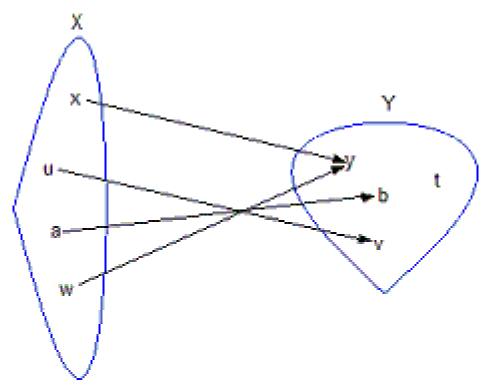
Functions

Definition (informal) Suppose X and Y are sets. A function from X to Y is a rule that associates to every element of X a unique element of Y. Sometimes a function is called a mapping.

For example, suppose $X = \{x, u, a, w\}$ and $Y = \{y, b, v, t\}$. The "pairings" set up by a function f are pictured below. This f pairs x to y, u to v, a to b and w to y. We write these facts as $y = f(x)$, $u = f(v)$, $b = f(a)$ and $y = f(w)$.



The (informal) definition of function from X to Y states that i) every element of X must be paired to something in Y: each point of X in the figure must be at the "tail" of some arrow. However, there might be "unpaired" elements (such as t) in Y.

The set X is called the domain of the function f and the set Y is called the codomain of f.
The set of all points in Y that are paired with an element of X is called the range of f. In the figure, range$(f) = \{y, b, v\}$. Always, the range is a subset of the codomain – but they might or might not be equal.

The statement that ii) every element in X must be associated to a unique element of Y means that, in the figure, there cannot be two arrows that begin at the same point. However, the definition does allow two arrows to end at the same point: for example $x \to y$ and $w \to y$ is OK.

Thinking more dynamically, we can imagine f as a machine that takes an "input" x, and creates an output "y." The domain, X, is the set of inputs that are allowed; the range of f is the set of all outputs actually produced by f using all the possible inputs from X. The codomain Y is simply the "universe" in which the outputs live.

The formal definition

In the preceding informal example, we know everything about the function f when we know all the pairs "(element of X, associated element from Y)" are. We can use a set of pairs to make the official definition of a function from X to Y.

Definition A set $f \subseteq X \times Y$ is called a function (or mapping) from X to Y if

i) $(\forall a \in X)\ (\exists b \in Y)\ \ (a, b) \in f$
ii) $(\forall a \in X)$ if $(a, b) \in f$ and $(a, c) \in f$, then $b = c$.

Here, i) and ii) are the formal version of i) and ii) in the earlier example.

Since any $f \subseteq X \times Y$ is a relation from X to Y, a function is a special kind of relation: one that satisfies i) and ii). If we followed the standard relation notation, we would write afb whenever $(a, b) \in f$. However, when f is a function, it's customary to write $f(a) = b$ instead.

Since f is a (special kind of) relation, we have already defined the domain and range:

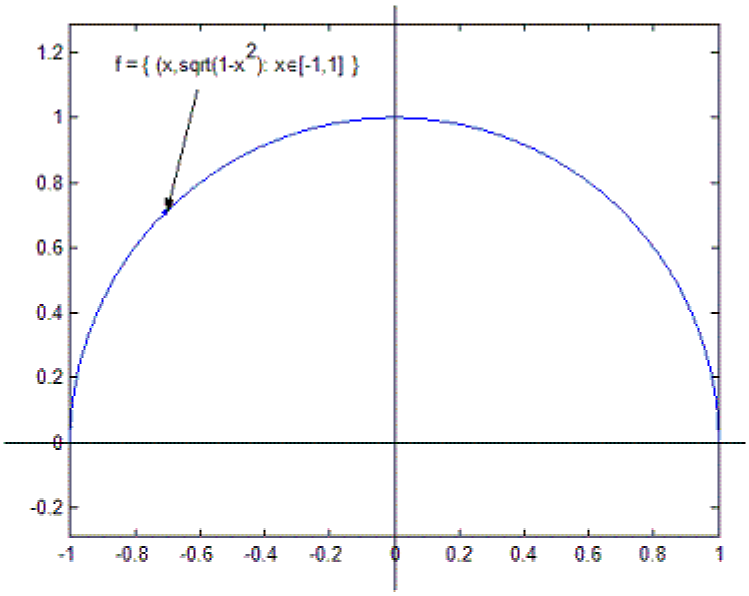
$$\text{domain}(f) = \underbrace{\{a \in X :\ \exists b \in Y\ (a, b) \in f\} = X}_{\uparrow \textit{ because of i)}}$$

and

$$\begin{aligned}\text{range}(f) &= \{b \in Y :\ (\exists a \in A)\ \ (a, b) \in f \\ &= \{b \in Y :\ (\exists a \in A)\ \ \ b = f(a)\}\end{aligned}$$

Example The function f in the earlier (informal) example is $f = \{(x, y),\ (u, v), (a, b),\ (w, y)\}$. From the set of ordered pairs we can read off domain$(f) =$ the set of all first coordinates $= \{x, u, a, w\}$. We can also describe the range : range$(f) =$ the set of all second coordinates $= \{y, v, b\}$. Notice that, using just the set f, we cannot tell whether there are additional points in the codomain. When a function is given as a set of ordered pairs, the domain and range are "built-in," but the codomain is a somewhat arbitrary choice: it could be any set containing the range.

Example Let $f = \{\, (x, \sqrt{1-x^2}\,) : x \in [1, 1]\, \}$. Domain$(f)$ is the set of all first coordinates of pairs in f : domain$(f) = [-1, 1]$. We can draw a picture of all the pairs (x, y) where $y = \sqrt{1-x^2}$:



More informally, we can write this function as $y = f(x) = \sqrt{1-x^2}$ where $-1 \leq x \leq 1$. The set of all second coordinates of pairs in f (the set of all "values" or "outputs" from f) is the range: range$(f) = [0, 1]$.

