Sets and functions

1 Sets

The language of sets and functions pervades mathematics, and most of the important operations in mathematics turn out to be functions or to be expressible in terms of functions. We will not define what a set is, but take as a basic (undefined) term the idea of a set X and of membership $x \in X$ (x is an element of X). The negation of $x \in X$ is $x \notin X$: x is not an element of X. Typically, the elements of a set will themselves be sets, underscoring the point that, in mathematics, everything is a set. A set can be described (i) as a list $\{x_1, \dots, x_n\}$ or (ii) by giving a description of its elements, e.g. the set of positive real numbers is described via

$$\{x \in \mathbb{R} : x > 0\},$$

where $\mathbb{R}$ denotes the set of all real numbers. A very important set is the *empty set* $\emptyset$: for all x, $x \notin X$. Thus $\emptyset$ has **no** elements. Two sets X and Y are by definition equal if they have the same elements: $X = Y$ if, for all x, $x \in X \iff x \in Y$. We can say this somewhat informally as follows: **a set is uniquely specified by its elements**. This implies by logic that $\emptyset$ is **uniquely** specified by the condition that, for all x, $x \notin X$: there is exactly one empty set. If, for every $x \in X, x \in Y$, then X is a *subset* of Y, written $X \subseteq Y$. Note that we always have $X \subseteq X$ and $\emptyset \subseteq X$. A subset A of X is called a *proper* subset if $A \neq X$. By the definition of equality of sets, $X = Y \iff X \subseteq Y$ and $Y \subseteq X$. If $X \subseteq Y$ and $Y \subseteq Z$, then $X \subseteq Z$; this is called the *transitivity property*. The notation $X \subset Y$ is sometimes used to mean that $X \subseteq Y$ but $X \neq Y$. A set of the form $X = \{x_1, \dots, x_n\}$ is a *finite* set. If for all i, j with $1 \leq i, j \leq n$, we have $x_i \neq x_j$ then we write $\#(X) = n$. By logic or convention, $\emptyset$ is finite and $\#(\emptyset) = 0$. Conversely, if X is a (finite) set with $\#(X) = 0$, then $X = \emptyset$. A set of the form $\{x\}$ has exactly one element. In particular, $\{\emptyset\}$ has a single element, namely $\emptyset$, and thus $\{\emptyset\} \neq \emptyset$.

Recall the standard operations on sets:

Definition 1.1. If X_1 and X_2 are two sets, then:

1. The *union* of X_1 and X_2 is the set

$$X_1 \cup X_2 = \{x : x \in X_1 \text{ or } x \in X_2\}.$$

Thus $X_1 \subseteq (X_1 \cup X_2)$ and $X_2 \subseteq (X_1 \cup X_2)$. The union of finitely many sets is defined similarly: If $X_1, \ldots, X_n$ are sets, then

$$\bigcup_{i=1}^{n} X_i = \{x : \text{ for some } i,\ x \in X_i\}.$$

2. The *intersection* of X_1 and X_2 is:

$$X_1 \cap X_2 = \{x : x \in X_1 \text{ and } x \in X_2\}.$$

Thus $(X_1 \cap X_2) \subseteq X_1$ and $(X_1 \cap X_2) \subseteq X_2$. Likewise

$$\bigcap_{i=1}^{n} X_i = \{x : \text{ for all } i,\ x \in X_i\}.$$

3. Given two sets X_1 and X_2, the *complement* of X_2 in X_1, written $X_1 - X_2$, is the set

$$\{x \in X_1 : x \notin X_2\}.$$

Thus $X_2 \cap (X_1 - X_2) = \emptyset$. If $X_2 \subseteq X_1$, then $X_2 \cup (X_1 - X_2) = X_1$. For example, $X - X = \emptyset$ and $X - \emptyset = X$.

By logic (deMorgan's laws), $Y - (X_1 \cap X_2) = (Y - X_1) \cup (Y - X_2)$ and $Y - (X_1 \cup X_2) = (Y - X_1) \cap (Y - X_2)$. (For example, if $x \in Y, x \notin X_1 \cap X_2$, then either $x \notin X_1$ or $x \notin X_2$ and conversely.)

