Section 1.3 Ordered pairs and functions

The ordered pair (a, b) of two things a and b is another thing that contains the information of both a and b, together the information that " a comes first, b second". Mathematically expressed, the essential property of the ordered-pair construction is

$$(a, b) = (c, d) \iff a = c \text{ and } b = d . \tag{1}$$

It is possible to construct the ordered pair set-theoretically; however, we will not do so here; all we ever use about ordered pairs is the fact expressed in (1). Let us note though that the pair-set $\{a, b\}$ would *not* work as the ordered pair: we have

$$\{0, 1\} = \{1, 0\} ,$$

but we want

$$(0, 1) \neq (1, 0) .$$

The use of the ordered pair is familiar in coordinate geometry; the points in the plane equipped with a Cartesian coordinate system are represented by ordered pairs of real numbers. Various geometric figures become sets of ordered pairs. Denoting the set of ordered pairs of real numbers by $\mathbb{R}^2$, the set

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

is the circle with center the origin, and radius the unit length; that is, the set of points on that circle. The set

$$\{(x, y)\in\mathbb{R}^2 : x^2 + y^2 < 1\}$$

is the open disc of radius 1 around the origin;

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

is the closed disc; the open disc does not, the closed one does, contain the circumference.

For sets A and B, $A \times B$, the *Cartesian product of* A *and* B, is the set of all ordered pairs (a, b) with first element a from A, second element b from B:

$$A \times B \underset{\text{def}}{=} \{(a, b) : a \in A \text{ and } b \in B\}.$$

Thus, what we wrote as $\mathbb{R}^2$ above is the same as $\mathbb{R} \times \mathbb{R}$; in general, we may write A^2 for $A \times A$.

A *function* f from a set A to another set B is a rule that assigns, to every element a of A, a definite element of B; this element is denoted by $f(a)$; it is called the *value* of the function f at the *argument* a. We write

$$f : A \longrightarrow B$$

to indicate that f is a function from A to B; A is the *domain* of f, B is the *codomain* of f. The codomain of f has to be distinguished from the *range* of f; the latter is the set $\{f(a) : a \in A\}$ of all values of f:

$$\text{range}(f) \underset{\text{def}}{=} \{f(a) : a \in A\}.$$

The range of f is a subset of the codomain of f; the range and the codomain may or may not be the same.

It is possible to construe functions as sets, in particular, as sets of ordered pairs: with $f : A \longrightarrow B$, we may consider the set of all pairs $(a, f(a))$ with $a \in A$; this set is called the *graph of* the function f:

$$\text{graph}(f) \underset{\text{def}}{=} \{(a, f(a)) : a \in A\}.$$

This is exactly the representation of functions that we use in coordinate geometry and calculus.

For instance, with the exponential function $\mathtt{exp}:\mathbb{R}\longrightarrow\mathbb{R}$ assigning e^x to x for all $x \in \mathbb{R}$, we associate its graph which is the exponential curve in the Cartesian plane.

Note that the range of $\mathtt{exp}:\mathbb{R}\longrightarrow\mathbb{R}$ is the set of all positive real numbers, $\mathbb{R}^+ = \{y\in\mathbb{R}: y>0\}$. This is true since the values of the exponential function are all positive, and every positive real number is the value of $\mathtt{exp}$ at a suitable argument $x\in\mathbb{R}$: if $y>0$, then there is $x\in\mathbb{R}$ namely, $x=\ln(y)$, for which $y=f(x)$. The range of $\mathtt{exp}:\mathbb{R}\longrightarrow\mathbb{R}$ does not coincide with its codomain: $\mathbb{R}^+ \subsetneqq \mathbb{R}$.

Usually, we do not distinguish between the function and its graph; the exponential function and the exponential curve are considered to be the same thing. There is one qualification to this rule though: two functions $f:A\longrightarrow B$ and $g:A\longrightarrow C$, with the same domain but with different codomains, may have the same graph. E.g., the $\sin$ function may be construed as $\sin:\mathbb{R}\longrightarrow\mathbb{R}$, from $\mathbb{R}$ to $\mathbb{R}$, or as $\sin:\mathbb{R}\longrightarrow[-1,1]$, from $\mathbb{R}$ to the closed interval $[-1, 1]$ (since all values of $\sin$ are in the latter interval); these two functions have the same graph. For us, these two functions are technically different; the specification of a function includes the specification of its domain as well as its codomain.

When is a set, say A , is the graph of a function? There are two conditions that are necessary and sufficient for this to hold:

(i) Every element a of A must be an ordered pair: a must equal to (x, y) for suitable (uniquely determined) x and y ;

(ii) For all x, y, z , $(x, y)\in A$ and $(x, z)\in A$ imply that $y=z$ [note the same x as first component in the two ordered pairs].