In calculus, one usually thinks of this function $y = f(x)$ given by some rule, and thinks of the set of points in the figure as the "graph of the function." In our formal definition, the set of points (the graph) IS (literally!) the function f.

Example For any function $f : X \rightarrow Y$, we can write down the inverse relation f^{-1} but f^{-1} might not be a function from Y to X.

In the preceding example, $f^{-1} = \{(y, x) : (x, y) \in f\}$

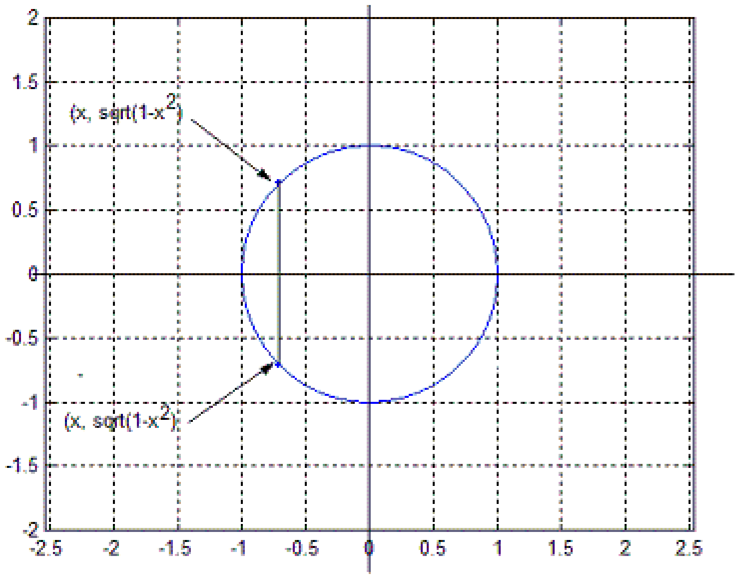
$= \{(y, x) : x \in [-1, 1]$ and $y = \sqrt{1-x^2}\,\}$ $=$ $\{(x, y) : y \in [-1, 1]$ and $x = \sqrt{1-y^2}\,\}$

$\uparrow$ *optional interchange of names x, y : handy in case you want to graph this relation in the same figure as f*)

$= \{(x, y) : y = \pm\sqrt{1-x^2}$ and $x \in [0, 1]\,\}$.

This relation f^{-1} is not a function since $(0, -1)$ and $(0, 1)$ are both in f^{-1} (*violating ii) in the definition of a function. Sometimes we describe the problem here by saying that $y = \pm\sqrt{1-x^2}$*) *is not single-valued, as a function needs to be.*)

Example The set $g = \{ (x, \pm \sqrt{1 - x^2}) : x \in [-1, 1] \}$ is a <u>relation</u> from $[-1, 1]$ to $\mathbb{R}$. We can picture g as:



g is <u>not</u> a function: there are pairs in g, such as $(0, 1)$ and $(0, -1)$ that violate condition ii) in the definition. (*How does this figure relate to f^{-1} is the preceding example?*)

When X and Y are sets of real numbers, we can picture a relation from X to Y. In such a picture, condition ii) in the definition of function means that "every vertical line intersects g is at most one point." When this is violated, the relation g is not a function.

<u>When are two functions f, g equal</u>? Since both f and g are sets of ordered pairs, $f = g$ just means that the sets are equal: $(a, b) \in f$ iff $(a, b) \in g$.

For example, consider the functions $f : \{ -1, 3\} \rightarrow \{ -1, 7\}$ given by $f(x) = 2x + 1$
$g : \{ -1, 3\} \rightarrow \{ -1, 7\}$ given by $g(x) = x^2 - 2$

The functions "look different" because the rules are different, but actually $f = g$:

$$f = \{(-1, -1), (3, 7)\} = g$$

Of course the functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ (given by the same rules) are <u>not</u> the same. When we change the domain, then, for example, $(0, 1) \in f$ but $(0, 1) \notin g$.

If $f = g$, then

i) $a \in$ domain(f) iff there is a pair $(a, b) \in f$ iff there is a pair $(a, b) \in g$ iff $a \in$ domain(g), and therefore domain$(f) =$ domain(g). Call the domain X.

ii) for every $a \in X$, $b = f(a)$ iff $(a, b) \in f$ iff $(a, b) \in g$ iff $b = g(a)$

So, if $f = g$ (viewed formally as sets of pairs), then f <u>and</u> g <u>have the same domain, and the functions have the same value at each point in the domain</u>. (*You should check that the converse is also true.*)

Example Here are examples of some other functions that you might have seen in Calculus III.

$f : \mathbb{R}^2 \to \mathbb{R}$ $\qquad$ $z = f(x, y) = x^2 - 2xy + y^2$
More formally,
$$f = \{(x, y, z) : \ z = x^2 - 2xy + y^2 \text{ and } x, y \in \mathbb{R}\} \subseteq \mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$$

$g : \mathbb{R} \to \mathbb{R}^3$ $\qquad$ $g(t) = (t^2, t^3, t^4) \in \mathbb{R}^3$
More formally,
$$g = \{(t, t^2, t^3, t^3) : t \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}^3 = \mathbb{R}^4$$
Here you can imagine t as "time." Then the points (t^2, t^3, t^4) in the range of g move along some curve in $\mathbb{R}^3$ as t varies.

$h : \ \mathbb{R}^2 \to \mathbb{R}^3$ $\qquad$ $h(x, y) = (2xy, x^2 - y^2, x^2 + y^2) \in \mathbb{R}^3$
More formally,
$$h = \{(x, y, 2xy, x^2 - y^2, x^2 + y^2) : x, y \in \mathbb{R}\} \subseteq \mathbb{R}^2 \times \mathbb{R}^3 = \mathbb{R}^5$$

For those who have had some linear algebra:

$T : \mathbb{R}^2 \to \mathbb{R}^3$ $\qquad$ $T(x, y) = \begin{bmatrix} 2 & 1 \\ 4 & 3 \\ 0 & -3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}.$

More formally, $T \ = \ ??$

Example Here are some examples of special functions that have names attached to them.