Definition 1.2. Given X and Y, we define $X \times Y$, the *Cartesian product* of X and Y, to be the set of ordered pairs (x, y) with $x \in X$ and $y \in Y$. Here x is the *first component* or *first coordinate* of the ordered pair (x, y) and y is the *second component* (or coordinate). Clearly, if $A \subseteq X$ and $B \subseteq Y$, then $A \times B \subseteq X \times Y$.

If $X = Y$, we abbreviate $X \times X$ by X^2. Likewise, if we have n sets $X_1, \ldots, X_n$, then $X_1 \times \cdots \times X_n$ is the set of ordered n-tuples $(x_1, \ldots, x_n)$ with $x_i \in X_i$ for every i, the i^{th} *component* (or coordinate) of $(x_1, \ldots, x_n)$ is x_i, and we again abbreviate $X \times \cdots \times X$ (n times) by X^n.

Remark 1.3. The operative properties of an ordered pair (x, y) are: 1) For all $x \in X$ and $y \in Y$, there exists an ordered pair $(x, y) \in X \times Y$, and 2) two ordered pairs (x_1, y_2) and (x_2, y_2) are equal $\iff$ they have the same first components and the same second components, i.e. $\iff$ $x_1 = x_2$ and $y_1 = y_2$; it is not enough to require that the sets $\{x_1, y_1\}$ and $\{x_2, y_2\}$ be equal. It is possible to give a formal definition of an ordered pair just using set theory. In fact, one can define $(x, y) = \{\{x\}, \{x, y\}\}$. (In other words, ordered pair does not have to be an undefined term.) However, we shall not really care what the precise definition is, but only that an ordered pair has the operative properties 1) and 2) above. Using functions, though, we can give a careful definition of an ordered n-tuple; we shall describe this later.

If X and Y are finite sets, then $X \times Y$ is also finite, and

$$\#(X \times Y) = \#(X)\#(Y),$$

and similarly for the product of n finite sets $X_1 \times \cdots \times X_n$. In particular, this formula says that $\#(\emptyset \times X) = \#(X \times \emptyset) = 0$ for every (finite) set X and hence that, if X is finite, then $\emptyset \times X = X \times \emptyset = \emptyset$. Of course, it is easy to check by logic that $\emptyset \times X = X \times \emptyset = \emptyset$ for every set X (not necessarily finite).

The set of all subsets of X is also a set, and is called the *power set* of X, often denoted $\mathcal{P}(X)$:

$$\mathcal{P}(X) = \{A : A \subseteq X\}.$$

By the transitivity property, if $Y \subseteq X$, then $\mathcal{P}(Y)$ is a subset of $\mathcal{P}(X)$. Note that $X \in \mathcal{P}(X)$ and that $\emptyset \in \mathcal{P}(X)$. If $X \neq \emptyset$ and $x \in X$, then $\{x\} \in \mathcal{P}(X)$. Examples: $\mathcal{P}(\emptyset) = \{\emptyset\}$. In particular, $\mathcal{P}(\emptyset) \neq \emptyset$; in fact, $\mathcal{P}(\emptyset)$ contains the unique element $\emptyset$, and thus $\#(\mathcal{P}(\emptyset)) = 1$. Likewise, $\mathcal{P}(\mathcal{P}(\emptyset)) = \mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$. In particular, $\#(\mathcal{P}(\mathcal{P}(\emptyset))) = 2$. Likewise,

$$\mathcal{P}(\{\emptyset, \{\emptyset\}\}) = \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\},$$

and hence $\#(\mathcal{P}(\{\emptyset, \{\emptyset\}\})) = 4$. More generally, we shall see that, if X is a finite set and $\#(X) = n$, then $\#(\mathcal{P}(X)) = 2^n$.

