1.2. Functions

A function $f : X \to Y$ between sets X, Y assigns to each $x \in X$ a unique element $f(x) \in Y$. Functions are also called maps, mappings, or transformations. The set X on which f is defined is called the domain of f and the set Y in which it takes its values is called the codomain. We write $f : x \mapsto f(x)$ to indicate that f is the function that maps x to $f(x)$.

Example 1.9. The identity function $\mathrm{id}_X : X \to X$ on a set X is the function $\mathrm{id}_X : x \mapsto x$ that maps every element to itself.

Example 1.10. Let $A \subset X$. The characteristic (or indicator) function of A,

$$\chi_A : X \to \{0, 1\},$$

is defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Specifying the function χ_A is equivalent to specifying the subset A.

Example 1.11. Let A, B be the sets in Example 1.4. We can define a function $f : A \to B$ by

$$f(2) = 7, \quad f(3) = 1, \quad f(5) = 11, f(7) = 3, \quad f(11) = 9,$$

and a function $g : B \to A$ by

$$g(1) = 3, \quad g(3) = 7, \quad g(5) = 2, g(7) = 2, \quad g(9) = 5, \quad g(11) = 11.$$

Example 1.12. The square function $f : \mathbb{N} \to \mathbb{N}$ is defined by

$$f(n) = n^2,$$

which we also write as $f : n \mapsto n^2$. The equation $g(n) = \sqrt{n}$, where $\sqrt{n}$ is the positive square root, defines a function $g : \mathbb{N} \to \mathbb{R}$, but $h(n) = \pm\sqrt{n}$ does not define a function since it doesn't specify a unique value for $h(n)$. Sometimes we use a convenient oxymoron and refer to h as a multi-valued function.

One way to specify a function is to explicitly list its values, as in Example 1.11. Another way is to give a definite rule, as in Example 1.12. If X is infinite and f is not given by a definite rule, then neither of these methods can be used to specify the function. Nevertheless, we suppose that a general function $f : X \to Y$ may be "defined" by picking for each $x \in X$ a corresponding value $f(x) \in Y$.

If $f : X \to Y$ and $U \subset X$, then we denote the restriction of f to U by $f|_U : U \to Y$, where $f|_U(x) = f(x)$ for $x \in U$.

In defining a function $f : X \to Y$, it is crucial to specify the domain X of elements on which it is defined. There is more ambiguity about the choice of codomain, however, since we can extend the codomain to any set $Z \supset Y$ and define a function $g : X \to Z$ by $g(x) = f(x)$. Strictly speaking, even though f and g have exactly the same values, they are different functions since they have different codomains. Usually, however, we will ignore this distinction and regard f and g as being the same function.

The graph of a function $f : X \to Y$ is the subset G_f of $X \times Y$ defined by

$$G_f = \{(x, y) \in X \times Y : x \in X \text{ and } y = f(x)\}.$$

For example, if $f : \mathbb{R} \to \mathbb{R}$, then the graph of f is the usual set of points (x, y) with $y = f(x)$ in the Cartesian plane $\mathbb{R}^2$. Since a function is defined at every point in its domain, there is some point $(x, y) \in G_f$ for every $x \in X$, and since the value of a function is uniquely defined, there is exactly one such point. In other words, for each $x \in X$ the "vertical line" $L_x = \{(x, y) \in X \times Y : y \in Y\}$ through x intersects the graph of a function $f : X \to Y$ in exactly one point: $L_x \cap G_f = (x, f(x))$.

Definition 1.13. The range, or image, of a function $f : X \to Y$ is the set of values

$$\operatorname{ran} f = \{y \in Y : y = f(x) \text{ for some } x \in X\}.$$

A function is onto if its range is all of Y; that is, if

$$\text{for every } y \in Y \text{ there exists } x \in X \text{ such that } y = f(x).$$

A function is one-to-one if it maps distinct elements of X to distinct elements of Y; that is, if

$$x_1, x_2 \in X \text{ and } x_1 \neq x_2 \text{ implies that } f(x_1) \neq f(x_2).$$

An onto function is also called a surjection, a one-to-one function an injection, and a one-to-one, onto function a bijection.