1. $f = \{(x, c) : \ x \in X\}$ is called a constant function: the second coordinate is the same in every pair in f. More informally, suppose $c \in Y$. Then $f : X \to Y$ is given by the formula $f(x) = c$ for every $x \in X$.

2. The function $i = \{(x, x) : x \in X\}$ is called the identity function on X. It can also be described as $i : X \to X$, where $i(x) = x$ for every $x \in X$.

3. If $f : X \to Y$ and $A \subseteq X$, we can define a new function $g : A \to Y$ by using the "restricted" formula $g(x) = f(x)$ for $x \in A$. We call g the restriction of f to A, denoted by $g = f|A$. $\quad$ f and g are, strictly speaking different functions because dom$(f) \neq$ dom(g) (when $A \neq X$).
In terms of the set definition of functions: $g|A = \{(x, y) \in f : \ x \in A\}$

For example, the restriction $g = \sin|\,[-\frac{\pi}{2}, \frac{\pi}{2}]$ refers to the function

$g(x) = \sin(x)$ with the domain restricted to $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

4. Suppose X is a set and $A \subseteq X$. Then $f : X \to \mathbb{R}$ defined by $f(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \in X - A \end{cases}$

f is called the characteristic function or indicator function of the set A (because the value of $f(x)$ "indicates" whether or not x is in A). More formally,

$$f = \{(x, y) : \ (x \in A \ \wedge \ y = 1) \ \vee \ (x \in X - A \ \wedge \ y = 0)\}$$

5. If $f : \mathbb{N} \to X$, then f is called a sequence in the set X. The terms of the sequence are $f(1),\ f(2),\ f(3),\ ...$ and we denote these values by $x_1 = f(1),\ x_2 = f(2),\ x_3 = f(3),$ $...,\ x_n = f(n),\ ...$ and sometimes refer to f as "the sequence (x_n)."

For example, the function $f(n) = \frac{1}{2^n}$ (with domain $\mathbb{N}$) is a sequence in $\mathbb{R}$. The terms of the sequence are $r_1 = f(1) = \frac{1}{2},\ r_2 = f(2) = \frac{1}{4},\ r_3 = f(3) = \frac{1}{8}, ...,\ r_n = \frac{1}{2^n},\ ...$
The function f might be referred to as the sequence (r_n) or the sequence $(\frac{1}{2^n})$.

Incidentally, a sequence is a function, that is, a certain set of ordered pairs. So here's another another mathematical concept, "sequence," that can be viewed as a set.

6. In a number system, the operations $+$ and $\cdot$ can be thought of as functions. For example, addition in $\mathbb{Z}$:

$$\begin{array}{lcl} (1,2) & \to & 3 \\ (3,-5) & \to & -2 \\ (-7,1) & \to & -6 \end{array}$$

Addition takes each pair in $\mathbb{Z} \times \mathbb{Z}$ and assigns to it number in $\mathbb{Z}$. In other words, we can think of $+$ as a function:

$$+ : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$$

Viewed as a function, "$+$" is a set of pairs with first coordinate from $\mathbb{Z} \times \mathbb{Z}$ and second coordinate from $\mathbb{Z}$:

$$+ = \{\ldots,\ ((1,2),3),\ ((3,-5),-2),\ ((-7,1),-6),\ ...\}$$

Similarly multiplication is a function

$$\cdot : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$$

So still more mathematical concepts (addition and multiplication) can be viewed as "really" being sets.

Definition Suppose $f : X \to Y$

i) If range$(f) = Y$, we say that f is a function <u>from</u> X <u>onto</u> Y, or simply that f <u>is</u> <u>onto</u> (*<u>provided</u> it's clear what set Y is meant. A function f <u>always</u> <u>onto the range</u> of f, but the range might not be the whole codomain Y.*)

Rephrased, f is a function from X <u>onto</u> Y if, for every $c \in Y$, there exists an $a \in X$ such that $f(a) = c$.

In some books, an onto function is also called a <u>surjection</u>.

ii) We say that f <u>is</u> <u>one-to-one</u> if, whenever $a, b \in$ dom(f) and $a \neq b$, then $f(a) \neq f(b)$. (equivalently, if $f(a) = f(b)$, then $a = b$).

In some books, an onto function is also called an <u>injection</u>.

(*CAUTION: <u>do not confuse the definition of one-to-one with its converse: the statement</u> "if $f(a) \neq f(b)$, then $a \neq b$" is a true statement about <u>every function</u>!*)

iii) If f is both one-to-one and onto, we call f a <u>bijection</u>.

Here is a more informal description:

<u>onto function</u>	every element in Y is a value (output) of the function
<u>one-to-one function</u>	different inputs to f always produce different outputs
<u>bijection</u>	the function f establishes a perfect pairing ("one-to-one correspondence") between all the elements in X and all the elements in Y. *Intuitively, a bijection from X to Y can exist iff "X and Y have the same number of elements."*

And here is a description worded in terms of functions as sets of ordered pairs:

i) because f is a function, each element $a \in X$ appears <u>once and only once</u> in a pair $(a, y) \in f$

ii) f is <u>onto</u> Y iff each $b \in Y$ appears <u>at least once</u> as the second coordinate of a pair $(x, b) \in f$

iii) f is <u>one-to-one</u> iff each $b \in Y$ occurs <u>at most once</u> as the second coordinate of a pair $(x, b) \in f$.

iv) (*Combining i), ii), and iii*) f is a bijection between X and Y iff each a in X occurs once and only once as a first coordinate of a pair in f <u>and</u> each $b \in Y$ occurs once and only once as the second coordinate of a pair in f.

You should compare the informal descriptions to the descriptions in terms of ordered pairs to be sure you see that they are saying the same thing,

Exercise: Write the negations of the definitions:

1) f is <u>not</u> one-to-one if ...
2) f is <u>not</u> onto if ...
3) f is <u>not</u> a bijection if ...