2 Functions

Next we define a function $f \colon X \to Y$. Although we can think of a function as a "rule" which assigns to every $x \in X$ a unique $y \in Y$, it is easier to make this concept precise by identifying the function f with its graph in $X \times Y$ (as we were taught **not** to do in calculus). Thus a function f is the same thing as a subset G_f of $X \times Y$ with the following property: for all $x \in X$, there is a unique $y \in Y$ such that $(x, y) \in G_f$ and we set $y = f(x)$. To say that there is a unique $y \in Y$ says that $f(x)$ is uniquely determined by x, and to say that for every $x \in X$ there exists an $(x, y) \in G_f$ says that in fact $f(x)$ is defined for all $x \in X$. This is the so-called *vertical line test*: for each $x \in X$, we have the subset $\{x\} \times Y$ of $X \times Y$. (In case $X = Y = \mathbb{R}$, such subsets are exactly the vertical lines.) Then G is the graph of a function f if and only if, for every $x \in X$, $(\{x\} \times Y) \cap G$ consists of exactly one point, necessarily of the form (x, y) for some $y \in Y$. The unique such y is then $f(x)$.

In the above notation, we call X the *domain* of f and Y the *range*. Thus the domain and range are a part of the information of a function. Note that a function *must be defined at all elements of its domain*; thus for example the function $f(x) = 1/x$ cannot have domain $\mathbb{R}$ without assigning some value to $f(0)$. (This is in contrast to the practice in some calculus courses where f is allowed to be not everywhere defined.) Two functions f_1 and f_2 are equal if and only if their graphs are equal, if and only if, for all $x \in X$, $f_1(x) = f_2(x)$. Thus, just as a set is specified by its elements, **a function is uniquely specified by its values**. We emphasize, though, that for two functions f_1 and f_2 to be equal, they **must** have the same domain and range. We shall use the word *map* or *mapping* as a synonym for function; typically maps are functions in some kind of geometric setting.

Definition 2.1. Let $f\colon X \to Y$ be a function. The set

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

is called the *image* of f and is sometimes written $f(X)$. (Sometimes people call the range the *codomain* and define the range to be the image.) More generally, if A is a subset of X, then we set

$$f(A) = \{y \in Y : \text{ there exists an } x \in A \text{ such that } f(x) = y\}.$$

We say that f is *surjective* or *onto* if $f(X) = Y$, in other words if the image of f is Y. In general the image of a function $f(x)$ will be a subset of the range, but **need not** equal the range.

We also define, for B a subset of Y, the subset

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

of X, called the *preimage* of B. If $B = \{y\}$ has just one element we write $f^{-1}(y)$ instead of $f^{-1}(\{y\})$. For example, $f^{-1}(Y) = X$ and $f^{-1}(y) \neq \emptyset$ if and only if $y \in f(X)$.

Note: if Y is a subset of another set Y', then a function $f \colon X \to Y$ defines (in an obvious way) a function from X to Y'. Technically these are two *different* functions, although we will occasionally (and incorrectly) blur the distinction. Also, given a function $f \colon X \to Y$, we can always replace it by a function from X to $f(X) \subseteq Y$, and this new function is always surjective.

Often we need to restrict the values of given function, leading to the following:

Definition 2.2. If $f \colon X \to Y$ is a function and $A \subseteq X$, then we define the *restriction* $(f|A)$ of f to A to be the function $(f|A) \colon A \to Y$ defined by $(A \times Y) \cap G_f$, where G_f is the graph of f. In other words, $(f|A)(a) = f(a)$ for all $a \in A$, and the domain of $(f|A)$ is exactly A. If moreover $f(A) \subseteq B$, there is the *induced* function $g \colon A \to B$, which technically is different from $(f|A)$. However we shall sometimes be a little careless.

Definition 2.3. A function $f \colon X \to Y$ is *injective* or *one-to-one* if, for all $x_1, x_2 \in X$, $f(x_1) = f(x_2)$ if and only if $x_1 = x_2$. Equivalently, for all $y \in Y$, the set $f^{-1}(y)$ has at most one element. Thus f is injective if, for all $y \in Y$, the equation $f(x) = y$ has at most one solution, or in other words if a solution exists, then it is unique. By contrast, f is surjective if the equation $f(x) = y$ has a solution (not necessarily unique) for every $y \in Y$. A function $f \colon X \to Y$ which is one-to-one and onto is called a *bijection* or a *one-to-one correspondence*.

