1. Introduction

The concept of a function is considered by many mathematicians to be the most important idea in all of Mathematics. Functions are introduced at the high school level or earlier, but in most cases the definition given is that a function associates with each x (in some set called the domain) a single y value. The association is usually defined by a formula like

$$f(x) \;=\; x^2 - 3x + 2$$

While this definition captures the essentials and may be adequate for introductory courses, it is not sufficiently general for more advanced study. It is also not very rigorous, because it is rather vague about the exact nature of the association between x and y, and also about what the possible values of x (and y) are.

Some functions are not easily represented by a closed formula. For example, suppose we define $f : \mathbb{R} \to \{0, 1\}$ by

$$f(x) \;=\; \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{otherwise} \end{cases}$$

or $f : \mathbb{R} \to \mathbb{R}$ by

$$f(x) \;=\; \sum_{n=1}^{\infty} \left(a_n \cos n\pi x + b_n \sin n\pi x\right), \quad n = 1, 2, \ldots$$

In some applications (probability theory), we want to associate a real number in the interval $[0, 1]$ with an element of the power set of the possible outcomes of some random experiment. This sounds like a function, but the elements of the domain of this function will be *sets*, not numbers.

It would be beneficial to have a definition of a function that is flexible enough to handle this kind of application, but still precisely specifies the nature of the association between the elements of the domain and range.

In the sequel, we will examine an alternative definition of a function that meets these requirements. This definition will allow us to define functions on arbitrary sets, not just numbers, and will make the exact nature of the association between elements of the domain and range crystal clear.

2. Cartesian Products

Before we can give a rigorous definition of a function, we need some preliminaries.

Definition 1 (Cartesian product). If A and B are sets, the **Cartesian product** of A and B, denoted by $A \times B$, is the set of all *ordered pairs* (a, b) where $a \in A$ and $b \in B$.

In *set builder notation*, the Cartesian product is:

$$A \times B \;=\; \{(a, b) \mid a \in A \quad \text{and} \quad b \in B\}$$

The Cartesian product is a simple but extremely useful idea.

Example 1. Suppose $A = \{1, 2, 3\}$ and $B = \{4, 5\}$. Then

$$A \times B \;=\; \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$$

Note that A and B are sets, and $A \times B$ is also a set.

Example 2. The elements of A and B do not have to be numbers: Let $A = \{\text{sun}, \text{moon}\}$ and $B = \{\text{dog}, \text{cat}\}$. Then

$$A \times B \;=\; \{(\text{sun}, \text{dog}), (\text{sun}, \text{cat}), (\text{moon}, \text{dog}), (\text{moon}, \text{cat})\}$$

Example 3. Here is a pairing of outcomes of a coin toss and real numbers: Let $A = \{\text{heads}, \text{tails}\}$ and $B = \{0.5\}$. Then

$$A \times B \;=\; \{(\text{heads}, 0.5), (\text{tails}, 0.5)\}$$

Example 4. The Cartesian product of $\mathbb{R}$ with itself is usually described as the (Cartesian) *coordinate plane*,

$$A \times B \;=\; \mathbb{R} \times \mathbb{R} \;=\; \{(x, y) \mid x, y \in \mathbb{R}\}$$

Example 5. The Cartesian product of $\mathbb{R}$ with the closed interval $[0, 2\pi]$ can be described as the *polar coordinate plane*,

$$A \times B \;=\; \mathbb{R} \times [0, 2\pi] \;=\; \{(r, \theta) \mid r \in \mathbb{R},\ \theta \in [0, 2\pi]\}$$

Example 6. The Cartesian product definition is easily extended to three sets. Suppose A, B, C are sets. Define

$$A \times B \times C \; = \; \{(x, y, z) \mid x \in A, \; y \in B, \; z \in C\}$$

Strictly speaking, the ordered triple (x, y, z) is not the same as the ordered pair $((x, y), z)$ which would belong to

$$(A \times B) \times C \; = \; \{((x, y), z) \mid (x, y) \in A \times B, \; z \in C\}$$

but if we agree to identify

$$((x, y), z) \text{ with } (x, y, z)$$

the difference between $A \times B \times C$ and $(A \times B) \times C$ will usually not be important and is often ignored.