Examples (check the details where necessary)

1. The function sin : $\mathbb{R} \to \mathbb{R}$ is not one-to-one because, for example, $\sin(0) = \sin(\pi) = 0$. But the restriction $g = \sin|\,[-\frac{\pi}{2}, \frac{\pi}{2}]$ is a one-to-one function: if $x, y \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $x \neq y$, then $\sin(x) \neq \sin(y)$.

   Neither sin nor the restriction, g, is onto $\mathbb{R}$.

2. Let $f : \mathbb{N} \times \mathbb{N}, \to \mathbb{N}$ where $f(m, n) = 2^m 3^n$. The Fundamental Theorem of Arithmetic shows that f is one-to-one. (*Why? Suppose* $f(m, n) = f(k, l)$, *that is,* $2^m 3^n = 2^k 3^l$. *Then* ...)

   However, f is not onto $\mathbb{N}$ because $8 \in \mathbb{N}$ but 8 is not in the range of f (*why?*).

3. Let $X =$ the interval $(1, \infty) \subseteq \mathbb{R}$ and let $Y = \mathbb{R}$. Define $f : X \to Y$ using the formula $f(x) = \int_1^\infty \frac{1}{t^x}\, dt$. (This definition makes sense because (*see Calculus II*) the improper integral converges for each $x > 1$.)

   For example, $f(2) = \int_1^\infty \frac{1}{t^2}\, dt = 1$. Then f is one-to-one but not onto (*Why? Convince yourself using an informal argument about areas.*)

4. Define $f : \{2, 3, 4, ...\} \to \mathbb{N}$ by the rule $f(x) =$ "the least integer ≥ 2 which divides x." For example, $f(2) = 2$, $f(3) = 3$, $f(4) = 2$, $f(5) = 5$, $f(6) = 2$, and $f(9) = 3$.

   The function f is not one -to-one (because $f(2) = f(4)$ but $2 \neq 4$).

   What is the range of f?

5. Define $f : \mathbb{N} \to \mathbb{N}$ by $f(n) =$ the n^{th} prime number. For example $f(1) = 2$, $f(2) = 3$, $f(3) = 5$, $f(4) = 7$. The function f is clearly one-to-one but not onto $\mathbb{N}$.

6. Let $\mathbb{S}$ be the set of prime numbers and define $f : \mathbb{N} \to \mathbb{S}$ using the same rule as in Example 6. Then f is a bijection between $\mathbb{N}$ and $\mathbb{S}$.

   This illustrates that whether $f : X \to Y$ is onto depends on what we are considering the set Y to be. We cannot say whether a function *f* is "onto" without knowing what set Y is intended.

7. Define $\pi : \mathbb{R} \to \omega = \{0, 1, 2, 3, \ldots\}$ by $\pi(x) =$ "the number of primes $\leq x$". For example, $\pi(1) = 0, \pi(2) = 1,\ \pi(2.5) = 1,\ \pi(5.6) = 3$.
The function π is onto ω (why?) but not one-to-one.

This function is a part of the famous Prime Number Theorem which states that

$$\lim_{x \to \infty} \frac{\pi(x) \ln(x)}{x} = 1.$$

A proof of the Prime Number Theorem is quite difficult. However, a much simpler fact is that $\lim_{x \to \infty} \pi(x) = \infty$. *(Why is this simpler fact true?)*

8. Define $g : \mathbb{R} \to \mathbb{R}$ by using the formula $g(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$. This function is one-to-one but not onto $\mathbb{R}$. (*Why? Can you describe* $g(x)$ *in a different way?*)

9. Let A be a set and define a function $f : A \to \mathcal{P}(A)$ by $f(a) = \{a\}$ for each $a \in A$

For example : *If* $A = \{1, 2\}$, *then* $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$

$f(1) = \{1\} \in \mathcal{P}(A)$ *and* $f(2) = \{2\} \in \mathcal{P}(A)$

For any set A, f is one-to-one but not onto (*why?*)

10. Let A be a set and define $f : A \to \mathcal{P}(\mathcal{P}(A))$ by

$f(a) =$ "the set of all subsets of A containing a" $= \{B \in \mathcal{P}(A) : a \in B\}$.

For example : *If* $A = \{1, 2, 3\}$, *then* $f(2) = \{\{2\}, \{1, 2\}, \{2, 3\}, \{1, 2, 3\}\}$.

This function is one-to-one. To see this, notice that if $a \neq b$, then $\{a\} \in f(a)$ and $\{a\} \notin f(b)$, so $f(a) \neq f(b)$. Is the function onto $\mathcal{P}(\mathcal{P}(A))$.

The formal view of a function as a set of ordered pairs is often not necessary: the informal definition of a function as "a rule that ... " often is sufficient. But the formal definition does "match up" with our the informal view, and it shows precisely how a function fits into the point of view that "everything is a set." In addition, the set notation is sometimes handy as the following example shows.

Example Since functions f and g are sets f and g are functions, it makes sense to form the sets $f \cup g$ and $f \cap g$. This is occasionally handy just as a quick and clean way to say something.

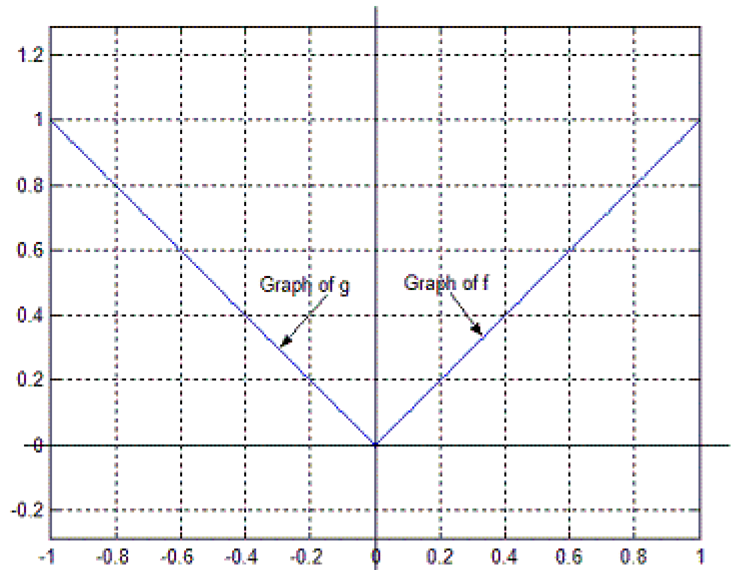
Suppose we have two functions defined informally by the rules

$f(x) = x$ for $x \geq 0$ (so $f : [0, \infty) \to \mathbb{R}$) and
$g(x) = -x$ for $x \leq 0$ (so $g : [-\infty, 0] \to \mathbb{R}$)

More formally, $f = \{(x,x) : x \in [0,\infty)\}$ and $g = \{(x.-x) : x \in (-\infty,0]\}$
Then $f \cup g$ is the set $\{(x,y) :\ x \in \mathbb{R}$ and $y = x$ if $x \geq 0$ and $y = -x$ if $x \leq 0\}$.