For example, taking $X = \mathbb{R}$, the function $f(x) = x^2$ is neither injective nor surjective. (When is $x_1^2 = x_2^2$? What is the image of f?) The function $f(x) = e^x$ is injective but not surjective. (What is the image of f?) The function $f(x) = x^3 + 1$ is a bijection. The identity function $\mathrm{Id}_X \colon X \to X$ is always a bijection.

The property of being injective or surjective can also be described via "horizontal lines," in other words by subsets of $X \times Y$ of the form $X \times \{y\}$ (which are exactly the horizontal lines in case $X = Y = \mathbb{R}$). A function $f \colon X \to Y$ is injective $\iff$ for every $y \in Y$, the intersection of the graph G_f with $X \times \{y\}$, i.e. $G_f \cap (X \times \{y\})$, has *at most one point*. The function f is surjective $\iff$ for every $y \in Y$, the intersection of the graph G_f with $X \times \{y\}$, i.e. $G_f \cap (X \times \{y\})$, has *at least one point*. Thus, f is a bijection

$\iff$ for every $y \in Y$, the intersection of the graph G_f with $X \times \{y\}$, i.e. $G_f \cap (X \times \{y\})$, has *exactly one point*. This can be interpreted as follows: for every subset A of $X \times Y$, we get a new subset ${}^tA$ of $Y \times X$ by:

$${}^tA = \{(y, x) : (x, y) \in A\}.$$

This is the abstract analogue of “reflection about the diagonal.” Then:

Proposition 2.4. *The function $f\colon X \to Y$ is a bijection $\iff$ the subset ${}^tG_g \subseteq Y \times X$ is the graph of a function from Y to X. This function is denoted f^{-1}.* □

Here are some basic examples of functions:

1. For any set X, the *identity function* $\mathrm{Id}_X\colon X \to X$ satisfies: $\mathrm{Id}_X(x) = x$ for every $x \in X$. Thus its graph in X^2 is the set $\Delta_X = \{(x, x) : x \in X\}$, which we can think of as the “diagonal” viewed as a subset of X^2. (Does the diagonal satisfy the test of being the graph of a function?) The preimage of $A \subseteq X$ is just A. When X is clear from the context, we will often abbreviate Id_X to Id.

2. A related example is inclusion: if $X \subseteq Y$, then $\{(x, x) : x \in X\}$ is a subset of $X \times Y$ which is the graph of the inclusion function from X to Y, which we shall often denote by i_X. Note that $i_X = \mathrm{Id}_Y|X$. The image $i_X(A)$ of a subset A of X is then A, viewed as a subset of Y, and the preimage $i_X^{-1}(B)$ of $B \subseteq Y$ is $B \cap X$.

3. Another example, if $Y \neq \emptyset$, is a *constant function*: choose $c \in Y$ and define $f(x) = c$ for all $x \in X$. (What is the graph of this function and why is it a function?) In this case, the preimage of a subset B of Y is $\emptyset$ if $c \notin B$ and is X if $c \in B$. The image $f(A)$ of a

4. Of course, all of the standard functions of calculus give examples of functions from $\mathbb{R}$ to $\mathbb{R}$. (If $f\colon \mathbb{R} \to \mathbb{R}$ is the function $f(x) = x^2$, what is the image of f? What is $f^{-1}(a)$?)

