assume reader has 7+ attention points

Definitions

A *function* is a mathematical object that takes numbers as inputs and produces numbers as outputs. We use the notation

$$f\colon A \to B$$

to denote a function from the input set A to the output set B. In this book, we mostly study functions that take real numbers as inputs and give real numbers as outputs: $f\colon \mathbb{R} \to \mathbb{R}$.

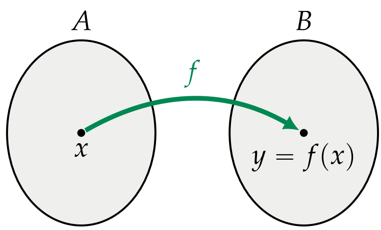

Figure 1.20: An abstract representation of a function f from the set A to the set B. The function f is the arrow which *maps* each input x in A to an output $f(x)$ in B. The output of the function $f(x)$ is also denoted y.

A function is not a number; rather, it is a *mapping* from numbers to numbers. We say "f maps x to $f(x)$." For any input x, the output value of f for that input is denoted $f(x)$, which is read as "f of x."

We'll now define some fancy technical terms used to describe the input and output sets of functions.

spend four points of f inputs and outputs

- A: the *source set* of the function describes the types of numbers that the function takes as inputs.
- $\text{Dom}(f)$: the *domain* of a function is the set of allowed input values for the function.
- B: the *target set* of a function describes the type of outputs the function has. The target set is sometimes called the *codomain*.
- $\text{Im}(f)$: the *image* of the function is the set of all possible output values of the function. The image is sometimes called the *range*.

See Figure 1.21 for an illustration of these concepts. The purpose of introducing all this math terminology is so we'll have words to distinguish the general types of inputs and outputs of the function (real numbers, complex numbers, vectors) from the specific properties of the function like its domain and image.

Let's look at an example to illustrate the difference between the source set and the domain of a function. Consider the square root

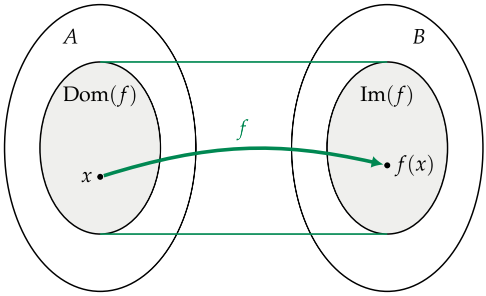

Figure 1.21: Illustration of the input and output sets of a function $f\colon A \to B$. The *source set* is denoted A and the *domain* is denoted $\text{Dom}(f)$. Note that the function's domain is a subset of its source set. The *target set* is denoted B and the *image* is denoted $\text{Im}(f)$. The image is a subset of the target set.

function $f\colon \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \sqrt{x}$, which is shown in Figure 1.22. The source set of f is the set of real numbers—yet only nonnegative real numbers are allowed as inputs, since $\sqrt{x}$ is not defined for negative numbers. Therefore, the domain of the square root function is only the nonnegative real numbers: $\text{Dom}(f) = \mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$. Knowing the domain of a function is essential to using the function correctly. In this case, whenever you use the square root function, you need to make sure that the inputs to the function are nonnegative numbers.

The complicated-looking expression between the curly brackets uses *set notation* to define the set of nonnegative numbers $\mathbb{R}_+$. In words, the expression $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$ states that "$\mathbb{R}_+$ is defined as the set of all real numbers x such that x is greater than or equal to zero." We'll discuss set notation in more detail in Section 1.23. For now, you can just remember that $\mathbb{R}_+$ represents the set of nonnegative real numbers.

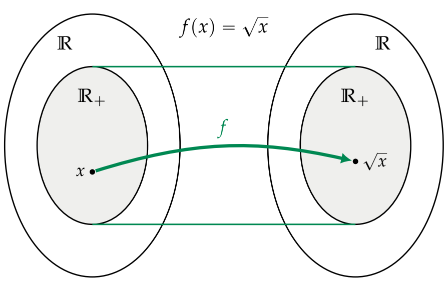

Figure 1.22: The input and output sets of the function $f(x) = \sqrt{x}$. The domain of f is the set of nonnegative real numbers $\mathbb{R}_+$ and its image is $\mathbb{R}_+$.

To illustrate the difference between the image of a function and its target set, let's look at the function $f(x) = x^2$ shown in Figure 1.23. The quadratic function is of the form $f\colon \mathbb{R} \to \mathbb{R}$. The function's source set is $\mathbb{R}$ (it takes real numbers as inputs) and its target set is $\mathbb{R}$ (the outputs are real numbers too); however, not all real numbers are possible outputs. The *image* of the function $f(x) = x^2$ consists only of the nonnegative real numbers $\mathbb{R}_+ = \{y \in \mathbb{R} \mid y \geqslant 0\}$, since $f(x) \geqslant 0$ for all x.

