Chapter 1. Real Limits, Continuity and Differentiation

Introduction

Real analysis is similar to calculus with a strong emphasis placed on rigorous mathematical proofs. In this first chapter, we shall prove some of the theorems, about limits, continuity and differentiation, which are normally omitted in calculus courses. We shall also introduce a few additional concepts. In the second chapter, we give a careful definition of the Riemann integral, which is usually described informally in calculus courses, and we prove many of the basic properties of the integral (which cannot be proven using only an informal definition of the integral). In the third chapter, we discuss sequences of functions and power series. We introduce the concept of uniform convergence (which is not usually mentioned in calculus courses) and prove various properties which allow us to show, for example, that power series can be differentiated and integrated term by term (a fact which is used often in calculus, physics and engineering, but which is not usually proven).

In the fourth chapter, we shall discuss inner product spaces and normed linear spaces (which students meet in linear algebra courses) and metric spaces. These include the Euclidean space $\mathbb{R}^n$, but they also include some infinite dimensional spaces such as spaces of sequences and spaces of functions. This is of interest because a differential or integral equation can often be interpreted as being given by a linear operator (a differential or integral operator) acting on an (infinite dimensional) vector space of functions, with the solutions forming a subspace. This can be studied using a combination of analysis and linear algebra.

In the next chapter, we study limits and continuity in metric spaces, and in the following chapter we introduce completeness and compactness.

We finish the course with some applications of our study of metric spaces. We shall prove an existence and uniqueness theorem for differential equations, which is often stated and used without proof in applied differential equations courses and in physics and engineering. We shall also prove results about polynomial approximations and trigonometric approximations and explore their relevance to Fourier series, which are a useful tool used in the solution of a number of differential equations.

Let us give a more detailed introduction to this first chapter. We shall prove the Monotone Convergence Theorem for sequences, the Intermediate Value Theorem and the Extreme Value Theorem for continuous functions, and the Chain Rule, the Inverse Function Theorem, and l'Hôpital's Rule for differentiable functions. These proofs are usually omitted in calculus courses. We shall also introduce a few new concepts including Cauchy sequences and uniform continuity.

What makes most of these theorems difficult to prove is that they involve a subtle property of $\mathbb{R}$ (involving the order relation $\leq$) which is not shared by $\mathbb{Q}$. This property can be vaguely described by saying that intervals in $\mathbb{R}$ do not have any holes while intervals in $\mathbb{Q}$ do have holes (the holes being the irrational numbers). The precise statement of this property is given in Theorem 1.5, which we shall accept, axiomatically, without proof. In mathematics, we must accept some properties, called axioms, without proof and then use these properties to prove other results.

Order Properties in $\mathbb{R}$

1.1 Theorem: *(Discreteness Property of $\mathbb{Z}$) For all $k, n \in \mathbb{Z}$ we have $k \leq n$ if and only if $k < n+1$. Equivalently, for all $n \in \mathbb{Z}$ there does not exist $k \in \mathbb{Z}$ with $n < k < n+1$.*

Proof: This theorem is accepted as true axiomatically, without proof.

1.2 Definition: Let $A \subseteq \mathbb{R}$. We say that A is **bounded above** (in $\mathbb{R}$) when there exists an element $b \in \mathbb{R}$ such that $x \leq b$ for all $x \in A$, and in this case we say that b is an **upper bound** for A. We say that A is **bounded below** (in $\mathbb{R}$) when there exists an element $a \in \mathbb{R}$ such that $a \leq x$ for all $x \in A$, and in this case we say that a is a **lower bound** for A. We say that A is **bounded** when A is bounded above and bounded below.

1.3 Definition: Let $A \subseteq \mathbb{R}$. We say that A has a **supremum** (or a **least upper bound**) when there exists an element $b \in \mathbb{R}$ such that b is an upper bound for A with $b \leq c$ for every upper bound $c \in \mathbb{R}$ for A, and in this case we say that b is the **supremum** (or the **least upper bound**) of A (note that if the supremum exists then it is unique since if b and c are both suprema then $b \leq c$ and $c \leq b$) and we write $b = \sup A$. When the supremum $b = \sup A$ exists and we have $b \in A$, then we also say that b is the **maximum element** of A and we write $b = \max A$.

We say that A has an **infimum** (or a **greatest lower bound**) when there exists an element $a \in \mathbb{R}$ such that a is a lower bound for A with $c \leq a$ for every lower bound c for A, and in this case we say that a is the **infimum** (or the **greatest lower bound**) of A and we write $a = \inf A$. When $a = \inf A \in A$ we also say that a is the **minimum element** of A and we write $a = \min A$.