5. Another example is the Cartesian product $X \times X$, which we can identify with the set of functions $f\colon \{1, 2\} \to X$. In fact, in the notation for an ordered pair (x_1, x_2), we can think of x_i as a function which assigns the value x_1 to 1 and x_2 to 2. If we wanted to define the Cartesian product of two possibly different sets in this way, we could define $X \times Y$ to be the set of all functions $f\colon \{1, 2\} \to X \cup Y$ such that $f(1) \in X$ and $f(2) \in Y$. Likewise X^n is identified with the set of

functions $f\colon \{1, 2, \dots, n\} \to X$. A sequence $x_1, x_2, \dots$ of real numbers is then the same as a function $\mathbb{N} \to \mathbb{R}$, where $\mathbb{N}$ is the set of natural numbers $\{1, 2, \dots\}$. Here, given a function $f\colon \mathbb{N} \to \mathbb{R}$, we define a sequence $x_1, x_2, \dots$ via $x_i = f(i)$, and conversely. More generally, if X is any set, possibly finite, then a sequence $x_1, x_2, \dots$ with values in X is the same thing as a function $f\colon \mathbb{N} \to X$.

6. A *function of two variables* is the same thing as a function $f\colon X \times Y \to Z$ of a single "variable," in other words we evaluate f on elements of $X \times Y$, which are ordered pairs. Traditionally, we write $f(x, y)$ instead of $f((x, y))$ for the value of f on (x, y). Similarly for functions of n variables.

7. If X and Y are two sets, then the set of all functions from X to Y is a new set, sometimes denoted by Y^X:

$$Y^X = \{f : \ f \text{ is a function from } X \text{ to } Y\}.$$

If X and Y are finite, say $\#(X) = n$ and $\#(Y) = m$, then Y^X is also finite and $\#(Y^X) = m^n$.

Given $x \in X$, we get a function ev_x from Y^X to Y by evaluating at x:

$$\mathrm{ev}_x(f) = f(x).$$

Thus, when we write $f(x)$ above, the symbol f has become the "variable." There is a similar function of two variables, $e\colon X \times Y^X \to Y$, defined by

$$e(x, f) = f(x).$$

Remark 2.5. If X and Y are finite sets, and $f\colon X \to Y$ is a bijection, then $\#(X) = \#(Y)$. In fact, we can define a finite set X in the following way: X is finite $\iff$ for some natural number n, there exists a bijection from the set $\{1, \dots, n\}$ to X, and in this case $\#(X) = n$. (By definition, therefore, an *infinite set* X is one for which, for every natural number n, there is no bijection from $\{1, \dots, n\}$ to X.)

More generally, if X and Y are finite sets, then

1. If there exists a bijection $f\colon X \to Y$, then $\#(X) = \#(Y)$.

2. If there exists an injection $f\colon X \to Y$, then $\#(X) \le \#(Y)$.

3. If there exists a surjection $g\colon X \to Y$, then $\#(X) \ge \#(Y)$.

4. If $\#(X) = \#(Y)$ and $h\colon X \to Y$ is a function, then h is an injection $\iff$ h is a surjection $\iff$ h is a bijection.

Any one of the last three facts is referred to as the *pigeonhole principle*.

It follows easily that, if X is finite and A is a proper subset of X, then $\#(A) < \#(X)$ and there is no bijection from A to X. It turns out that infinite sets can be characterized by the opposite property: X is infinite $\iff$ a proper subset A of X and a bijection from A to X.

Given functions $f\colon X \to Y$ and $g\colon Y \to Z$, we have the *composition* $g \circ f\colon X \to Z$ defined by

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

for all $x \in X$. For example, given $f\colon X \to Y$, $\mathrm{Id}_Y \circ f = f \circ \mathrm{Id}_X = f$. Thus, the identity function behaves very much like an identity element under composition, as long as we are careful about the domains of the relevant identity functions.

The operation of function composition is somewhat like an algebraic operation, in that we can sometimes "combine" two functions and get a third. But we can't always do so: we can only define $g \circ f$ when the range of f is equal to the domain of g.