The second condition expresses the fact that for a function f , the value $y=f(x)$ is *uniquely determined* by the argument x . If (i) and (ii) hold true, then there is a function $f:X\longrightarrow Y$ for which $\mathrm{graph}(f)=A$. Here, X , the domain of f , is the set of all x for which there is y such that $(x, y)\in A$; Y , the codomain, is any set that *contains* as a subset the set R of all y for which there is x such that $(x, y)\in A$ (R is the *range* of f); and we have

$$y = f(x) \iff (x, y) \in A .$$

The usual notation for a function is to give its value at an indeterminate argument; thus, e^x

denotes the exponential function. This notation is ambiguous, however; it may also mean the value of the function at a certain argument-value of x . A more explicit notation e.g. for the exponential function is

$$x \longmapsto e^x \qquad\qquad (x \in \mathbb{R})$$

Note here the vertical line at the beginning of the arrow; this kind of arrow is to be distinguished from the arrow that connects the domain and codomain of the function. If we write $\mathtt{exp}$ for the exponential function, a full notation and description of the function $\mathtt{exp}$ is this:

$$\begin{aligned} \mathtt{exp} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto e^x \ . \end{aligned}$$

If we have two functions $f{:}A \longrightarrow B$ and $g{:}A \longrightarrow B$ between the same two sets, f and g are the *same function*, $f = g$, just in case for all $a \in A$, $f(a) = g(a)$:

$$f = g \iff \text{for all } a \in A\ ,\ f(a) = g(a)\ .$$

This is in agreement with the construal of functions as sets of ordered pairs: $f = g$ just in case $\mathtt{graph}(f) = \mathtt{graph}(g)$; note that this is valid only if the two functions f and g are given already with the same domain and the same codomain.

Here is a notation for specifying a function when the domain of the function is a reasonably small finite set. I'll explain this on an example. For instance, the symbolic expression

$$\begin{pmatrix} 1 & 3 & 5 & 7 & 9 & 11 \\ 0 & 5 & 4 & 20 & 3 & 3 \end{pmatrix} \tag{2}$$

denotes the function whose domain is the set $\{1, 3, 5, 7, 9, 11\}$, the set which is listed in the upper row, and whose value for each argument in the domain is given in the second row underneath the particular argument; in the case of (2), if the function is called f , then $f(1)=0$, $f(3)=5$, $f(5)=4$, etc.

To be precise, we should note that this notation exhibits only the *graph* of the function. In the example (2), the function f may have any codomain (which then has to be specified separately) that contains the set $\{0, 5, 4, 20, 3\}$, the range of the function f.

If we have two functions, $f:A \longrightarrow B$ and $g:B \longrightarrow C$, such that the codomain of the first is the same as the domain of the second, we can form their *composite* $g \circ f:A \longrightarrow C$; the definition of $g \circ f$ is:

$$(g \circ f)(a) \underset{\text{def}}{=} g(f(a)) \qquad\qquad (a \in A).$$

We may omit the circle in the notation of composition, and write simply gf. To see the domain/codomain relationships of the functions involved, we may draw the three functions f, g, and gf in the diagram

$$\begin{tikzcd} & B \arrow[dr, "g"] & \\ A \arrow[ur, "f"] \arrow[rr, "gf"'] & & C \end{tikzcd}$$

The composite of two functions is defined only if the codomain of one coincides with the domain of the other.

E.g., consider the functions

$$\begin{array}{rcl} f : \mathbb{N} & \longrightarrow & \mathcal{P}(\mathbb{N}) \\ n & \longmapsto & \{n\} \end{array} \qquad \text{and} \qquad \begin{array}{rcl} g : \mathcal{P}(\mathbb{N}) & \longrightarrow & \mathcal{P}(\mathbb{N}) \\ X & \longmapsto & \mathbb{N} - X \end{array}$$

Then, gf is the following function:

$$\begin{array}{rcl} gf : \mathbb{N} & \longrightarrow & \mathcal{P}(\mathbb{N}) \\ n & \longmapsto & \{x \in \mathbb{N} \mid x \neq n\} \end{array}$$

When, in the calculus, we talk about a function like $\sin(e^x)$, we have in mind a composite; in the case at hand, the composite $\sin \circ \exp$:

$$
\begin{array}{ccccc}
\mathbb{R} & \xrightarrow{\exp} & \mathbb{R} & \xrightarrow{\sin} & \mathbb{R} \\
x & \longmapsto & e^x & & \\
 & & y & \longmapsto & \sin(y) \\
x & \longmapsto & e^x & \longmapsto & \sin(e^x)
\end{array}
$$

