Chapter 1

Functions and Continuity

1.1 Introduction

How do we think of functions? Newton and Leibniz used "functions" but not in the sense that we now use them. Cauchy, in 1821, gave a definition of *function* that made the dependence between variables central to the function concept. Despite the generality of Cauchy's definition, he was still thinking of a function in terms of a formula. The problem here is that the function is tied up with its particular representation. What this means is that there is still the question of whether there is a difference in the function, say e^x, and its Taylor series representation. Do these both represent the same function? Cauchy gave an early example when he noted that

$$f(x) = \begin{cases} e^{-1/x^2} & \text{for} x \neq 0, \\ 0 & x = 0, \end{cases}$$

is a continuous function which has all its derivatives at 0 equal to 0. It therefore has a Taylor series which converges everywhere but only equals the function at 0. It was not until 1923 when Goursat gave the definition which appears in most textbooks today: *One says that y is a function of x if to a value of x corresponds a value of y. One indicates this correspondence by the equation $y = f(x)$.*

1.2 Modern Definitions

As you can see, the concept of function has been a long time in the making. The definition of functions that we give in our high school classrooms have been through many, many changes. We are offering the students the final product of centuries of thought. We offer a mathematically exact, precise definition. The use of other representations can (and will) be used to try to understand what the definitions mean and why we have chosen these definitions.

From what we have discussed above, the idea of a function is to express a relationship between the elements of two sets. If A and B are sets, then a *function from A to B* is often described as a *rule* or *process* that associates each element of A with one and only one element of B.

Definition 1.1 *A* function *is a rule that assigns to each element of a set A a unique element of a set B, where B may or may not equal A.*

The set A is called the **domain** of the function f, the set B the **codomain**, and the subset of B consisting of those elements that are images under the function f of some element of its domain is called the **range** of the function f. If f associates a in A to b in B, then the element b is called the **image of a under f** or **the value of f at a**, and a is called the **preimage of b under f**.

There are a number of notations for functions in use in mathematics. The common notations are when f associates a with b, then the **functional notation** or $f(x)$ notation is written ***f(a)=b***. The **arrow** or **mapping** notation is written $f\colon a \to b$.

One positive aspect of the arrow notation is that it conveys the idea that there is an action that associates the elements from A to the corresponding elements of B. This can be written as $f\colon A \to B$ only to indicate the domain and codomain, in which case the notation for elements is $f\colon a \mapsto b$. When the arrow notation is used, we will say that the function f **maps** the element a to b and we call f a **mapping** or **map** from A to B. We say that f maps A **onto** B if every element of B is in the range; i.e., $f(A) = B$.

The value in the domain of a function is called an **argument** of f. Then, the variable that we use to stand for the argument is called the **independent variable**. The variable that stands for the value of the function f is called the **dependent variable**. These are also referred to as *input* and *output* variables.

Example 1.1 The rule that assigns to each number the square of that number. Here we can express the function as a formula, either $y = x^2$ or $f(x) = x^2$.

For the function $f\colon x \to y$ many authors consistently use the single letter f to name the function and distinguish this from the symbol $f(x)$ used to identify the values of the function. But more broadly in mathematics this distinction is not made and $f(x)$ my stand for a function and also its values. Using the symbol $f(x)$ to stand for a function allows the independent variable to be explicitly identified.

There might be a "rule" to describe a function, but that might not be easily discovered or written. In the case of a correspondence where the idea of a function as a rule we need a better definition. This is done in the language of sets.

Definition 1.2 *For any sets A and B a function f from A to B, $f\colon A \to B$ is a subset f of the Cartesian product $A \times B$ such that every $a \in A$ appears once and only once as the first element of an ordered pair $(a, b) \in f$.*

This characterization of function now allows us to associate a *graph* with a function. Notice also that this is a very precise definition, but it is removed from the concept of a function "doing something". This is a more static definition and does not give us the feeling that the function is moving or mapping. Notice, though, that this definition is independent of any *a priori* knowledge of the sets A and B. This is the most general definition of function, and so is easily generalized to other settings.