Example 1.14. The function $f : A \to B$ defined in Example 1.11 is one-to-one but not onto, since $5 \notin \operatorname{ran} f$, while the function $g : B \to A$ is onto but not one-to-one, since $g(5) = g(7)$.

1.3. Composition and inverses of functions

The successive application of mappings leads to the notion of the composition of functions.

Definition 1.15. The composition of functions $f : X \to Y$ and $g : Y \to Z$ is the function $g \circ f : X \to Z$ defined by

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

The order of application of the functions in a composition is crucial and is read from from right to left. The composition $g \circ f$ can only be defined if the domain of g includes the range of f, and the existence of $g \circ f$ does not imply that $f \circ g$ even makes sense.

Example 1.16. Let X be the set of students in a class and $f : X \to \mathbb{N}$ the function that maps a student to her age. Let $g : \mathbb{N} \to \mathbb{N}$ be the function that adds up the digits in a number e.g., $g(1729) = 19$. If $x \in X$ is 23 years old, then $(g \circ f)(x) = 5$, but $(f \circ g)(x)$ makes no sense, since students in the class are not natural numbers.

Even if both $g \circ f$ and $f \circ g$ are defined, they are, in general, different functions.

Example 1.17. If $f : A \to B$ and $g : B \to A$ are the functions in Example 1.11, then $g \circ f : A \to A$ is given by

$$(g \circ f)(2) = 2, \quad (g \circ f)(3) = 3, \quad (g \circ f)(5) = 11,$$
$$(g \circ f)(7) = 7, \quad (g \circ f)(11) = 5.$$

and $f \circ g : B \to B$ is given by

$$(f \circ g)(1) = 1, \quad (f \circ g)(3) = 3, \quad (f \circ g)(5) = 7,$$
$$(f \circ g)(7) = 7, \quad (f \circ g)(9) = 11, \quad (f \circ g)(11) = 9.$$

A one-to-one, onto function $f : X \to Y$ has an inverse $f^{-1} : Y \to X$ defined by

$$f^{-1}(y) = x \text{ if and only if } f(x) = y.$$

Equivalently, $f^{-1} \circ f = \mathrm{id}_X$ and $f \circ f^{-1} = \mathrm{id}_Y$. A value $f^{-1}(y)$ is defined for every $y \in Y$ since f is onto, and it is unique since f is one-to-one. If $f : X \to Y$ is one-to-one but not onto, then one can still define an inverse function $f^{-1} : \operatorname{ran} f \to X$ whose domain in the range of f.

The use of the notation f^{-1} to denote the inverse function should not be confused with its use to denote the reciprocal function; it should be clear from the context which meaning is intended.

Example 1.18. If $f : \mathbb{R} \to \mathbb{R}$ is the function $f(x) = x^3$, which is one-to-one and onto, then the inverse function $f^{-1} : \mathbb{R} \to \mathbb{R}$ is given by

$$f^{-1}(x) = x^{1/3}.$$

On the other hand, the reciprocal function $g = 1/f$ is given by

$$g(x) = \frac{1}{x^3}, \qquad g : \mathbb{R} \setminus \{0\} \to \mathbb{R}.$$

The reciprocal function is not defined at $x = 0$ where $f(x) = 0$.

If $f : X \to Y$ and $A \subset X$, then we let

$$f(A) = \{y \in Y : y = f(x) \text{ for some } x \in A\}$$

denote the set of values of f on points in A. Similarly, if $B \subset Y$, we let

$$f^{-1}(B) = \{x \in X : f(x) \in B\}$$

denote the set of points in X whose values belong to B. Note that $f^{-1}(B)$ makes sense as a set even if the inverse function $f^{-1} : Y \to X$ does not exist.

Example 1.19. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. If $A = (-2, 2)$, then $f(A) = [0, 4)$. If $B = (0, 4)$, then

$$f^{-1}(B) = (-2, 0) \cup (0, 2).$$

If $C = (-4, 0)$, then $f^{-1}(C) = \varnothing$.

Finally, we introduce operations on a set.

Definition 1.20. A binary operation on a set X is a function $f : X \times X \to X$.

We think of f as "combining" two elements of X to give another element of X. One can also consider higher-order operations, such as ternary operations $f : X \times X \times X \to X$, but will will only use binary operations.