The operation of composition of functions satisfies the *associative law*: in the situation

$$A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D\ ,$$

we have that

$$h(gf) = (hg)f\ .$$

Indeed, $h(gf)$ applied to any $a \in A$ gives

$$[h(gf)](a) = h([gf](a)) = h(g(f(a)))\ ;$$

and, $(hg)f$ applied to a gives

$$[(hg)f](a) = (hg)(f(a)) = h(g(f(a)))\ ,$$

which is the same value. Since the two functions, both from A to D, give the same value at each argument $a \in A$, they are equal.

With any set A, there is a particular function associated, namely the *identity function on* A:

$$
\begin{array}{rcl}
1_A : A & \longrightarrow & A \\
a & \longmapsto & a
\end{array}
$$

($1_A(a) = a$ for $a \in A$). This has the property that its composite with any function, provided it is well-defined, is the function itself:

$$B \xrightarrow{f} A \xrightarrow{1_A} A \quad :: \quad 1_A \circ f = f\ ,$$

$$A \xrightarrow{1_A} A \xrightarrow{g} C \; :: \; g \circ 1_A = g \, .$$

Another operation on functions is *restriction*. Suppose $f{:}A \longrightarrow B$ and $A' \subseteq A$. Then the *restriction of* f *to* A' is the function denoted as $f{\restriction}A' : A' \longrightarrow B$ for which $(f{\restriction}A')(a) = f(a)$ for all $a \in A'$. E.g., for the absolute-value function $|\text{-}| : \mathbb{Z} \longrightarrow \mathbb{N}$, its restriction to the subset $\mathbb{N}$ of its domain is $|\text{-}|{\restriction}\mathbb{N} = 1_{\mathbb{N}}$, the identity function on $\mathbb{N}$; the reason is that $|n| = n$ for all $n \in \mathbb{N}$.

With any subset A' of any set A, one can associate the *inclusion function* $\varphi{:}A' \longrightarrow A$, which acts like the identity: $\varphi(a) = a \; (a \in A')$; what makes it different from the identity function is that its domain and codomain are not (necessarily) equal. Note that, with the notation of this and the previous paragraph, $f{\restriction}A' = f \circ \varphi$.

A function $f{:}A \longrightarrow B$ is *injective*, or *one-to-one*, or f is an *injection*, if it maps distinct arguments to distinct values:

$$a \neq a' \implies f(a) \neq f(a') \quad \text{for any } a, \; a' \in A .$$

A more positive, but equivalent, way of putting the definition of injectivity is that

$$f(a) = f(a') \implies a = a' \quad \text{for any } a, \; a' \in A .$$

E.g., the exponential function $\exp{:}\mathbb{R} \longrightarrow \mathbb{R}$ is injective: if x and y are two distinct real numbers, then either $x < y$, or $y < x$; in the first case $e^x < e^y$ (the exponential function is strictly increasing), in the second case the other way around; thus, at any rate, $e^x \neq e^y$. But, the $\sin$ function is not injective: $0 \neq \pi$ but $\sin(0) = \sin(\pi) = 0$.

$f{:}A \longrightarrow B$ is *surjective*, or *onto*, or f is a *surjection*, if for any $b \in B$, there is at least one $a \in A$ such that $f(a) = b$. f is surjective just in case its range equals its codomain.

E.g., the range of $\sin{:}\mathbb{R} \longrightarrow \mathbb{R}$ is

$$[-1,1] \underset{\text{def}}{=} \{y \in \mathbb{R} \mid -1 \leq y \leq 1\} ;$$

thus, $\sin : \mathbb{R} \longrightarrow \mathbb{R}$ is not surjective (e.g., for $y = 2$, there is no x such that $\sin(x) = y = 2$). But, if we consider $\sin$ to be a function from $\mathbb{R}$ to the interval $[-1, 1]$, $\sin : \mathbb{R} \longrightarrow [-1,1]$, then $\sin$, in this sense, is surjective. (Note that for us, the information of the codomain is part of the data defining the function. Thus, $\sin : \mathbb{R} \longrightarrow \mathbb{R}$ and $\sin : \mathbb{R} \longrightarrow [-1,1]$ are, strictly speaking, not the same function.)

If $A \underset{f}{\overset{g}{\leftrightarrows}} B$, and $gf = 1_A$, we say that g is a *left inverse of* f, or that f is a *right inverse of* g. If $gf = 1_A$ and $fg = 1_B$ both hold, g is a *two-sided inverse*, or simply, an *inverse*, of f (and then, of course, f is an inverse of g).