1.3 Properties of Functions

Definition 1.3 *A function, $f\colon A \to B$ is a* one-to-one *function if and only if every element $b \in B$ is the image of at most one element $a \in A$. Symbolically, this can be expressed as f is one-to-one if and only if for all $x_1, x_2 \in A$, $f(x_1) = f(x_2)$ implies $x_1 = x_2$.*

Definition 1.4 *If $f = \{(x, y) \mid y = f(x)\}$ and f is one-to-one, then the function $\{(y, x) \mid (x, y) \in f\}$ is called the* inverse *of f and denoted by f^{-1}.*

Lemma 1.1 *If $f\colon A \to B$ is a one-to-one function with range $f(A)$, then*

$$f^{-1} = \{(y, x) \in f(A) \times A \mid (x, y) \in f\}$$

is a one-to-one function with domain $f(A)$ and range A.

Definition 1.5 *If $f\colon A \to B$ and $g\colon B \to C$ are functions, then the* composite function *$(g \circ f)\colon A \to C$ is the subset $g \circ f \subset A \times C$ defined as follows:*

$$g \circ f = \{(a, g(f(a))) \in A \times C \mid a \in A\}.$$

If $f\colon X \to Y$ and $U \subset X$ is a subset of X, then the set

$$f_U = \{(x, y) \mid x \in U\}$$

is a function from U to Y called the *restriction of f to U*. The restriction $f_U\colon U \to Y$ has the equation

$$f_U(x) = f(x) \text{ for all } x \in U.$$

For any set C the symbol I_C denotes the identity function on C given by

$$I_C = \{(x, x) \mid x \in C\}.$$

Lemma 1.2 *Suppose $f\colon A \to B$ is a given function. Then there is a function $g\colon B \to A$ such that*

$$g \circ f = I_A \text{ and } f \circ g = I_B$$

if and only if f is a one-to-one function and $g = f^{-1}$.

1.3.1 Graph of a Function

Let $f\colon X \to Y$ be a function. Define the *graph of f* to be the set

$$gr(f) = \{(x, y) \in X \times Y \mid y = f(x)\} = \{(x, f(x) \mid x \in X\}.$$

We normally identify the graph of a function with its graphical representation.

One of the exercises we should always do is to study how any new idea or concept interacts with what we already know. So, one thing that we should do is to undertake to study how the graph of a function relates to our usual operations on real numbers.

Let's say that you know what the graph of $y = f(x)$ looks like in the usual xy-plane. Then for a real number $a > 0$, what do the graphs of the following look like?

1. $f(x) + a \qquad f(x) - a$;
2. $f(x + a) \qquad f(x - a)$;
3. $f(ax) \qquad af(x)$;
4. $f(x/a) \qquad f(x)/a$

1.4 Topology or ***Analysis Situ***

[Topology] We like to look at properties in a particular situation and see which of those are generalizable. Such is the case of functions from $\mathbb{R} \to \mathbb{R}$. What are the properties of the functions that we should study and try to generalize? Are there certain properties of the reals that we should try to generalize and study? In the latter case, we know that when we look at the plane or 3-space, one of the properties that seems to be indispensable is the concept of a distance between to points. This is a very important property that we will discuss shortly. At this point, however, the author should point out that it is NOT indispensable and we topologists can do without a distance function quite well.

Let X be a set and suppose that d is a real-valued function defined for all pairs (x, y) of elements of X, i.e. $d\colon X \times X \to \mathbb{R}$, satisfying:

1. $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$;
2. $d(x, y) = d(y, x)$ for all $x, y \in X$;
3. $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

Such a function d on X is called a *metric* on X, and (X, d) is called a *metric space*. The reals are a metric space where $d(x, y) = |x - y|$ and the plane with its usual distance function is a metric space. In fact the space of all k-tuples

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_k) \quad \text{where } x_i \in \mathbb{R}$$

together with the metric

$$d(\boldsymbol{x}, \boldsymbol{y}) = \left[\sum_{i=1}^{k} (x_i - y_i)^2 \right]^{1/2}$$

is the usual *k-dimensional Euclidean space* and is denoted by $\mathbb{R}^k$.