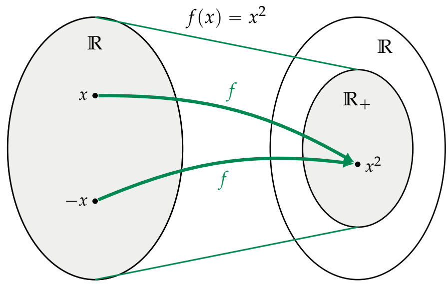

Figure 1.23: The function $f(x) = x^2$ is defined for all reals: $\mathrm{Dom}(f) = \mathbb{R}$. The image of the function is the set of nonnegative real numbers: $\mathrm{Im}(f) = \mathbb{R}_+$.

Function properties

We'll now introduce some additional terminology for describing three important function properties. Every function is a mapping from a source set to a target set, but what kind of mapping is it?

spend
- A function is *injective* if it maps two different inputs to two different outputs. If x_1 and x_2 are two input values that are not equal $x_1 \neq x_2$, then the output values of an injective function will also not be equal $f(x_1) \neq f(x_2)$.

three

more
- A function is *surjective* if its image is equal to its target set. For every output y in the target set of a surjective function, there is at least one input x in its domain such that $f(x) = y$.

here
- A function is *bijective* if it is both injective and surjective.

I know this seems like a lot of terminology to get acquainted with, but it's important to have names for these function properties. We'll need this terminology to give a precise definition of the *inverse function* in the next section.

Injective property We can think of *injective* functions as pipes that transport fluids between containers. Since fluids cannot be compressed, the "output container" must be at least as large as the "input container." If there are two distinct points x_1 and x_2 in the input container of an injective function, then there will be two distinct points

$f(x_1)$ and $f(x_2)$ in the output container of the function as well. In other words, injective functions don't smoosh things together.

In contrast, a function that doesn't have the injective property can map several different inputs to the same output value. The function $f(x) = x^2$ is not injective since it sends inputs x and $-x$ to the same output value $f(x) = f(-x) = x^2$, as illustrated in Figure 1.23.

The maps-distinct-inputs-to-distinct-outputs property of injective functions has an important consequence: given the output of an injective function y, there is only one input x such that $f(x) = y$. If a second input x' existed that also leads to the same output $f(x) = f(x') = y$, then the function f wouldn't be injective. For each of the outputs y of an injective function f, there is a *unique* input x such that $f(x) = y$. In other words, injective functions have a unique-input-for-each-output property.

Surjective property A function is *surjective* if its outputs cover the entire target set: every number in the target set is a possible output of the function for some input. For example, the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is surjective: for every number y in the target set $\mathbb{R}$, there is an input x, namely $x = \sqrt[3]{y}$, such that $f(x) = y$. The function $f(x) = x^3$ is surjective since its image is equal to its target set, $\text{Im}(f) = \mathbb{R}$, as shown in Figure 1.24.

On the other hand, the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by the equation $f(x) = x^2$ is not surjective since its image is only the nonnegative numbers $\mathbb{R}_+$ and not the whole set of real numbers (see Figure 1.23). The outputs of this function do not include the negative numbers of the target set, because there is no real number x that can be used as an input to obtain a negative output value.

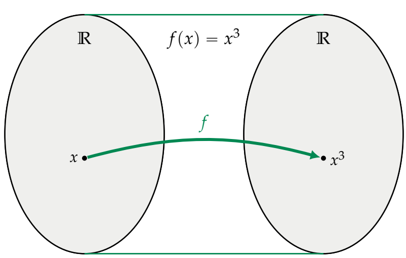

Figure 1.24: For the function $f(x) = x^3$ the image is equal to the target set of the function, $\text{Im}(f) = \mathbb{R}$, therefore the function f is surjective. The function f maps two different inputs $x_1 \neq x_2$ to two different outputs $f(x_1) \neq f(x_2)$, so f is injective. Since f is both injective and surjective, it is a *bijective* function.

Bijective property A function is bijective if it is both injective and surjective. When a function $f : A \to B$ has both the injective and surjective properties, it defines a *one-to-one correspondence* between the numbers of the source set A and the numbers of the target set B. This means for every input value x, there is exactly one corresponding output value y, and for every output value y, there is exactly one input value x such that $f(x) = y$. An example of a bijective function is the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ (see Figure 1.24). For every input x in the source set $\mathbb{R}$, the corresponding output y is given by $y = f(x) = x^3$. For every output value y in the target set $\mathbb{R}$, the corresponding input value x is given by $x = \sqrt[3]{y}$.

A function is not bijective if it lacks one of the required properties. Examples of non-bijective functions are $f(x) = \sqrt{x}$, which is not surjective and $f(x) = x^2$, which is neither injective nor surjective.

Counting solutions Another way to understand the injective, surjective, and bijective properties of functions is to think about the solutions to the equation $f(x) = b$, where b is a number in the target set B. The function f is injective if the equation $f(x) = b$ has *at most one* solution for every number b. The function f is surjective if the equation $f(x) = b$ has *at least one* solution for every number b. If the function f is bijective then it is both injective and surjective, which means the equation $f(x) = b$ has *exactly one* solution.

Inverse function

We used inverse functions repeatedly in previous chapters, each time describing the inverse function informally as an "undo" operation. Now that we have learned about bijective functions, we can give a the precise definition of the inverse function and explain some of the details we glossed over previously.

