operations on $\mathbb{Z}/n\mathbb{Z}$.
- ★ Let $M_n(\mathbb{R})$ be the set of $n \times n$ matrices with real entries. Then addition of matrices, subtraction of matrices, and multiplication of matrices are all binary operations on $M_n(\mathbb{R})$.
- ★ Let $\mathscr{F}$ be the set of all functions $\mathbb{R} \to \mathbb{R}$. Then composition is a binary operation on $\mathscr{F}$. Here $\circ\colon \mathscr{F} \times \mathscr{F} \to \mathscr{F}$ is given by $\circ(g, h) = g \circ h$.

42 Exercise Let V be a real vector space. Is scalar multiplication a binary operation on V? Is vector addition? Explain.

Binary operations are usually denoted by symbols such as $+$, $\times$, $\circ$, $\oplus$, $\otimes$, $\cdot$ and written *between* their two arguments.[4] That is, we write $x+y$ instead of $+(x,y)$, $f\circ g$ instead of $\circ(f,g)$ etc.

If A is a (small) finite set and $*$ is a binary operation on A it is possible to describe $*$ using a table, called a Cayley table. Let $A=\{a_1,\ldots,a_n\}$. In position (i,j) we write entry $a_i * a_j$. We have already seen examples like this, such as the table for $\oplus$ on $\mathbb{Z}/3\mathbb{Z}$:

$\oplus$	$[0]$	$[1]$	$[2]$
$[0]$	$[0]$	$[1]$	$[2]$
$[1]$	$[1]$	$[2]$	$[0]$
$[2]$	$[2]$	$[0]$	$[1]$

Most binary operations encountered in mathematics satisfy additional constraints. The most common are as follows.

3.6.3 Definition A binary operation $*$ on a set A is *associative* if for every $a,b,c\in A$

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

It is *commutative* if for every a, $b\in A$

$$a*b=b*a.$$

3.6.4 Example

- ★ $+$ and $\cdot$ (multiplication) of real numbers are associative and commutative. This is why we can write an expression like $a_1+\cdots+a_n$ without worrying about where the parentheses go. We can also change the order the a_i occur in.
- ★ Sum and product of matrices is associative. Sum is commutative, but product is not.
- ★ Sum and product of congruence classes mod n is associative and commutative (theorem 2.2.8).
- ★ Composition of functions is associative:

  $$f\circ(g\circ h)=(f\circ g)\circ h.$$

  by theorem 3.2.13. It is not commutative. For example, if $f,g\colon\mathbb{R}\to\mathbb{R}$ are given by $f(x)=x+1$, $g(x)=2x$ then $f\circ g(x)=f(g(x))=f(2x)=2x+1$, but $g\circ f(x)=g(f(x))=g(x+1)=2(x+1)$.
- ★ Subtraction of real numbers is not associative or commutative: $a-(b-c)=a-b+c\neq(a-b)-c$. [To simplify notation we adopt the *convention* that $a-b-c$ is to be interpreted as $(a-b)-c$, but this is only a useful convention.]
- ★ Exponentiation of positive real numbers is not associative: $(a^b)^c=a^{bc}\neq a^{b^c}$. To clarify, temporarily write $a\uparrow b$ for a^b. Then $(a\uparrow b)\uparrow c=(a^b)^c=a^{bc}=a\uparrow(bc)$ but $a\uparrow(b\uparrow c)=a\uparrow(b^c)$.

  [Again we adopt a convention, and decree that $a^{b^c}=a\uparrow b\uparrow c$ means $a^{b^c}=a\uparrow(b\uparrow c)$ and not $(a^b)^c$, that is, $(a\uparrow b)\uparrow c$. Note that this is the opposite convention to that used for subtraction.]
- ★ Vector cross product in $\mathbb{R}^3$ is not associative or commutative.

Almost all important binary operations are associative. We shall study associative binary operators in the next chapter. Many important operations are not commutative (matrix multiplication, composition of functions).

Associativity is only defined in terms of products of length 3. What happens for longer products?