Using the concept of a metric, we can now generalize what it means for a sequence to converge.

Definition 1.6 *A sequence $\{x_n\}$ in a metric space (X, d)* converges *to $x \in X$ if*

$$\lim_{n \to \infty} d(x_n, x) = 0.$$

In other words, given any $\epsilon > 0$ there is an $N \in \mathbb{Z}^+$ so that $d(x_n, x) < \epsilon$ whenever $n > N$.

A sequence is a *Cauchy sequence* is for each $\epsilon > 0$ there is an $n \in \mathbb{Z}^+$ so that if $m, n > N$ then $d(x_m, x_n) < \epsilon$. The metric space is said to be *complete* if every Cauchy sequence in X converges to some $x \in X$.

The Least Upper Bound Axiom guarantees that the reals with the usual metric is complete. Later, we may show that $\mathbb{R}^n$ is complete for each $n \in \mathbb{Z}^+$.

Let (X, d) be a metric space, $a \in X$, and $r \in \mathbb{R}^+$. The set $B_r(a) = \{x \in X \mid d(a, x) < r\}$ is called the (*open*) *ball* of radius r centered at a.

Let (X, d) be a metric space and let $U \subset X$. An element $x_0 \in U$ is *interior* to U if for some $r > 0$ we have that $B_r(x_0) \subset U$. We write U° for the set of points in U that are in the interior of U and call it the *interior of* U. A set $A \subset X$ is *open in* (X, d) if every point of A is interior to A, i.e., $A = A^\circ$.

Theorem 1.1 *The following conditions are met.*

1. *X is open in (X, d).*
2. *$\emptyset$ is open in (X, d).*
3. *The union of any collection of open sets is open.*
4. *The intersection of any finite collection of open sets is open.*

Note: We can state the definition of *limit* anew. The sequence $\{x_n\}$ converges to $x \in X$ if for any open set, U, containing x there is an $N \in \mathbb{Z}^+$ so that for $n > N$ we have $x_n \in U$.

Definition 1.7 *Let (X, d) be a metric space. A subset $F \subset X$ is* closed *if its complement $X \setminus F$ is open in (X, d).*

From the above theorem and de Morgan's Laws we have that any intersection of closed sets is closed and any finite union of closed sets is closed.

Example 1.2 Let $A_n = [\frac{1}{n}, 1)$. Note that $\cup_{n=2}^{\infty} A_n = (0,1)$ which is open, so the infinite union of closed sets **does not have to be** closed.

Since the intersection of any number of closed sets is closed, for any set $U \subset X$, we can define the *closure* of U, denoted U^-, to be the intersection of all closed sets containing U.[1]

The *boundary* of U, denoted ∂U, is defined to be $\partial U = U^- \setminus U^\circ$ and the points of this set are call *boundary points of* U.

Theorem 1.2 *Let U be a subset of a metric space (X, d).*

i) The set U is closed if and only if $U = U^-$.

ii) The set U is closed if and only if it contains the limit of every convergent sequence of points in U.

iii) An element is in U^- if and only if it is the limit of some sequence of points in U.

iv) A point is in the boundary of U if and only if it belongs to the closure of both U and its complement.

A set $U \subset X$ is said to be *dense* in X if $U^- = X$. We have seen that the rationals are dense in the reals from last semester.