Therefore $h = f \cup g$ is the function with domain $\mathbb{R}$ where $h(x) = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x \leq 0 \end{cases}$.

In other words, $h(x) = |x|$. Using " $\cup$ " has simply "pasted together" these functions into a new one. The domain of $f \cup g$ is domain$(f) \cup$ domain(g)



Graph of h(x) = |x| is the union of the graphs of f and g

Example Suppose $f : \{a,b,c\} \to \{u,v,w\}$ where $f(a) = u,\ f(b) = u$ and $f(c) = w$
and $g : \{a,d,e\} \to \{u,s,t,\}$ where $g(a) = s,\ g(d) = t,\ g(e) = u$

The $h = f \cup g = \{(a,u), (b,u),\ (c,w),\ (a,s),\ (d,t),\ (e,u)\}$ is not a function because the pairs (a,u) and (a,s) are both in h (*so* $h(a) =$ *???; h is not "single-valued."*)

In general, $f \cup g$ is a function iff whenever $x \in$ domain$(f) \cap$ domain(g), then $f(x) = g(x)$, so that f and g do not "conflict" at a point x where both are defined, In the preceding example, $a \in$ domain$(f) \cap$ domain(g) but $f(a) \neq g(a)$; the two functions can't be "pieced together" into one new function.

If f and g are functions, then $f \cap g$ is always a function: its domain is domain$(f) \cap$ domain(g)

Definition For sets X and Y, we use the symbol Y^X to represent the set of all functions from X to Y.

Example 1) $\mathbb{R}^{\mathbb{R}}$ denotes the set of all real-valued functions with domain $\mathbb{R}$.

2) Suppose $n, m \in \mathbb{N}$. If X has n elements and Y has m elements, then Y^X has m^n elements. (Why? – if $f \in Y^X$, then for each $x \in X$, there must be a pair $(x, y) \in f$; there are n choices for x and m choices for y. How many ways could such a function be defined?)

3) In part 2), what happens if X or Y is empty, that is, if n or m is 0?

i) If $n = 0$, then $X = \emptyset$. Then exactly one function in Y^X – namely, the empty function $\emptyset$. (*This is true whether* $m = 0$ *or* $m \neq 0$.)

To see whether this assertion is true, we have to consider the definition: a function from X to Y is a subset of $X \times Y = \emptyset \times Y = \emptyset$, so the only possible candidate for a function from X to Y is $\emptyset$.

And $\emptyset$ does satisfy the definition for a function from X to Y:

a) each $x \in X = \emptyset$ is the first coordinate of a pair in the relation $\emptyset \subseteq X \times Y$
(*If you're skeptical, show me an* $x \in X$ *that violates this statement*.)

b) If (a, u) and $(a, v) \in \emptyset$, then $u = v$.
(*Again, show me a counterexample if you disagree.*)

So, in this case $Y^X = \{\emptyset\}$ and Y^X contains $m^0 = 1$ elements (functions).

ii) if $n \neq 0$ (so $X \neq \emptyset$) and $m = 0$, then there are no functions from X to Y. Just as in part i), the only possible candidate is "$\emptyset$" but it violates the definition of a function from X to Y: if $x \in X$, then x does not occur as the first coordinate of a pair in $\emptyset$.

So in this case $Y^X = \emptyset$ and Y^X contains $0^n = 0$ elements (functions).

Inverse Functions

For any function $f : X \to Y$ we can always form the inverse relation f^{-1}. We saw earlier that f^{-1} might not be a function from Y to X. Here are two simple examples showing what can go wrong.

Example i) Suppose $f : \underset{\substack{\| \\ X}}{\{a, b, c\}} \to \underset{\substack{\| \\ Y}}{\{u, v\}}$ where $f = \{(a, u), (b, u), (c, v)\}$

Then $f^{-1} = \{(u, a),\ (u, b),\ (v, c)\}$ is not a function at all because there are two pairs (u, a) and (u, b) in f^{-1} having the same first coordinate, u.

ii) Suppose $f : \underset{\substack{\| \\ X}}{\{a, b\}} \to \underset{\substack{\| \\ Y}}{\{u, v, w\}}$ where $f = \{(a, u), (b, v)\}$. Then

Then $f^{-1} = \{(u, a), (v, b)\}$ is a function, but its domain is $\{u, v\} \neq Y$. So f^{-1} is not a function from Y to X.

> In i), the issue is that the function f contains the pairs (a, u) and (b, u) : the problem for f^{-1} arises from the fact the f is not one-to-one.
>
> In ii), the issue is that there is not also a pair in f with w as second coordinate : the problem for f^{-1} arises from the fact that f is not onto.

If f is both one-to-one and onto, the both problems disappear.

Theorem Suppose $f : X \to Y$. f^{-1} is a function from $Y \to X$ iff f is a bijection.

Proof Suppose$f$ is a function from X to Y. Then we know that

> (*) each $x \in X$ occurs once and only once as the first coordinate of an ordered pair $(x, y) \in f$

Suppose f^{-1} is a function from Y to X. Then each $y \in Y$ occurs once and only once as the first coordinate of a pair $(y, x) \in f^{-1}$, and therefore once and only once as the second coordinate of a pair $(x, y) \in f$. This, together with (*), says that f is a bijection.