Example 1.21. Addition $a : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and multiplication $m : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ are binary operations on $\mathbb{N}$ where

$$a(x, y) = x + y, \qquad m(x, y) = xy.$$

1.4. Indexed sets

We say that a set X is indexed by a set I, or X is an indexed set, if there is an onto function $f : I \to X$. We then write

$$X = \{x_i : i \in I\}$$

where $x_i = f(i)$. For example,

$$\{1, 4, 9, 16, \dots\} = \left\{n^2 : n \in \mathbb{N}\right\}.$$

The set X itself is the range of the indexing function f, and it doesn't depend on how we index it. If f isn't one-to-one, then some elements are repeated, but this doesn't affect the definition of the set X. For example,

$$\{-1, 1\} = \{(-1)^n : n \in \mathbb{N}\} = \left\{(-1)^{n+1} : n \in \mathbb{N}\right\}.$$

If $\mathcal{C} = \{X_i : i \in I\}$ is an indexed collection of sets X_i, then we denote the union and intersection of the sets in $\mathcal{C}$ by

$$\bigcup_{i\in I} X_i = \{x : x \in X_i \text{ for some } i \in I\}, \qquad \bigcap_{i\in I} X_i = \{x : x \in X_i \text{ for every } i \in I\},$$

or similar notation.

Example 1.22. For $n \in \mathbb{N}$, define the intervals

$$\begin{aligned} A_n &= [1/n, 1 - 1/n] = \{x \in \mathbb{R} : 1/n \leq x \leq 1 - 1/n\}, \\ B_n &= (-1/n, 1/n) = \{x \in \mathbb{R} : -1/n < x < 1/n\}). \end{aligned}$$

Then

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n=1}^{\infty} A_n = (0, 1), \qquad \bigcap_{n\in\mathbb{N}} B_n = \bigcap_{n=1}^{\infty} B_n = \{0\}.$$

The general statement of De Morgan's laws for a collection of sets is as follows.

Proposition 1.23 (De Morgan)**.** If $\{X_i \subset X : i \in I\}$ is a collection of subsets of a set X, then

$$\left(\bigcup_{i\in I} X_i\right)^c = \bigcap_{i\in I} X_i^c, \qquad \left(\bigcap_{i\in I} X_i\right)^c = \bigcup_{i\in I} X_i^c.$$

Proof. We have $x \notin \bigcup_{i\in I} X_i$ if and only if $x \notin X_i$ for every $i \in I$, which holds if and only if $x \in \bigcap_{i\in I} X_i^c$. Similarly, $x \notin \bigcap_{i\in I} X_i$ if and only if $x \notin X_i$ for some $i \in I$, which holds if and only if $x \in \bigcup_{i\in I} X_i^c$. □

The following theorem summarizes how unions and intersections map under functions.

Theorem 1.24. Let $f : X \to Y$ be a function. If $\{Y_j \subset Y : j \in J\}$ is a collection of subsets of Y, then

$$f^{-1}\left(\bigcup_{j\in J} Y_j\right) = \bigcup_{j\in J} f^{-1}(Y_j), \qquad f^{-1}\left(\bigcap_{j\in J} Y_j\right) = \bigcap_{j\in J} f^{-1}(Y_j);$$

and if $\{X_i \subset X : i \in I\}$ is a collection of subsets of X, then

$$f\left(\bigcup_{i\in I} X_i\right) = \bigcup_{i\in I} f(X_i), \qquad f\left(\bigcap_{i\in I} X_i\right) \subset \bigcap_{i\in I} f(X_i).$$

Proof. We prove only the results for the inverse image of a union and the image of an intersection; the proof of the remaining two results is similar.