Theorem 1.3 *Let $\{F_n\}$ be a decreasing sequence of closed, bounded, nonempty sets in $\mathbb{R}^n$. Then $F = \bigcap_{n=1}^{\infty} F_n$ is also closed, bounded and nonempty.*

PROOF: F is closed since it is an intersection of closed sets. It is bounded since it must be contained in F_1 which was bounded. We need to show that it is nonempty. Let $\boldsymbol{x}_n \in F_n$ for each n. Since this sequence is bounded, it has a convergent subsequence by the Bolzano-Weierstrauss Theorem. Call this subsequence $\{\boldsymbol{x}_{n_m}\}$ and let it converge to $\boldsymbol{x}_0 \in \mathbb{R}^n$. We need to show that $\boldsymbol{x}_0 \in F$. If $m \geq k$, then $n_m \geq k$ so $\boldsymbol{x}_{n_m} \in F_{n_m} \subset F_k$. Thus, the subsequence $\{x_{n_m}\}_{m=k}^{\infty}$ consists of points in F_k and converges to $\boldsymbol{x}_0$. By Part 2 of the above theorem, this puts $\boldsymbol{x}_0 \in F_k$ for each k and hence it is in F. ∎

Example 1.3 [Cantor Set] Let $I = [0,1] = T_0$. Define T_1 to be $T_0 \setminus (1/3, 2/3)$. We can see that $T_1 = [0, 1/3] \cup [2/3, 1]$. Now, take the middle third away from each of these pieces, i.e.,

$$T_2 = T_1 \setminus (1/9, 2/9) \cup (7/9, 8/9) = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$$

To get T_3 we will remove the middle third of each of these four intervals. Continue in this process and take $T_\infty = \bigcap_{n=0}^{\infty} T_n$. The set $T = T_\infty$ is called the Cantor set or Cantor's Dust.

[1] The interior of U can be defined to be the union of all open sets contained in U.

It looks like we shouldn't be left with much. We first removed an interval of length 1/3 and then two intervals of length 1/9 and then four intervals of length 1/27, and so on. Thus, we removed a set that has total length

$$\frac{1}{3}+\frac{2}{9}+\frac{4}{27}+\frac{8}{81}+\cdots=\sum_{n=0}^{\infty}\frac{2^n}{3^{n+1}}=\frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^n=\frac{1}{3}\left(\frac{1}{1-2/3}\right)=1.$$

Thus, what is left must have length 0.

A closer look at generating this set shows that there must be something left, since removing the "middle third" of each interval involved removing only open intervals that do not include their endpoints. So removing the line segment $(1/3, 2/3)$ from the original interval $[0, 1]$ leaves behind the points 1/3 and 2/3. Our subsequent steps do not remove these, or any other endpoints. So the Cantor set is not empty.

It can be shown that the points remaining in the Cantor set are exactly those that can be represented as

$$\sum_{k=0}^{\infty}\frac{a_k}{3^k}$$

where $a_k = 0$ or 2. It can then be shown that the Cantor set is *uncountable*.

The Cantor set is the prototype of a fractal. It is self-similar, because it is equal to two copies of itself, if each copy is shrunk by a factor of 3 and translated. Its Hausdorff dimension is equal to $\ln(2)/\ln(3)$.

Since the Cantor set is the complement of a union of open sets, it itself is a closed subset of the reals. The interior of T is the empty set so $T = \partial T$. Every point in the Cantor set is a limit point, but none is an interior point. A closed set in which every point is an limit point is also called a perfect set in topology, while a closed subset of the interval with no interior points is nowhere dense in the interval.

Definition 1.8 *Let (X, d) be a metric space. A family $\mathscr{U}$ of open sets is called an* open cover *for X if each point of X belongs to at least one set in $\mathscr{U}$, i.e.,*

$$X = \bigcup_{\mathscr{U}} U.$$

A subcover *of $\mathscr{U}$ is any subfamily of $\mathscr{U}$ that also covers X. A cover or subcover is* finite *if it contains only finitely many sets.*

A set K is compact *if every open cover of K has a finite subcover.*

While this is a very general definition, in $\mathbb{R}^n$ we have the following result.

Theorem 1.4 (Heine-Borel) *A subset K of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Other results about compact sets in $\mathbb{R}^n$ follow without proof.