Example 7. The Cartesian coordinate system in three dimensions, usually denoted by $\mathbb{R}^3$, is

$$\mathbb{R} \times \mathbb{R} \times \mathbb{R} \; = \; \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$$

Note that the exponent of $\mathbb{R}$ refers to the number of factors in the Cartesian product, not to multiplication of numbers.

Example 8. The cylindrical coordinate system in three dimensions can be thought of as:

$$\mathbb{R} \times \mathbb{R} \times [0, 2\pi] \; = \; \{(r, h, \theta) \mid r, h \in \mathbb{R}, \; \theta \in [0, 2\pi]\}$$

Example 9. The spherical coordinate system in three dimensions can be thought of as:

$$\mathbb{R} \times [0, 2\pi] \times [0, 2\pi] \; = \; \{(r, \theta, \phi) \mid r \in \mathbb{R}, \; \theta, \phi \in [0, 2\pi]\}$$

3. Functions

3.1. **Definition of a Function.**

Definition 2 (function). Let A and B be any two sets. A **function** f from A into B, denoted by $f : A \to B$, is defined to be a subset of

$$A \times B \; = \; \{(x, y) \mid x \in A, \; y \in B\}$$

with the property that *every element of* A *belongs to exactly one ordered pair in* $A \times B$.

A function $f : A \to B$ is also referred to as a **mapping** from A to B.

Note that $A \times B$ is a set of ordered pairs (x, y) and f is a subset of $A \times B$, so it is entirely proper to write

$$(x, y) \in f$$

to indicate that f associates the element y of B with the element x of A. In the sequel we will do so when it suits our purpose. In practice this notation is seldom used, and instead we write

$$f(x) \;=\; y$$

You should understand and be comfortable with the fact that the precise meaning of this notation is that the ordered pair (x, y) belongs to the subset of $A \times B$ that defines the function f.

3.2. **Images.**

Definition 3 (image of x under f)*.* If f is a function and x is an element of the domain of f, we say that y **is the image of** x **under** f if the ordered pair (x, y) belongs to the subset of $A \times B$ that defines f:

$$y \text{ is the image under } f \text{ of } x \Leftrightarrow (x, y) \in f$$

In the usual notation, we say that y is the image of x under f whenever

$$f(x) \;=\; y$$

We can extend the idea of the image under f to subsets of the domain of f:

Definition 4 (image of S under f)*.* If $f : A \to B$ is a function and $S \subset A$ is a subset of the domain of f, we define the **image of** S **under** f, denoted by $f[S]$, to be the set of all $y \in B$ for which there exists an ordered pair (x, y) in f whose first entry x belongs to S:

$$y \in f[S] \;\Leftrightarrow\; \exists x \in S \text{ s.t. } (x, y) \in f$$

In the usual notation,

$$y \in f[S] \quad \text{iff} \quad f(x) = y \text{ for some } x \in S$$

Definition 5 (range of a function). If $f : A \to B$ is a function, the image of the domain A under f is called the **range** of f:

$$\text{range}(f) \;=\; f[A] \;=\; \{y \in B \mid \exists\, x \in A \text{ s.t. } (x, y) \in f\}$$

The range of f may or may not be all of B. Some authors refer to the set B as the *codomain* of f, but others do not give it a name at all.

Example 10. If $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = x^2$$

and $S = [-3, 3]$, then

$$f[S] \;=\; [0, 9]$$

Example 11. If $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = |x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$$

then the range of f is

$$\text{range}(f) \;=\; f[\mathbb{R}] \;=\; [0, \infty)$$

3.3. Inverse Images.

Definition 6 (inverse image). If $f : A \to B$ is a function and $S \subset B$ is a subset of the codomain B, the **inverse image of** S, denoted by $f^{-1}[S]$, is the set of all elements in the domain A whose image belongs to S:

$$f^{-1}[S] \;=\; \{x \in A \,|\, f(x) \in S\}$$

Example 12. Suppose $f : \mathbb{R} \to \mathbb{R}$ is defined by the rule of assignment

$$f(x) = x^2$$

and $S = \{1, 2, 3\}$. Then the inverse image $f^{-1}[S]$ is

$$f^{-1}[S] \;=\; \{1, -1, \sqrt{2}, -\sqrt{2}, \sqrt{3}, -\sqrt{3}\}$$

Example 13. Suppose $f : \mathbb{R} \to \mathbb{R}$ is defined by the rule of assignment

$$f(x) = \sin x$$

and $S = \{1, -1\}$. Then the inverse image $f^{-1}[S]$ is:

$$f^{-1}[S] \;=\; \left\{ \ldots, \frac{-5\pi}{2}, \frac{-3\pi}{2}, \frac{-\pi}{2}, \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \ldots \right\}$$