Function composition has the important property that it is associative where defined:

Proposition 2.6. *Suppose given functions $f\colon X \to Y$, $g\colon Y \to Z$, and $h\colon Z \to W$. Then*

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

Proof. For all $x \in X$, $(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x)))$, and likewise $((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x)))$. Thus $(h \circ (g \circ f))(x) = ((h \circ g) \circ f)(x)$ for all $x \in X$ and so $h \circ (g \circ f) = (h \circ g) \circ f$. □

In general function composition is not commutative. For example, given $f\colon X \to Y$, we can only compose it with $g\colon Y \to Z$ in both orders when $X = Z$. In this case $g \circ f\colon X \to X$ and $f \circ g\colon Y \to Y$, and we can only compare these when $X = Y$. Finally, very simple examples show that even when $Y = X$, if we pick two random functions $f\colon X \to X$ and $g\colon X \to X$, then $g \circ f \neq f \circ g$ (as long as X has more than one element). In other words, the composition of two random functions, whose domain and range are both equal to a fixed set X, will depend on the order (for example, take $X = \mathbb{R}$, $f(x) = e^x, g(x) = x^2 + 1$, and check that $g \circ f \neq f \circ g$).

We have seen that identity functions work much like identity elements for addition or multiplication of real numbers. We can also ask about inverse functions. It turns out that inverse functions are related to bijections. Let $f\colon X \to Y$ be a function. An *inverse function* $g\colon Y \to X$ is a function g such that $g \circ f = \mathrm{Id}_X$ and $f \circ g = \mathrm{Id}_Y$. As we will show soon, if an inverse function exists it is unique and is denoted f^{-1}. This should **not** be confused with the preimage which can be defined for any function, and it should never be confused with $1/f$, which can be defined for a real-valued function which is never zero. For example, if $f\colon X \to Y$ is a bijection with inverse f^{-1}, then $f^{-1}(y)$ could potentially mean the value of f^{-1} on y or the **preimage** of y, which is the **subset** $f^{-1}(\{(y)\}) \subseteq X$.

In a similar vein, a *left inverse* for f is a function g such that $g \circ f = \mathrm{Id}_X$, and a *right inverse* for f is a function g such that $f \circ g = \mathrm{Id}_Y$. It is possible for a function to have a right inverse but not a left inverse, and vice-versa. However, if a function has both a right and a left inverse they are equal:

Proposition 2.7. *Suppose that $f\colon X \to Y$ is a function, and that $g\colon Y \to X$ and $h\colon Y \to X$ are functions such that $g \circ f = \mathrm{Id}_X$ and $f \circ h = \mathrm{Id}_Y$. Then $g = h$ and so g is an inverse function for f.*

Proof. Consider $g \circ f \circ h$. Since function composition is associative, this is

$$(g \circ f) \circ h = \mathrm{Id}_X \circ h = h$$

but associating the other way says that it is also equal to

$$g \circ (f \circ h) = g \circ \mathrm{Id}_Y = g.$$

Hence $g = h$. □

Corollary 2.8. *If g_1 and g_2 are two inverse functions for f, then $g_1 = g_2$. In other words, an inverse function, if it exists, is unique.*

Proof. Since an inverse function is both a left and a right inverse, we can apply the previous proposition, viewing, say, g_1 as a right inverse and g_2 as a left inverse, to conclude that $g_1 = g_2$. □

Note that g is a left inverse for $f \iff f$ is a right inverse for g, and similarly for right inverses. In particular, if f has an inverse f^{-1}, then f is a right and a left inverse for f^{-1}, hence an inverse to f^{-1}. We can express this by:

Proposition 2.9. *Suppose that $f\colon X \to Y$ has an inverse function $f^{-1}\colon Y \to X$. Then f^{-1} also has an inverse function, and in fact it is necessarily equal to f. In other words,*

$$(f^{-1})^{-1} = f. \quad \square$$

The relation between left and right inverses and injectivity and surjectivity, is given by the following:

Proposition 2.10. *Let $f\colon X \to Y$ be a function.*

1. *Suppose that $X \neq \emptyset$. Then f has a left inverse if and only if f is injective.*

2. *f has a right inverse if and only if f is surjective.*

3. *f has an inverse if and only if f is a bijection.*

Proof. (1), (2): left as homework.
(3) (Outline.) This follows from (1) and (2). However, in this important case, we can give a more direct argument as follows. Use the fact that $f\colon X \to Y$ is a bijection $\iff$ ${}^tG_f \subseteq Y \times X$ is the graph of a function $g\colon Y \to X$ and check that, necessarily, $g \circ f = \mathrm{Id}_X$ and $f \circ g = \mathrm{Id}_Y$. Conversely, if $f^{-1}\colon Y \to X$ is an inverse function, then it is easy to see that $G_{f^{-1}} = {}^tG_f$, hence f is a bijection. $\square$

In particular, a very efficient way to show that a function is a bijection is to exhibit an inverse function for it.