Theorem 1.5 *i) Let U be a subset of $\mathbb{R}^n$. U is compact if and only if every sequence in U has a subsequence that converges to a point in U.*

ii) If U is a closed subset of a compact set V, then U is also compact.

iii) The finite union of compact sets is compact.

iv) If U is nonempty and compact in $\mathbb{R}^n$, then $\operatorname{lub} U$ and $\operatorname{glb} U$ belong to U.

1.5 Continuity

For functions $f\colon \mathbb{R} \to \mathbb{R}$ we finally were able to show that starting from a "sequence" definition of continuity —
f is continuous at $a \in \mathbb{R}$ if for every sequence $\{x_n\}$ that converges to a, the sequence $\{f(x_n)\}$ converges to $f(a)$ —
we eventually got to the usual definition of continuity.

Definition 1.9 *Let $f\colon \mathbb{R} \to \mathbb{R}$ be defined at $a \in \mathbb{R}$. We say that f is continuous at a if given any $\epsilon > 0$ there exists a $\delta > 0$ so that whenever $|x - a| < \delta$ then $|f(x) - f(a)| < \epsilon$.*

We carried this further to get a stronger condition that is satisfied by most of the functions that we will want to study, that of *uniform continuity.* A function $f\colon \mathbb{R} \to \mathbb{R}$ is *uniformly continuous* on the set $A \subset \mathbb{R}$ if for any $\epsilon > 0$ there is a $\delta > 0$ so that if $x, y \in A$ and $|x - y| < \delta$ then $|f(x) - f(y)| < \epsilon$.

In these definitions, we see that one of the important components is the distance function. Thus, we should be able to easily generalize these concepts to metric spaces.

Let (X, d_x) and (Y, d_y) be metric spaces and let $f\colon X \to Y$. We say that f is continuous at $a \in X$ if for each $\epsilon > 0$ there is a $\delta > 0$ so that if $d_x(x, a) < \delta$ then $d_y(f(x), f(a)) < \epsilon$. The function is *uniformly continuous* on a subset $U \subset X$ if if for any $\epsilon > 0$ there is a $\delta > 0$ so that if $x, y \in U$ and $d_x(x, y) < \delta$ then $d_y(f(x), f(y)) < \epsilon$.

One particulary useful application of this consists of functions with domain $\mathbb{R}$ having values in $\mathbb{R}^2$ or $\mathbb{R}^3$, or generally $\mathbb{R}^n$. For ease let's just consider the case of $\mathbb{R}^2$. Then we have $X = \mathbb{R}$, $d_x(x, y) = |x - y|$, $Y = \mathbb{R}^2$ and $d_y((x), (y)) = \sqrt{(x_1 - y_1)^2 + (y_1 - y_2)^2}$. The images of these functions are usually called curves or paths. We want to distinguish the function from its image, though. Let $\gamma\colon \mathbb{R} \to \mathbb{R}^n$ be continuous. We will call γ a *path*; its image $\gamma(\mathbb{R})$ in $\mathbb{R}^n$ will be called a *curve.*

As an example consider the function $\gamma_1(t) = (\cos t, \sin t)$. This function maps $\mathbb{R}$ onto the circle of radius 1 centered at the origin. Note that $\gamma_2(t) = (-\cos t, -\sin t)$ has the same image yet $\gamma_1(t) \neq \gamma_2(t)$ for all real numbers. The first function starts at $(1, 0)$ and wraps the real line around the circle in a "counterclockwise" direction while γ_2 starts at $(-1, 0)$ and wraps around the circle in a "clockwise" direction.

Theorem 1.6 *If $f_1, f_2, \ldots, f_n$ are continuous, real-valued functions on $\mathbb{R}$, then*

$$\gamma(t) = (f_1(t), f_2(t), \ldots, f_n(t))$$

defines a path in $\mathbb{R}^n$.

The following result really pushes the concept of continuity in its strongest direction, removing it from any reliance on a distance function.