If $x \in f^{-1}\left(\bigcup_{j\in J} Y_j\right)$, then there exists $y \in \bigcup_{j\in J} Y_j$ such that $f(x) = y$. Then $y \in Y_j$ for some $j \in J$ and $x \in f^{-1}(Y_j)$, so $x \in \bigcup_{j\in J} f^{-1}(Y_j)$. It follows that

$$f^{-1}\left(\bigcup_{j\in J} Y_j\right) \subset \bigcup_{j\in J} f^{-1}(Y_j).$$

Conversely, if $x \in \bigcup_{j\in J} f^{-1}(Y_j)$, then $x \in f^{-1}(Y_j)$ for some $j \in J$, so $f(x) \in Y_j$ and $f(x) \in \bigcup_{j\in J} Y_j$, meaning that $x \in f^{-1}\left(\bigcup_{j\in J} Y_j\right)$. It follows that

$$\bigcup_{j\in J} f^{-1}(Y_j) \subset f^{-1}\left(\bigcup_{j\in J} Y_j\right),$$

which proves that the sets are equal.

If $y \in f\left(\bigcap_{i\in I} X_i\right)$, then there exists $x \in \bigcap_{i\in I} X_i$ such that $f(x) = y$. Then $x \in X_i$ and $y \in f(X_i)$ for every $i \in I$, meaning that $y \in \bigcap_{i\in I} f(X_i)$. It follows that

$$f\left(\bigcap_{i\in I} X_i\right) \subset \bigcap_{i\in I} f(X_i).$$

□

The only case in which we don't always have equality is for the image of an intersection, and we may get strict inclusion here if f is not one-to-one.

Example 1.25. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$. Let $A = (-1, 0)$ and $B = (0, 1)$. Then $A \cap B = \varnothing$ and $f(A \cap B) = \varnothing$, but $f(A) = f(B) = (0, 1)$, so $f(A) \cap f(B) = (0, 1) \neq f(A \cap B)$.

Next, we generalize the Cartesian product of finitely many sets to the product of possibly infinitely many sets.

Definition 1.26. Let $\mathcal{C} = \{X_i : i \in I\}$ be an indexed collection of sets X_i. The Cartesian product of $\mathcal{C}$ is the set of functions that assign to each index $i \in I$ an element $x_i \in X_i$. That is,

$$\prod_{i\in I} X_i = \left\{ f : I \to \bigcup_{i\in I} X_i : f(i) \in X_i \text{ for every } i \in I \right\}.$$

For example, if $I = \{1, 2, \dots, n\}$, then f defines an ordered n-tuple of elements $(x_1, x_2, \dots, x_n)$ with $x_i = f(i) \in X_i$, so this definition is equivalent to our previous one.

If $X_i = X$ for every $i \in I$, then $\prod_{i\in I} X_i$ is simply the set of functions from I to X, and we also write it as

$$X^I = \{f : I \to X\}.$$

We can think of this set as the set of ordered I-tuples of elements of X.

Example 1.27. A sequence of real numbers $(x_1, x_2, x_3, \dots, x_n, \dots) \in \mathbb{R}^{\mathbb{N}}$ is a function $f : \mathbb{N} \to \mathbb{R}$. We study sequences and their convergence properties in Chapter 3.

Example 1.28. Let $\mathbf{2} = \{0, 1\}$ be a set with two elements. Then a subset $A \subset I$ can be identified with its characteristic function $\chi_A : I \to \mathbf{2}$ by: $i \in A$ if and only if $\chi_A(i) = 1$. Thus, $A \mapsto \chi_A$ is a one-to-one map from $\mathcal{P}(I)$ onto $\mathbf{2}^I$.

Before giving another example, we introduce some convenient notation.

Definition 1.29. Let

$$\Sigma = \{(s_1, s_2, s_3, \dots, s_k, \dots) : s_k = 0, 1\}$$

denote the set of all binary sequences; that is, sequences whose terms are either 0 or 1.

Example 1.30. Let $\mathbf{2} = \{0, 1\}$. Then $\Sigma = \mathbf{2}^{\mathbb{N}}$, where we identify a sequence $(s_1, s_2, \dots s_k, \dots)$ with the function $f : \mathbb{N} \to \mathbf{2}$ such that $s_k = f(k)$. We can also identify Σ and $\mathbf{2}^{\mathbb{N}}$ with $\mathcal{P}(\mathbb{N})$ as in Example 1.28. For example, the sequence $(1, 0, 1, 0, 1, \dots)$ of alternating ones and zeros corresponds to the function $f : \mathbb{N} \to \mathbf{2}$ defined by

$$f(k) = \begin{cases} 1 & \text{if } k \text{ is odd}, \\ 0 & \text{if } k \text{ is even}, \end{cases}$$

and to the set $\{1, 3, 5, 7, \dots\} \subset \mathbb{N}$ of odd natural numbers.