The composition of two injections is an injection and the composition of two surjections is a surjection (homework). Thus the composition of two bijections is a bijection. However, in light of the above remark, it is better to show this last statement by describing the inverse function to the composition, which has the advantage of also giving a formula for the inverse. Note the reversal of order in the formula, which is a basic fact of life.

Proposition 2.11. *Suppose that $f\colon X \to Y$ has an inverse function $f^{-1}\colon Y \to X$ and that $g\colon Y \to Z$ has an inverse function $g^{-1}\colon Z \to Y$. Then $g \circ f$ has an inverse, and it is equal to $f^{-1} \circ g^{-1}$.*

Proof. We must check **both** equalities

$$(g \circ f) \circ (f^{-1} \circ g^{-1}) = \mathrm{Id}_Z;$$
$$(f^{-1} \circ g^{-1}) \circ (g \circ f) = \mathrm{Id}_X\,.$$

Since these are similar, we shall just check the first: by associativity,

$$\begin{aligned}(g \circ f) \circ (f^{-1} \circ g^{-1}) &= g \circ (f \circ (f^{-1}) \circ g^{-1} \\ &= (g \circ \mathrm{Id}_Y) \circ g^{-1} \\ &= g \circ g^{-1} = \mathrm{Id}_Z\,.\end{aligned}$$

□

Bijections express the idea that two sets have the same number of elements. We have already discussed this for finite sets. For infinite sets this can be used to *define* what it means for two infinite sets to have the same number of elements (Cantor). But such bijections might be very non-obvious. For example, one can show that there is a one-to-one correspondence from $\mathbb{R}$ to $\mathbb{R}^2$, or in fact to $\mathbb{R}^n$ for any $n > 0$, but such a bijection does not have geometric properties and is hard to write down in any explicit way. Its existence says that $\mathbb{R}$ and $\mathbb{R}^2$ have the same number of elements from a purely quantitative point of view, but in no geometric sense do $\mathbb{R}$ and $\mathbb{R}^2$ resemble each other.

On the other hand, especially in algebra, we often look for "good" bijections, which might tell us that two sets might be essentially the same even if technically different. For example, the sets $X \times Y$ and $Y \times X$ are different sets if $X \neq Y$, but there is a natural function $F\colon X \times Y \to Y \times X$ defined by $F(x, y) = (y, x)$. This function is a bijection: if $F(x_1, y_1) = F(x_2, y_2)$, then by definition $(y_1, x_1) = (y_2, x_2)$ as ordered pairs in $Y \times X$. Hence by the operative property of equality of ordered pairs, $y_1 = y_2$ and $x_1 = x_2$, and thus $(x_1, y_1) = (x_2, y_2)$. Hence F is injective. To see that it is surjective, let (y, x) be an arbitrary element of $Y \times X$. Then $(y, x) = F(x, y)$. Thus F is surjective and hence a bijection. Without verifying directly that F is both injective and surjective, we could try to find an inverse function $G\colon Y \times X \to X \times Y$. What is the inverse function? Likewise, there is a one-to-one correspondence from $X_1 \times (X_2 \times X_3)$ to $(X_1 \times X_2) \times X_3$, and from either of these sets to $X_1 \times X_2 \times X_3$. In fact, these bijections are so obvious that we don't always write them down explicitly.