Theorem 1.7 *Let (X, d_x) and (Y, d_y) be metric spaces and let $f \colon X \to Y$. The function f is continuous on X if and only if $f^{-1}(U)$ is an open subset of X for every open set U in Y.*

Recall that $f^{-1}(U) = \{x \in X \mid f(x) \in U\}$.

Proof: Suppose f is continuous on X. Let $U \subset Y$ be open and let $x_0 \in f^{-1}(U)$. We need to show that x_0 is in the interior of U.

Now, since $f(x_0) \in U$ and U is open, we know that for some $\epsilon > 0$ we have that $B_\epsilon(x_0) \subset U$. Since f is continuous at x_0, there exists a $\delta > 0$ so that if $d_x(x, x_0) < \delta$ then $d_y(f(x), f(x_0)) < \epsilon$. Thus, $f(x) \in U$ and hence $x \in f^{-1}(U)$. Thus, $B_\delta(x_0) = \{x \in X \mid d_x(x_0, x) < \delta\} \subset f^{-1}(U)$. Thus, x_0 is in the interior of $f^{-1}(U)$.

Now suppose that $f^{-1}(U)$ is an open set in X for every open set $U \subset Y$. Let $x_0 \in X$ and let $\epsilon > 0$. Then $U = B_\epsilon(f(x_0))$ is open in Y, so $f^{-1}(U)$ is open in X. Therefore, since $x_0 \in f^{-1}(U)$ there is some $\delta >$) so that $B_\delta(x_0) \subset f^{-1}(U)$, from which it follows that if $d_x(x, x_0) < \delta$ then $d_y(f(x), f(x_0)) < \epsilon$. Thus, f is continuous at x_0.∎

The following are offered without proof (at the present).

Theorem 1.8 *Let (X, d_x) and (Y, d_y) be metric spaces and let $f \colon X \to Y$ be continuous. Let K be a compact subset of X. Then*

i) $f(K)$ is a compact subset of Y, and

ii) f is uniformly continuous on K.

Corollary 1.1 *Let f be a continuous, real-valued function on a metric space (X, d). If K is a compact subset of X, then*

i) f is bounded on K,

ii) f assumes its maximum and its minimum on K.

We will find most of this more useful when we begin the study of multivariable calculus.

1.5.1 Connectedness

Let U be a subset of a metric space (X, d). The set C is *disconnected* if there are disjoint open subsets U_1 and U_2 in X so that $C \subset U_1 \cup U_2$, $C \cap U_1 \neq \emptyset$, and $C \cap U_2 \neq \emptyset$. A set is *connected* if it is not disconnected.

In the reals the intervals are connected sets and any set that is not an interval is disconnected.

Theorem 1.9 *Let (X, d_x) and (Y, d_y) be metric spaces and let $f\colon X \to Y$ be continuous. If E is a connected set of X, then $f(E)$ is a connected subset of Y.*

PROOF: Assume that $f(E)$ is not connected in Y. Then there exist disjoint, nonempty, open sets U_1 and U_2 in Y such that $f(E) \subset U_1 \cup U_2$.

Let $V_i = f^{-1}(U_i)$ for $i = 1, 2$. Then V_1 and V_2 are disjoint, nonempty, open sets in X with $E \subset V_1 \cup V_2$, and $E \cap V_1 \neq \emptyset$, and $E \cap V_2 \neq \emptyset$. This contradicts the fact that E is connected. ∎

Corollary 1.2 *Let f be a continuous, real-valued function on a metric space (X, d). If E is a connected subset of X, then $f(E)$ is an interval in $\mathbb{R}$. In particular, f has the intermediate value property.*

Note then that curves in $\mathbb{R}^n$ are connected.

Definition 1.10 *A subset E of a metric space (X, d) is* path connected *if for each pair of point $x, y \in E$, there is a continuous function $\gamma\colon [a, b] \to E$ such that $\gamma(a) = x$ and $\gamma(b) = y$. We call γ a path.*

Theorem 1.10 *If E in (X, d) is path-connected, then E is connected.*