The expression $a_1*a_2*a_3*a_4$ has five possible interpretations: $A_1=(a_1*a_2)*(a_3*a_4)$, $A_2=((a_1*a_2)*a_3)*a_4$, $A_3=(a_1*(a_2*a_3))*a_4$, $A_4=a_1*((a_2*a_3)*a_4)$, $A_5=a_1*(a_2*(a_3*a_4))$. We

[4] This is sometimes called *infix* notation.

show these are all equal, using associativity for products of length 3 at each step:

$$\begin{aligned} A_1 &= (a_1 * a_2) * (a_3 * a_4) = A_1 \\ &= ((a_1 * a_2) * a_3) * a_4 = A_2 \\ &= (a_1 * (a_2 * a_3)) * a_4 = A_3 \\ &= a_1 * ((a_2 * a_3) * a_4) = A_4 \\ &= a_1 * (a_2 * (a_3 * a_4)) = A_5 \end{aligned}$$

3.6.5 Theorem If $*$ is associative, then any expression of the form $a_1 * a_2 * \cdots * a_n$ is uniquely defined, no matter how parentheses are inserted.

Proof The proof is by (strong) induction. We assume the result for all products of length $m < n$, and show that the result then also holds for products of length n.

Base cases: if $n \leq 2$ there is nothing to prove, since there are no different ways of associating such products. If $n = 3$ there are two possible ways of inserting parentheses, and the associative law exactly says these give the same result.

Assume the result holds for all products of length $m < n$. Let P be a product $a_1 * a_2 * \cdots * a_n$ where parentheses have been inserted in any legal manner. We shall show $P = (\cdots((a_1{*}a_2){*}a_3){*}\cdots a_{n-1}){*}a_n$, where the products are grouped (associated) to the left. Thus all parenthesizations of length n are equal.

Now, no matter how the parentheses are inserted in $a_1 * a_2 * \cdots * a_n$, there must be an outermost $*$ operation, the one that is applied last. That is, we can write $P = A * B$ where $A = a_1 * \cdots * a_m$ and $B = a_{m+1} * \cdots * a_n$, both parenthesized in some unknown way, with $0 < m < n$.

By the inductive hypothesis, $A = (\cdots((a_1 * a_2) * a_3) * \cdots) * a_m$ and $B = (\cdots((a_{m+1} * a_{m+2}) * a_{m+3}) * \cdots) * a_n$.

If $m = n - 1$ then $P = A * a_n = (\ldots((a_1 * a_2) * a_3) * \ldots) * a_{n-1}) * a_n$ and we are done. Otherwise, Let $C = (\ldots((a_{m+1} * a_{m+2}) * a_{m+3}) * \ldots) * a_{n-1}$, so $B = C * a_n$. Thus

$$P = A * B = A * (B * a_n) \overset{\ddagger}{=} (A * B) * a_n,$$

where ‡ indicates the use of associativity for 3 arguments.

Since $A * B$ has length $< n$ so it is equal to the product with all associations to the left: $A * B = (\ldots((a_1 * a_2) * a_3) * \ldots) * a_{n-1}$, and hence $P = (\ldots((a_1 * a_2) * a_3) * \ldots a_{n-1}) * a_n$ as claimed. □

3.6.6 Theorem Let $*$ be a commutative and associative binary operation on a set A, and suppose a_1, a_2, $\ldots$, $a_n \in A$. Then all products involving each of the a_i exactly once are equal, no matter in which order the a_i occur.

Proof Similar to theorem 3.6.5. Details omitted. □

43 Exercise How many binary operations are there on a set with n elements?

44 Exercise Let $S = \{a, b, c, d\}$. Suppose $*$ is an associative binary operation on S. Complete the table for $*$ below:

$*$	a	b	c	d
a	a	b	c	d
b	b	a	c	d
c	c	d	c	d
d	?	?	?	?

45 Exercise Let X be a set and let $\mathscr{P}(X)$ be the set of subsets of X. If $A, B \in \mathscr{P}(X)$ the *symmetric difference* of A and B, denoted $A \Delta B$ is $(A \setminus B) \cup (B \setminus A)$. That is, $A \Delta B$ is the set of elements in A or B but not both.

Is Δ an associative binary operator on $\mathscr{P}(X)$?