Recall that a *bijective* function $f : A \to B$ is a *one-to-one correspondence* between the numbers in the source set A and numbers in the target set B: for every output y, there is exactly one corresponding input value x such that $f(x) = y$. The *inverse function*, denoted f^{-1}, is the function that takes any output value y in the set B and finds the corresponding input value x that produced it $f^{-1}(y) = x$.

For every bijective function $f : A \to B$, there exists an inverse function $f^{-1} : B \to A$ that performs the *inverse mapping* of f. If we start from some x, apply f, and then apply f^{-1}, we'll arrive—full circle—back to the original input x:

$$f^{-1}(\, f(x)\,) = x.$$

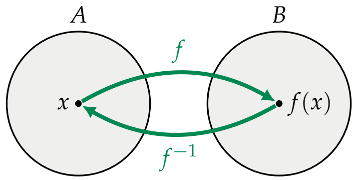

Figure 1.25: The inverse f^{-1} undoes the operation of the function f.

In Figure 1.25 the function f is represented as a forward arrow, and the inverse function f^{-1} is represented as a backward arrow that puts the value $f(x)$ back to the x it came from.

Similarly, we can start from any y in the set B and apply f^{-1} followed by f to get back to the original y we started from:

$$f(\, f^{-1}(y)\,) = y.$$

In words, this equation tells us that f is the "undo" operation for the function f^{-1}, the same way f^{-1} is the "undo" operation for f.

If a function is missing the injective property or the surjective property then it isn't bijective and it doesn't have an inverse. Without the injective property, there could be two inputs x and x' that both produce the same output $f(x) = f(x') = y$. In this case, computing $f^{-1}(y)$ would be impossible since we don't know which of the two possible inputs x or x' was used to produce the output y. Without the surjective property, there could be some output y' in B for which the inverse function f^{-1} is not defined, so the equation $f(f^{-1}(y)) = y$ would not hold for all y in B. The inverse function f^{-1} exists only when the function f is bijective.

Wait a minute! We know the function $f(x) = x^2$ is not bijective and therefore doesn't have an inverse, but we've repeatedly used the square root function as an inverse function for $f(x) = x^2$. What's going on here? Are we using a double standard like a politician that espouses one set of rules publicly, but follows a different set of rules in their private dealings? Is mathematics corrupt?

Don't worry, mathematics is not corrupt—it's all legit. We can use inverses for non-bijective functions by imposing *restrictions* on the source and target sets. The function $f(x) = x^2$ is not bijective when defined as a function $f : \mathbb{R} \to \mathbb{R}$, but it *is* bijective if we define it as a function from the set of nonnegative numbers to the set of nonnegative numbers, $f : \mathbb{R}_+ \to \mathbb{R}_+$. Restricting the source set to $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geqslant 0\}$ makes the function injective, and restricting the target set to $\mathbb{R}_+$ also makes the function surjective. The function $f : \mathbb{R}_+ \to \mathbb{R}_+$ defined by the equation $f(x) = x^2$ is bijective and its inverse is $f^{-1}(y) = \sqrt{y}$.

It's important to keep track the of restrictions on the source set we applied when solving equations. For example, solving the equation $x^2 = c$ by restricting the solution space to nonnegative numbers will give us only the positive solution $x = \sqrt{c}$. We have to manually add the negative solution $x = -\sqrt{c}$ in order to obtain the complete solutions: $x = \sqrt{c}$ or $x = -\sqrt{c}$, which is usually written $x = \pm\sqrt{c}$. The possibility of multiple solutions is present whenever we solve equations involving non-injective functions.

functions vs. equations

Discussion

To describe a function we specify its source and target sets $f\colon A \to B$, then give an equation of the form $f(x) =$ "expression involving x" that defines the function. Since functions are defined using equations, does this mean that functions and equations are the same thing? Let's take a closer look.

In general, any equation containing two variables describes a *relation* between these variables. For example, the equation $x - 3 = y - 4$ describes a relation between the variables x and y. We can isolate the variable y in this equation to obtain $y = x + 1$ and thus find the value of y when the value of x is given. We can also isolate x to obtain $x = y - 1$ and use this equation to find x when the value of y is given. In the context of an equation, the relationship between the variables x and y is symmetrical and no special significance is attached to either of the two variables.

We also can describe the same relationship between x and y as a function $f : \mathbb{R} \to \mathbb{R}$. We choose to identify x as the input variable and y as the output variable of the function f. Having identified y with the output variable, we can interpret the equation $y = x + 1$ as the definition of the function $f(x) = x + 1$.

Note that the equation $x - 3 = y - 4$ and the function $f(x) = x + 1$ describe the same relationship between the variables x and y. For example, if we set the value $x = 5$ we can find the value of y by solving the equation $5 - 3 = y - 4$ to obtain $y = 6$, or by computing the output of the function $f(x)$ for the input $x = 5$, which gives us the same answer $f(5) = 6$. In both cases we arrive at the same answer, but modelling the relationship between x and y as a function allows us to use the whole functions toolbox, like function composition and function inverses.