Suppose f is a bijection from X to Y. Then f is one-to-one and onto. So each $y \in Y$ occurs once and only once as the second coordinate of a pair $(x, y) \in f$, and therefore once and only once as the first coordinate of a pair $(y, x) \in f^{-1}$. There f^{-1} is a function from Y to X. •

Corollary If $f : X \to \text{range}(f)$, then f^{-1} is a function from range(f) to X iff f is one-to-one.

Proof In the theorem, let the codomain $Y = \text{range}(f)$. The f is automatically onto. So the theorem then says that f^{-1} is a function from range(f) to X iff f is one-to-one. •

We saw earlier how to form a composition $S \circ R$ when R relation from X to Y and S is a relation from Y to Z. So, if we have functions $f : X \to Y$ and $g : Y \to Z$, we can form a composite relation $g \circ f$ from X to Z.

Theorem If $f : X \to Y$ and $g : Y \to Z$ are functions, then $g \circ f : X \to Z$ is a function,

Proof a) For each $x \in X$ there is a $y \in Y$ and a $z \in Z$ such that $(x, y) \in f$ and $(y, z) \in g$. Then $(x, z) \in g \circ f$ so the relation $g \circ f$ has domain X.

b) If (x, z_1) and $(x, z_2) \in g \circ f$, then there are elements $y_1, y_2 \in Y$ such that

$(x, y_1) \in f$ and $(y_1, z_1) \in g$ and
$(x, y_2) \in f$ and $(y_2, z_2) \in g$

But since f is a function, this means that $y_1 = y_2$; so $(y_1, z_1) \in g$ and $(y_1, z_2) \in g$. Since g is a function, it must be that $z_1 = z_2$.

Parts a) and b) show that $g \circ f$ is a function from X to Z. •

<u>Notation</u>: Notice that if $(x, z) \in g \circ f$, then there is a $y \in Y$ such that $(x, y) \in f$ and $(y, z) \in g$. In the more standard, informal function notation: $(g \circ f)(x) = z = g(y) = g(f(x))$.

If $f : X \to X$ is a bijection, then $f^{-1} : X \to X$ is a function and we can consider the composite functions

$$f^{-1} \circ f : X \to X \qquad \text{and} \qquad f \circ f^{-1} : Y \to Y$$

For $x \in X$, let $y = f(x)$. Then $(x, y) \in f$ and $(y, x) \in f^{-1}$ so $(x, x) \in f^{-1} \circ f$. In standard function notation, $(f^{-1} \circ f)(x) = f^{-1}(f(x)) = x$.

Similarly, for every $y \in Y$, $(f \circ f^{-1})(y) = f(f^{-1}(y)) = y$.

The equations i) $(\forall x \in X)\ (f^{-1} \circ f)(x) = f^{-1}(f(x)) = x$
ii) $(\forall y \in Y)\ (f \circ f^{-1})(y) = f(f^{-1}(y)) = y$

display the interaction between f and f^{-1} : each "undoes" the effect of the other.

$f^{-1} \circ f : X \to X$ is the same as the identity mapping $i(x) = x$ on X.
$f \circ f^{-1} : Y \to Y$ is the same as the identity mapping $i(y) = y$ on Y.

Example The function $f : \mathbb{R} \to (0, \infty)$ given by $f(x) = e^x$ is a bijection. (*Why?*) The inverse function is called $\ln : (0, \infty) \to \mathbb{R}$. Then the equations i) and ii) say that

i) $(\forall x \in \mathbb{R})$ $\qquad \ln(e^x) = x,$ $\qquad$ and

ii) $(\forall y \in (0, \infty))$ $\qquad e^{\ln(y)} = y$

Example Let $f : [0, \infty) \to \mathbb{R}$ by given by $f(x) = \sqrt{x}$ and $g : \mathbb{R} \to [0, \infty)$ by $g(x) = x^2$.

Then $g(f(x)) = g(\sqrt{x}) = (\sqrt{x})^2 = x,$ but $g \neq f^{-1}$. The relation f^{-1} is not even a function $-$ according to the preceding theorem, it <u>cannot</u> be because f is not one-to-one.

If $y \in \mathbb{R}$ and $y < 0$, then $f(g(y)) = f(y^2) = \sqrt{y^2} = |y| \neq y$. The function $f \circ g$ <u>is not</u> the identity mapping on $Y = \mathbb{R}$.

Exercise Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = mx + b$.
f has a inverse $\Leftrightarrow$???. If f has an inverse, what is a formula for $f^{-1}(x)$?

Example (*for those who have had a course in linear algebra*) Suppose A is an $n \times n$ matrix. Define $T : \mathbb{R}^n \to \mathbb{R}^n$ by $T(x) = A \cdot x$. Then T is a <u>linear</u> function and T is invertible (= "has an inverse")

iff T is one-to-one and onto
iff the system of linear equations $A \cdot x = b$ has a unique solution for every choice of $b \in \mathbb{R}^n$
iff the matrix A is row equivalent to the identity matrix I_n
iff the matrix A is invertible
iff $\det A \neq 0$
iff (*several other equivalent conditions*)

Example The function $\sin : \mathbb{R} \to [-1, 1]$ is not one-to-one and therefore has no inverse. But the function $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1]$ is a bijection, and its inverse is often called arcsin.
So $\arcsin : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$.

For example, $\sin(\frac{\pi}{4}) = \frac{\sqrt{2}}{2}$ and $\arcsin(\frac{\sqrt{2}}{2}) = \frac{\pi}{4}$. Notice that $\arcsin(-1) = -\frac{\pi}{2}$, <u>not</u> $\frac{3\pi}{2}$, because the <u>values</u> of arcsin are in $[-\frac{\pi}{2}, \frac{\pi}{2}]$

Example Let $\mathbb{E} = \{2, 4, 6, 8, ...\}$. The function $f : \mathbb{N} \to \mathbb{E}$ given by $f(n) = 2n$ is a bijection. The inverse function is $f^{-1}(n) = \frac{n}{2}$.