The inverse image under a function $f : A \to B$ should not be confused with the function inverse, which is defined as the function $f^{-1}(y)$ with the property that

$$x = f^{-1}(y) \quad \text{if} \quad y = f(x) \quad \forall y \in \text{range}(f)$$

If you think in terms of our new definition, $f : A \to B$ is a subset of $A \times B$,

$$f \;=\; \{(x, y) \mid x \in A,\; y \in B\}$$

with the restriction that every element of A appears as the first element of *exactly one* ordered pair.

Consider what would happen if we took each of the ordered pairs in f and interchanged x and y.

We would have a new set of ordered pairs

$$\{(y, x) \mid y \in B,\; x \in A\}$$

Would this new set of ordered pairs represent a function $g : B \to A$?

In most cases it would not, because in order to qualify as a function, there can be no repetition in the first entries of the ordered pairs (y, x). Nothing in the definition of a function prohibits repetition in the *second* entries; it's entirely possible that there are multiple ordered pairs in f with identical values of y.

Think of $y = x^2$. Every positive real number y is the image of exactly two real numbers, $x = \sqrt{y}$ and $x = -\sqrt{y}$, so there are exactly two ordered pairs in f whose second element is y:

$$(\sqrt{y}, y) \in f \quad \text{and} \quad (-\sqrt{y}, y) \in f) \quad \forall y > 0$$

When we interchange x and y, the resulting set of ordered pairs will contain both

$$(y, \sqrt{y}) \quad \text{and} \quad (y, -\sqrt{y})$$

so our set will have two elements with the same first element. This violates the definition of a function, so this subset of $B \times A$ does not qualify as a function $g : B \to A$.

However, in the special case where there *are no repetitions* in y values, interchanging the elements of each ordered pair in f will actually produce a function, and that function will be the familiar function inverse of f.

Regardless of whether the function inverse exists or not, the inverse image of any subset of the codomain B **always** exists (though it may be the empty set). This is an important distinction between the inverse image of a set $f^{-1}[S]$ and the function inverse $f^{-1}(y)$.

3.4. **1-1 and Onto Functions.**

Definition 7 (onto function). The function $f : A \to B$ is said to be **onto** if the range of f is all of B. That is,

$$f \text{ is onto } \Leftrightarrow \forall\, y \in B,\ \exists\, x \in A \quad \text{s.t.} \quad (x, y) \in f$$

In the usual notation, the above definition could be stated as

$$f \text{ is onto } \Leftrightarrow \forall\, y \in B,\ \exists\, x \in A \quad \text{s.t.} \quad f(x) = y$$

Definition 8 (1-1 function). The function $f : A \to B$ is said to be **one to one** (denoted by 1-1) if

$$f(a) = f(b) \Rightarrow a = b \quad \forall a, b \in A$$

An equivalent statement would be that if $f : A \to B$ is $1 - 1$ and $y \in \text{range}(f)$ is any element of the range of f, then the inverse image of y is a singleton (a set with exactly one element).

If a function $f : A \to B$ is not $1-1$, then the function inverse $f^{-1}(y)$ does not exist.

If a function $f : A \to B$ is $1 - 1$, then the inverse function

$$f^{-1} : f[A] \to A$$

exists (recall that $f[A]$ is the range of f, the image under f of the domain A).

If $f : A \to B$ is both $1-1$ and onto, then the inverse function

$$f^{-1} : B \to A$$

exists (for an onto function, $f[A] = B$).

3.5. **Restrictions and Extensions.** Since $f : A \to B$ is defined as a subset of $A \times B$, we can treat f as a set. Suppose $C \subset A$ is a subset of the domain of f. In this case, consider the subset of $A \times B$ defined by:

$$g \;=\; \{(x, y) \mid (x, y) \in f \quad \text{and} \quad x \in C\}$$

A bit of thought should convince you that $g : C \to B$ is a function in its own right. In this situation g is called the **restriction** of f to C, and f is called the **extension** of g to f.