Consider

$$\begin{array}{ccc} [\frac{k}{2}] & \longleftarrow\!\shortmid & k \\ \mathbb{N} & \underset{f}{\overset{g}{\leftrightarrows}} & \mathbb{N} \\ n & \longmapsto & 2n \end{array}$$

(here, $[\frac{k}{2}]$ denotes the largest integer not greater than $\frac{k}{2}$). Then $gf = 1_{\mathbb{N}}$, since

$$(gf)(n) = g(2n) = [\frac{2n}{2}] = [n] = n = 1_{\mathbb{N}}(n) .$$

However, $fg \neq 1_{\mathbb{N}}$; e.g., $(fg)(1) = 2[\frac{1}{2}] = 2 \cdot 0 = 0 \neq 1$. Thus, in this case, g is a left inverse of f, but it is not a right inverse of it.

We claim that in the situation:

$$A \underset{f}{\overset{g}{\leftrightarrows}} B, \text{ and } gf = 1_A ,$$

f is injective and g is surjective. Indeed, if $a, a' \in A$, and $f(a) = f(a')$, then

$$a \underset{gf=1_A}{\overset{\uparrow}{=}} g(f(a)) = g(f(a')) = a' ,$$

which shows the injectivity of f. On the other hand, if $a \in A$ is an arbitrary element of A, then for $b \underset{def}{=} f(a)$, we have $g(b) = g(f(a)) = a$ (again since $gf = 1_A$); this shows that g is surjective.

We have shown that

if a function (f in the previous situation*) has a left inverse, then it is injective, and if a function* (g above) *has a right inverse, it is surjective.*

The converses of the last two assertions are *almost* true. First,

if $f: A \longrightarrow B$ *is injective, and if* A *is not empty, then* f *has a left inverse*:

given any $b \in B$, define $g(b)$ to be $a \in A$ for which $f(a) = b$ if there is (necessarily at most) one such a; if however there is no such a, let $g(b)$ be any element in A (since A is not empty, there is at least one such). Then $(gf)(a) = g(f(a)) = a$ by the definition of g on $b = f(a)$; thus $gf = 1_A$.

Secondly,

if $g: B \longrightarrow A$ *is surjective, then it has a right inverse.*

Namely, we define $f: A \longrightarrow B$ in the following way. Given any $a \in A$, we pick an arbitrary $b \in B$ such that $g(b) = a$; by the assumption of g being surjective, there is certainly at least one such b; we make $f(a)$ equal this b. Then, with $f: A \longrightarrow B$ so defined, $(gf)(a) = g(b)$ for the b described above; but the choice of that b was such that $g(b) = a$; this shows that $(gf)(a) = a$ for any $a \in A$, which is to say that f is a right inverse of g. [In a foundational setting, this argument requires the so-called *Axiom of*

Choice.]

Returning to the previous assertion, the additional assumption of A being non-empty is necessary: *any* $f:\emptyset \longrightarrow B$ is injective, but there is a function $g:B \longrightarrow \emptyset$ at all only if B is also empty.

If a function is both injective and surjective, it is called *bijective*, or a *bijection*. Here are two examples for bijection:

$$\begin{aligned} f : \mathbb{Z} &\longrightarrow \mathbb{Z} \\ x &\longmapsto x - 1 \end{aligned}$$

$$\begin{aligned} g : \mathbb{Q} &\longrightarrow \mathbb{Q} \\ x &\longmapsto 2x \end{aligned}$$

The symbol $\cong$ is used to indicate a bijection: $f:A \xrightarrow{\cong} B$.

To say that a function is a bijection is the same as to say that it has an inverse.

Indeed, if it has a (two-sided) inverse, then, by what we said above, it is both injective and surjective. On the other hand, if $f:A \longrightarrow B$ is bijective, and for a moment, we assume that A is non-empty, then f has a left inverse $g:B \longrightarrow A$ and a right inverse $h:B \longrightarrow A$: $gf = 1_A$, $fh = 1_B$. But then

$$h = 1_A \circ h = (gf)h = g(fh) = g \circ 1_B = g ,$$

which shows that $h = g$ is a two-sided inverse of f . If A happens to be empty, then, with $f:A \longrightarrow B$ bijective, in particular, surjective, B must also be empty; in this case, $f = 1_\emptyset$, the "empty function", is a two-sided inverse of itself.