Similarly, it is possible to find a bijection between $\mathbb{N}$ and, say, $\{5, 10, 15, 20, ...\}$: use $g(n) = 5n$.

A bijection between the sets $\mathbb{N}$ and $\{3, 4, 5, 6, ...\}$ is given by $h(n) = n + 2$; the inverse function is $h^{-1} : \{3, 4, 5, 6, ...\} \to \mathbb{N}$ given by $h^{-1}(n) = n - 2$.

The existence of a bijection between two <u>finite sets</u> clearly means that the sets have the same number of elements. The bijection sets up a one-to-one correspondence between the elements of the two sets. We might also consider saying that "two infinite sets have the same number of elements" iff there if a bijection between the two sets. If we use that terminology, then $\mathbb{N}$, $\mathbb{E}$, $\{5, 10, 15, 20, ...\}$ and $\{3, 4, 5, 6, 7, ...\}$ all have the same number of elements. For example, $g \circ h^{-1}$ is a bijection between $\{3, 4, 5, 6, ...\}$ and $\{5, 10, 15, 20, ...\}$: $(g \circ h^{-1})(n) = 5(n-2)$.

All this seems a little odd: because $\mathbb{E}$ is a <u>proper</u> subset of $\mathbb{N}$, to say that these sets have "the same number of elements" makes the whole equal to a part of itself. But that's only because we are comparing "size" in different senses. The "whole" $\mathbb{N}$ <u>is bigger</u> than $\mathbb{E}$ <u>in the sense</u> that $\mathbb{E}$ is a proper subset of $\mathbb{N}$; but they are "the same size" <u>in a different (but perfectly reasonable) sense</u>: that there is a bijection between them.

Here is a result that seems even odder.

Example There is a bijection $f : \mathbb{N} \to \mathbb{Q}^+$ ($=$ the set of <u>positive</u> rationals)

We will not actually give a formula for f, but we will show how to calculate $f(n)$ for every n. Imagine writing an "infinite matrix" that contains all the positive rationals. The first row contains will contain all the positive rationals that can be written with denominator 1, the second row will contain all that can be written with denominator 2, etc. The nth row contains all positive rationals that can be written with denominator $n \in \mathbb{N}$.

Although we can actually write down only a small piece of this "infinite matrix," it is clear that every positive rational occurs in it: for example, the rational $\frac{51}{314}$ appears in column 51 of row 314.

We then "move diagonally" through the matrix using the pattern indicated by the arrows to define $f : \mathbb{N} \to \mathbb{Q}^+$ as follows:

$$f(1) = \tfrac{1}{1},\ f(2) = \tfrac{2}{1},\ f(3) = \tfrac{1}{2},\ f(4) = \tfrac{1}{3},\ f(5) = \tfrac{3}{1},\ f(6) = \tfrac{4}{1},\ f(7) = \tfrac{3}{2},\ ...$$

In defining $f(5)$, we skip past $\frac{2}{2}$ because $\frac{2}{2} = \frac{1}{1}$: letting $f(5) = \frac{2}{2}$ would repeat a value of f unnecessarily and it would make the function f not one-to-one. In the same way, we skip over every rational we come to if it's a repeat of an earlier rational: for example, $f(10) = \frac{1}{5}$ but $f(11) = \frac{5}{1}$.

$$\begin{array}{cccccccccccccccc}
\frac{1}{1} & \to & \frac{2}{1} & & \frac{3}{1} & \to & \frac{4}{1} & & \frac{5}{1} & \to & \frac{6}{1} & & \frac{7}{1} & & \frac{8}{1} & \cdots\cdots \\
 & \swarrow & & \nearrow & & \swarrow & & & & & & & & & & \\
\frac{1}{2} & & \frac{2}{2} & & \frac{3}{2} & & \frac{4}{2} & & \frac{5}{2} & & \frac{6}{2} & & \frac{7}{2} & & \frac{8}{2} & \cdots\cdots \\
\downarrow & \nearrow & & \swarrow & & & & & & & & & & & & \\
\frac{1}{3} & & \frac{2}{3} & & \frac{3}{3} & & \frac{4}{3} & & \frac{5}{3} & & \frac{6}{3} & & \frac{7}{3} & & \frac{7}{3} & \cdots\cdots \\
 & \swarrow & & & & & & & & & & & & & & \\
\frac{1}{4} & & \frac{2}{4} & & \frac{3}{4} & & \frac{4}{4} & & \frac{5}{4} & & \frac{6}{4} & & \frac{7}{4} & & \frac{8}{4} & \cdots\cdots \\
\downarrow & & & & & & & & & & & & & & & \\
\frac{1}{5} & & \frac{2}{5} & & \frac{3}{5} & & \frac{4}{5} & & \frac{5}{5} & & \frac{6}{5} & & \frac{7}{5} & & \frac{8}{5} & \cdots\cdots \\
\vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots &
\end{array}$$

Although we don't have a formula for f, it is clear that we have clearly defined a function: for any n, we have enough information to compute $f(n)$ if we want to: we (or a computer) can compute $f(100)$ or $f(156732)$ if we want to make the effort.

The fact that we "skip repeats" as we calculate the values of f means that, by definition, f is one-to-one. And clearly every positive rational is in range(f). For example, with a little work, you could find n so that $f(n) = \frac{51}{314}$.

So, we can say that $\mathbb{N}$ and $\mathbb{Q}^+$ "have the same number of elements." Since both sets are infinite, you might say "no surprise." However, as we will see soon, it is not true that there is a bijection between every pair of infinite sets: for example, we will see later that no bijection exists between $\mathbb{N}$ and $\mathbb{R}$. In that sense, "$\mathbb{R}$ and $\mathbb{N}$ do not have the same number of elements" – infinite sets come in "different sizes" !

Images and Inverse Images (Preimages) of Sets

If $f\colon X \to Y$, then each element $a \in A$ has a image $f(a)$ in Y. If $b \in Y$, then b might have a preimage in X, that is, an element $a \in X$ for which $f(a) = b$. (*To say "every $b \in Y$ has at least one preimage in X" is equivalent to saying "f is onto."*)

If $A \subseteq X$, we can consider the set of all images of elements a in A; similarly, for a set $B \subseteq Y$, we can consider the set of all preimages of elements b in B.