1.5. Relations

A binary relation R on sets X and Y is a definite relation between elements of X and elements of Y. We write xRy if $x \in X$ and $y \in Y$ are related. One can also define relations on more than two sets, but we shall consider only binary relations and refer to them simply as relations. If $X = Y$, then we call R a relation on X.

Example 1.31. Suppose that S is a set of students enrolled in a university and B is a set of books in a library. We might define a relation R on S and B by:

$$s \in S \text{ has read } b \in B.$$

In that case, sRb if and only if s has read b. Another, probably inequivalent, relation is:

$$s \in S \text{ has checked } b \in B \text{ out of the library.}$$

When used informally, relations may be ambiguous (did s read b if she only read the first page?), but in mathematical usage we always require that relations are definite, meaning that one and only one of the statements "these elements are related" or "these elements are not related" is true.

The graph G_R of a relation R on X and Y is the subset of $X \times Y$ defined by

$$G_R = \{(x, y) \in X \times Y : xRy\}.$$

This graph contains all of the information about which elements are related. Conversely, any subset $G \subset X \times Y$ defines a relation R by: xRy if and only if $(x, y) \in G$. Thus, a relation on X and Y may be (and often is) defined as subset of $X \times Y$. As for sets, it doesn't matter how a relation is defined, only what elements are related.

A function $f : X \to Y$ determines a relation F on X and Y by: xFy if and only if $y = f(x)$. Thus, functions are a special case of relations. The graph G_R of a general relation differs from the graph G_F of a function in two ways: there may be elements $x \in X$ such that $(x, y) \notin G_R$ for any $y \in Y$, and there may be $x \in X$ such that $(x, y) \in G_R$ for many $y \in Y$.

For example, in the case of the relation R in Example 1.31, there may be some students who haven't read any books, and there may be other students who have read lots of books, in which case we don't have a well-defined function from students to books.

Two important types of relations are orders and equivalence relations, and we define them next.

1.5.1. Orders. A primary example of an order is the standard order $\leq$ on the natural (or real) numbers. This order is a linear or total order, meaning that two numbers are always comparable. Another example of an order is inclusion $\subset$ on the power set of some set; one set is "smaller" than another set if it is included in it. This order is a partial order (provided the original set has at least two elements), meaning that two subsets need not be comparable.

Example 1.32. Let $X = \{1, 2\}$. The collection of subsets of X is

$$\mathcal{P}(X) = \{\varnothing, A, B, X\}, \qquad A = \{1\}, \quad B = \{2\}.$$

We have $\varnothing \subset A \subset X$ and $\varnothing \subset B \subset X$, but $A \not\subset B$ and $B \not\subset A$, so A and B are not comparable under ordering by inclusion.

The general definition of an order is as follows.

Definition 1.33. An order $\preceq$ on a set X is a binary relation on X such that for every $x, y, z \in X$:

(a) $x \preceq x$ (reflexivity);

(b) if $x \preceq y$ and $y \preceq x$ then $x = y$ (antisymmetry);

(c) if $x \preceq y$ and $y \preceq z$ then $x \preceq z$ (transitivity).

An order is a linear, or total, order if for every $x, y \in X$ either $x \preceq y$ or $y \preceq x$, otherwise it is a partial order.

If $\preceq$ is an order, then we also write $y \succeq x$ instead of $x \preceq y$, and we define a corresponding strict order $\prec$ by

$$x \prec y \text{ if } x \preceq y \text{ and } x \neq y.$$

There are many ways to order a given set (with two or more elements).

Example 1.34. Let X be a set. One way to partially order the subsets of X is by inclusion, as in Example 1.32. Another way is to say that $A \preceq B$ for $A, B \subset X$ if and only if $A \supset B$, meaning that A is "smaller" than B if A includes B. Then $\preceq$ in an order on $\mathcal{P}(X)$, called ordering by reverse inclusion.

1.5.2. Equivalence relations. Equivalence relations decompose a set into disjoint subsets, called equivalence classes. We begin with an example of an equivalence relation on $\mathbb{N}$.