The last argument also shows that

the (two-sided) inverse, if exists, is uniquely determined:

if g and h are both inverses of f, then g is a left inverse, h is a right inverse, of f, and the calculation above shows that $g = h$. Moreover, we have also shown that

if f *has a left inverse* g, *and also a right inverse* h, *then* $g=h$, *and thus* f *has a two-sided inverse, namely* $g = h$.

The inverse of $f:A \longrightarrow B$, if exists, is denoted by f^{-1}. Thus, the defining properties of $f^{-1}:B \longrightarrow A$ are:

$$f \circ f^{-1} = 1_B \text{ and } f^{-1} \circ f = 1_A .$$

The composite of two injections (if well-defined) is an injection; the composite of two surjections is a surjection; the composite of two bijections is a bijection.

We leave the easy proofs to the reader.

A *permutation* of a set A is any bijection from A to A itself. The following denotes a permutation of the set $\{1, 2, 3, 4, 5\}$:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 5 & 3 & 1 \end{pmatrix} ;$$

(Recall this notation for a function from above, at (2): if σ is the name of the permutation at hand, then $\sigma(1)=4$, $\sigma(2)=2$, $\sigma(3)=5$, $\sigma(4)=3$, $\sigma(5)=1$.)

The set of all functions $A \longrightarrow B$ is denoted by the exponential notation B^A .

Sequences are particular functions. E.g., the 5-term sequence $\langle a_1, a_2, a_3, a_4, a_5 \rangle$ may be identified with the function whose domain is the set $\{1, 2, 3, 4, 5\}$, and whose value at i is a_i . The notation $\langle a_i \rangle_{1 \le i \le n}$ means the n-term sequence whose i^{th} term is a_i . A^n denotes the set of all n-term sequences of members of A . Thus,

$$A^n = \{\langle a_i \rangle_{1 \le i \le n} \mid a_i \in A \text{ for all } i < n\} .$$

Note that for $n=0$, for any set A , there is exactly one 0-term sequence of elements of A , the *empty sequence* $\perp$; $A^0 = \{\perp\}$.

If A is an *alphabet*, that is, a set of characters, then *strings* over A are essentially the same as finite sequences of elements of A ; strings of length n are the same as n-term sequences of elements of A . $A^* \underset{\text{def}}{=} \bigcup_{n \in \mathbb{N}} A^n$ is the set of all strings over A . Note that A^* always contains $\perp$, the empty string, as an element.

If, for instance, $A = \{a, b, c\}$, then the strings $aabccba$ and $bccb$ are members of A^* ; the first belongs to A^7 , the second to A^4 .

Infinite sequences are the same as functions with domain $\mathbb{N}$; $\mathbb{R}^{\mathbb{N}}$ is the set of all infinite sequences $\langle r_i \rangle_{i \in \mathbb{N}} = \langle r_0, r_1, \ldots, r_n, \ldots \rangle$ of reals. Infinite sequences of reals are important in the calculus.

Some more notation related to functions. Let $f : A \longrightarrow B$. If $X \subseteq A$, the *image of* X *under* f , denoted $f[X]$, is the set of all values of f while the argument of f ranges over X :

$$f[X] \underset{\text{def}}{=} \{f(a) \mid a \in X\} .$$

E.g., when

$$f : \mathbb{N} \longrightarrow \mathbb{N}$$
$$n \longmapsto 2n$$

and $X_1 = \{n \in \mathbb{N} \mid n \text{ is even}\}$, $X_2 = \{n \in \mathbb{N} \mid n \text{ is odd}\}$, then

$$f[X_1] = \{n \in \mathbb{N} \mid n \text{ is divisible by } 4\} ,$$

$$f[X_2] = \{n \in \mathbb{N} \mid n \text{ is even, but not divisible by } 4\} .$$

If $Y \subseteq B$, the *inverse image of* Y *under* f , $f^{-1}[Y]$ (warning: this notation does not imply that the inverse of f , f^{-1} , exists!) is the set of all $a \in A$ that are mapped into Y by f :

$$f^{-1}[Y] \underset{\mathrm{def}}{=} \{a \in A : f(a)\} .$$

E.g., with continuing the previous example, $f^{-1}[X_1] = \mathbb{N}$ and $f^{-1}[X_2] = \emptyset$.

In the general case $f : A \longrightarrow B$, let $b \in B$. Then $f^{-1}[\{b\}]$ is the set of those $a \in A$ whose f-image is b , $f(a) = b$. Thus, f is injective iff for all $b \in B$, $f^{-1}[\{b\}]$ has at most one element; f is surjective iff for all $b \in B$, $f^{-1}[\{b\}]$ has at least one element; and f is bijective iff for all $b \in B$, $f^{-1}[\{b\}]$ has exactly one element.