Definition Suppose $f\colon X \to Y$.

i) If $A \subseteq X$, then

the image of the set $A = f[A] = \{y \in Y : y = f(a) \text{ for some } a \in A\}$.

By definition, $f[A] \subseteq Y$.

ii) If $B \subseteq Y$, then

the inverse image set of $B = f^{-1}[B] = \{x \in X : f(x) \in B\}$

By definition, $f^{-1}[B] \subseteq A$

Example Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$.

If $A =$ the interval $[-2, 1] \subseteq \mathbb{R}$, then $f[A] = \{y : y = f(a) \text{ for some } a \in [-2, 1]\}$
$= [0, 4]$

$f[\{2\}] = \{4\}$

$f[\,[-3, -2]\,] = [4, 9]$

If $B = [3, 4] \subseteq \mathbb{R} = \text{codomain}(f)$, then $f^{-1}[B] = \{x : f(x) \in [3, 4]\}$
$= [-2, -\sqrt{3}] \cup [\sqrt{3}, 2]$

$f^{-1}[\,[0, 9]\,] = [-3, 3]$

$f^{-1}[\{4\}] = \{-2, 2\}$

$f^{-1}[\,[-5, -4]\,] = \emptyset$

The function f is not onto since $f[\mathbb{R}] \neq \mathbb{R}$.

Notation: When B is a one point set such as $\{2\}$, we will often write $f^{-1}[2]$ or even $f^{-1}(2)$ rather than the more formal $f^{-1}[\{2\}]$.

Example Let $f : \mathbb{Z} \to \mathbb{Z}_5$ be the canonical map $f(x) = [x]$. Then

For $A = \{0, 5\} \subseteq \mathbb{Z},\ f[A] = \{f(0), f(5)\} = \{\ [0]\ \}$

For $A = \{0, 2, 5\} \subseteq \mathbb{Z},\ f[A] = \{f(0), f(2), f(5)\} = \{[0], [2]\ \}$

For $A = \{0,\ \pm 5,\ \pm 10,\ \pm 15, ...\} \subseteq \mathbb{Z},\ f[A] = \{\ [0]\ \}$

For $B = \{\ [0]\} \subseteq \mathbb{Z}_5,\ f^{-1}[B] = \{z \in \mathbb{Z} : f(z) = [0]\} = \{0,\ \pm 5,\ \pm 10,\ \pm 15, ...\}$

For $B = \{\ [0],\ [1]\} \subseteq \mathbb{Z}_5,\ f^{-1}[B] = \{z \in \mathbb{Z} :\ z \equiv_5 0$ or $z \equiv_5 1\}$

Example Suppose $f : \mathbb{N} \to \mathbb{R}$, that $A \subseteq \mathbb{N}$ and that $B \subseteq \mathbb{R}$. Be sure you understand each of the following (true) statements:

i) if $2 \in A$, then $f(2) \in f[A]$

ii) if $5 \in f^{-1}[B]$, then $f(5) \in B$

iii) if $f(2) \in B$. then $2 \in f^{-1}[B]$

iv) if $f(2) \in f[A]$, then 2 might or might not be in A:

a) Suppose f is a constant function: say $f(n) = 1$ for all $n \in \mathbb{N}$.
Let $A = \{2, 3, 4\}$. Then $f[A] = \{f(2), f(3), f(4)\} = \{1\}$,
so $1 = f(1) \in f[A]$ is true; but $1 \notin A$.

b) If $f(2) \in f[A]$ and f is one-to-one, then $2 \in A$.

Proof Suppose $f(2) = y \in f[A]$. By definition of $f[A]$, then there is an $a \in A$ for which $f(a) = y = f(2)$. If f is one-to-one, this implies that $2 = a \in A$. $\bullet$

The following theorem gives some handy properties of images and inverse images of sets.

Theorem Suppose $f \colon X \to Y$, that $A_s \subseteq X$ and $B_s \subseteq Y$ ($s \in$ some indexing set S)

1) $f[\emptyset] = \emptyset$ and $f[X] \subseteq Y$

1′) $f^{-1}[\emptyset] = \emptyset$ and $f^{-1}[Y] = X$

2) $A_1 \subseteq A_2$ implies $f[A_1] \subseteq f[A_2]$

2′) $B_1 \subseteq B_2$ implies $f^{-1}[B_1] \subseteq f^{-1}[B_2]$

3) $f[\bigcup_{s\in S} A_s] = \bigcup_{s\in S} f[A_s]$

3′) $f^{-1}[\bigcup_{s\in S} B_s] = \bigcup_{s\in S} f^{-1}[B_s]$

4) $f[\bigcap_{s\in S} A_s] \subseteq \bigcap_{s\in S} f[A_s]$

4′) $f^{-1}[\bigcap_{s\in S} B_s] = \bigcap_{s\in S} f^{-1}[B_s]$

Proof The proof of each part is fairly easy. As an example, we prove 4′).

$x \in f^{-1}[\bigcap_{s \in S} B_s]$ iff $f(x) \in \bigcap_{s \in S} B_s$ iff $f(x) \in B_s$ for every $s \in S$

iff $x \in f^{-1}[B_s]$ for every $s \in S$ iff $x \in \bigcap_{s \in S} f^{-1}[B_s]$

so, in 4′), the sets on the left and right have the same members (and therefore are equal). •

Example In part 4) of the theorem, "$\subseteq$" cannot be replaced by "$=$". For example, let $A_1 = \{x \in \mathbb{R}: x < 1\}$ and $A_2 = \{x \in \mathbb{R}: x > 1\}$ and let f be the constant function $f(x) = 1$. Then $f[A_1] = f[A_2] = \{1\}$, so $f[A_1] \cap f[A_2] = \{1\}$, whereas $f[A_1 \cap A_2] = f[\emptyset] = \emptyset$.