Example 1.35. Fix $N \in \mathbb{N}$ and say that $m \sim n$ if

$$m \equiv n \pmod{N},$$

meaning that $m - n$ is divisible by N. Two numbers are related by $\sim$ if they have the same remainder when divided by N. Moreover, $\mathbb{N}$ is the union of N equivalence classes, consisting of numbers with remainders 0, 1,... $N - 1$ modulo N.

The definition of an equivalence relation differs from the definition of an order only by changing antisymmetry to symmetry, but order relations and equivalence relations have completely different properties.

Definition 1.36. An equivalence relation $\sim$ on a set X is a binary relation on X such that for every $x, y, z \in X$:

(a) $x \sim x$ (reflexivity);

(b) if $x \sim y$ then $y \sim x$ (symmetry);

(c) if $x \sim y$ and $y \sim z$ then $x \sim z$ (transitivity).

For each $x \in X$, the set of elements equivalent to x,

$$[x/\sim] = \{y \in X : x \sim y\},$$

is called the equivalence class of x with respect to $\sim$. When the equivalence relation is understood, we write the equivalence class $[x/\sim]$ simply as $[x]$. The set of equivalence classes of an equivalence relation $\sim$ on a set X is denoted by $X/\sim$. Note that each element of $X/\sim$ is a subset of X, so $X/\sim$ is a subset of the power set $\mathcal{P}(X)$ of X.

The following theorem is the basic result about equivalence relations. It says that an equivalence relation on a set partitions the set into disjoint equivalence classes.

Theorem 1.37. Let $\sim$ be an equivalence relation on a set X. Every equivalence class is non-empty, and X is the disjoint union of the equivalence classes of $\sim$.

Proof. If $x \in X$, then the symmetry of $\sim$ implies that $x \in [x]$. Therefore every equivalence class is non-empty and the union of the equivalence classes is X.

To prove that the union is disjoint, we show that for every $x, y \in X$ either $[x] \cap [y] = \varnothing$ (if $x \not\sim y$) or $[x] = [y]$ (if $x \sim y$).

Suppose that $[x] \cap [y] \neq \varnothing$. Let $z \in [x] \cap [y]$ be an element in both equivalence classes. If $x_1 \in [x]$, then $x_1 \sim z$ and $z \sim y$, so $x_1 \sim y$ by the transitivity of $\sim$, and therefore $x_1 \in [y]$. It follows that $[x] \subset [y]$. A similar argument applied to $y_1 \in [y]$ implies that $[y] \subset [x]$, and therefore $[x] = [y]$. In particular, $y \in [x]$, so $x \sim y$. On the other hand, if $[x] \cap [y] = \varnothing$, then $y \notin [x]$ since $y \in [y]$, so $x \not\sim y$. □

There is a natural projection $\pi : X \to X/\sim$, given by $\pi(x) = [x]$, that maps each element of X to the equivalence class that contains it. Conversely, we can index the collection of equivalence classes

$$X/\sim = \{[a] : a \in A\}$$

by a subset A of X which contains exactly one element from each equivalence class. It is important to recognize, however, that such an indexing involves an arbitrary choice of a representative element from each equivalence class, and it is better to think in terms of the collection of equivalence classes, rather than a subset of elements.

Example 1.38. The equivalence classes of $\mathbb{N}$ relative to the equivalence relation $m \sim n$ if $m \equiv n \pmod 3$ are given by

$$I_0 = \{3, 6, 9, \dots\}, \quad I_1 = \{1, 4, 7, \dots\}, \quad I_2 = \{2, 5, 8, \dots\}.$$

The projection $\pi : \mathbb{N} \to \{I_0, I_1, I_2\}$ maps a number to its equivalence class e.g. $\pi(101) = I_2$. We can choose $\{1, 2, 3\}$ as a set of representative elements, in which case

$$I_0 = [3], \qquad I_1 = [1], \qquad I_2 = [2],$$

but any other set $A \subset \mathbb{N}$ of three numbers with remainders 0, 1, 2 (mod 3) will do. For example, if we choose $A = \{7, 15, 101\}$, then

$$I_0 = [15], \qquad I_1 = [7], \qquad I_2 = [101].$$