For another example, we can identify the power see $\mathcal{P}(X)$ with the set of all functions from X to $\{0, 1\}$, i.e. with $\{0, 1\}^X$. Given $A \in \mathcal{P}(X)$, define the *characteristic function* $\chi_A\colon X \to \{0, 1\}$ by:

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

Then $\chi_A \in X^{\{0,1\}}$. Conversely, if $f \in \{0,1\}^X$, i.e. f is a function from X to $\{0,1\}$, define $S_f = f^{-1}(1) = \{x \in X : f(x) = 1\}$. More formally, we have defined a bijection $F\colon \mathcal{P}(X) \to \{0,1\}^X$ by the formula

$$F(A) = \chi_A.$$

The function $G\colon \{0,1\}^X \to \mathcal{P}(X)$ defined by

$$G(f) = f^{-1}(1)$$

is the inverse function to F, as one can check by verifying **both** statements

$$G \circ F = \mathrm{Id}_{\mathcal{P}(X)};$$
$$F \circ G = \mathrm{Id}_{\{0,1\}^X}.$$

For example, the first statement is the statement that, for every subset A of X,

$$(G \circ F)(A) = \chi_A^{-1}(1) = A,$$

where $\chi_A^{-1}(1)$ denotes the preimage of 1 under the function χ_A. The second statement is the statement that, if $f\colon X \to \{0,1\}$ is a function, then

$$(F \circ G)(f) = \chi_{f^{-1}(1)} = f.$$

These statements are checked by unwinding the definitions, and are part of the homework.

Thus, by Remark 2.5, if X is a finite set and $\#(X) = n$, then

$$\#(\mathcal{P}(X)) = \#(\{0,1\}^X) = 2^n.$$

The above examples illustrate a general pattern in mathematics. Given two sets X and Y, a bijection from X to Y is often found by giving (1), for every element $x \in X$ a "construction" of an element $y \in Y$, which we interpret as a function $F\colon X \to Y$; (2) similarly, for every element $y \in Y$ a "construction" of an element $x \in X$, which we interpret as a function $G\colon Y \to X$; (3) a proof that these are "inverse constructions," i.e. if we construct y from x and then do the corresponding construction on y, we get back the element $x \in X$ that we started with, and similarly in the other order. This last statement is equivalent to the assertions that $G \circ F = \mathrm{Id}_X$ and $F \circ G = \mathrm{Id}_Y$. Of course, we hope that the "constructions," in other words the functions F and G, are natural ones to consider in some vague sense.

Finally, we discuss the set of all bijections from a set to itself. This object will recur throughout the semester.

Definition 2.12. Let X be a set. We define S_X, the *set of permutations of* X, to be the set of all bijections $f\colon X \to X$. Thus $S_X \subseteq X^X$, the set of all functions from X to X.

Note that $\mathrm{Id}_X \in S_X$. If $f, g \in S_X$ then $g \circ f \in S_X$, and if $f \in S_X$, then $f^{-1} \in S_X$. In other words, S_X is closed under composition and every element of S_X has an inverse, which is also in S_X. For a finite set X with $\#(X) = n$, we usually take for X the standard finite set with n elements, namely $\{1, \dots, n\}$, and abbreviate $S_{\{1,\dots,n\}}$ by S_n. By counting, $\#(S_n) = n!$, since, to define a bijection $f\colon \{1, \dots, n\} \to \{1, \dots, n\}$, there are n possible choices for $f(1)$, but only $n - 1$ choices for $f(2)$ since the value $f(1)$ is excluded (as f is injective, we can't have $f(1) = f(2)$. Continuing, there are only $n - 2$ choices for $f(3)$, …, 2 choices for $f(n - 1)$, and one choice for $f(n)$. This says that the total number of injections $\{1, \dots, n\} \to \{1, \dots, n\}$ is

$$n(n-1)\cdots 2 \cdot 1 = n!.$$

But by Remark 2.5, an injection $\{1, \dots, n\} \to \{1, \dots, n\}$ is the same thing as a bijection $\{1, \dots, n\} \to \{1, \dots, n\}$. Thus $\#(S_n) = n!$. Of course, a similar argument shows that $\#(S_X) = n!$ for any finite set X with $\#(X) = n$.