1.4 Example: Let $A = (0, \infty) = \{x \in \mathbb{R} \,|\, 0 < x\}$ and $B = [1, \sqrt{2}) = \{x \in \mathbb{R} \,|\, 1 \leq x < \sqrt{2}\}$. The set A is bounded below but not bounded above. The numbers -1 and 0 are both lower bounds for A and we have $\inf A = 0$. The set A has no minimum element and no maximum element. The set B is bounded above and below. The numbers 0 and 1 are both lower bounds for B and the numbers $\sqrt{2}$ and 3 are both upper bounds for B. We have $\inf B = 1$ and $\sup B = \sqrt{2}$. The set B has a minimum element, namely $\min B = \inf B = 1$, but B has no maximum element.

1.5 Theorem: *(The Supremum and Infemum Properties of $\mathbb{R}$)*

(1) Every nonempty subset of $\mathbb{R}$ which is bounded above in $\mathbb{R}$ has a supremum in $\mathbb{R}$.
(2) Every nonempty subset of $\mathbb{R}$ which is bounded below in $\mathbb{R}$ has an infimum in $\mathbb{R}$.

Proof: We accept this axiomatically, without proof.

1.6 Theorem: *(Approximation Property of Supremum and Infimum) Let $\emptyset \neq A \subseteq \mathbb{R}$.*

(1) $b = \sup A$ if and only if b is an upper bound for A and $\forall \epsilon > 0 \;\; \exists x \in A \;\; b - \epsilon < x$.
(2) $b = \inf A$ if and only if b is a lower bound for A and $\forall \epsilon > 0 \; \exists x \in A \;\; x < b + \epsilon$.

Proof: We prove Part 1. Suppose that $b = \sup A$. Since b is an upper bound for A, it remains to prove that $\forall \epsilon > 0 \; \exists x \in A \;\; b - \epsilon < x$. Suppose not. Then we can choose $\epsilon > 0$ such that for all $x \in A$ we have $b - \epsilon \geq x$. But then $b - \epsilon$ is an upper bound for A, and since $b - \epsilon < b$, this contradicts the fact that b is the smallest upper bound for A.

Suppose, conversely, that b is an upper bound for A and $\forall \epsilon > 0 \; \exists x \in A \;\; b - \epsilon < x$. We need to prove that if c is any upper bound for A then $b \leq c$. We prove the contrapositive. Let $b > c$. From the assumption $\forall \epsilon > 0 \; \exists x \in A \;\; b - \epsilon < x$, taking $\epsilon = b - c > 0$, we can choose $x \in A$ such that $b - (b - c) < x$, that is $c < x$. Thus c is not an upper bound for A.

1.7 Theorem: *(Well-Ordering Properties of $\mathbb{Z}$ in $\mathbb{R}$)*

(1) Every nonempty subset of $\mathbb{Z}$ which is bounded above in $\mathbb{R}$ has a maximum element.
(2) Every nonempty subset of $\mathbb{Z}$ which is bounded below in $\mathbb{R}$ has a minimum element, in particular every nonempty subset of $\mathbb{N}$ has a minimum element.

Proof: We prove Part 1. Let A be a nonempty subset of $\mathbb{Z}$ which is bounded above. By the Supremum Property of $\mathbb{R}$, A has a supremum in $\mathbb{R}$. Let $n = \sup A$. We must show that $n \in A$. Suppose, for a contradiction, that $n \notin A$. By the Approximation Property (using $\epsilon = 1$), we can choose $a \in A$ with $n - 1 < a \leq n$. Note that $a \neq n$ since $a \in A$ and $n \notin A$ and so we have $a < n$. By the Approximation Property again (using $\epsilon = n - a$) we can choose $b \in A$ with $a < b \leq n$. Since $a < b$ we have $b - a > 0$. Since $n - 1 < a$ and $b \leq n$ we have $1 = n - (n-1) > b - a$. But then we have $b - a \in \mathbb{Z}$ with $0 < b - a < 1$ which contradicts the Discreteness Property of $\mathbb{Z}$. Thus $n \in A$ so A has a maximum element.

1.8 Theorem: *(Floor and Ceiling Properties of $\mathbb{Z}$ in R)*

(1) (Floor Property) For every $x \in \mathbb{R}$ there exists a unique $n \in \mathbb{Z}$ with $x - 1 < n \leq x$.
(2) (Ceiling Property) For every $x \in \mathbb{R}$ there exists a unique $m \in \mathbb{Z}$ with $x \leq m < x + 1$.

Proof: We prove Part 1. First we prove uniqueness. Let $x \in \mathbb{R}$ and suppose that $n, m \in \mathbb{Z}$ with $x - 1 < n \leq x$ and $x - 1 < m \leq x$. Since $x - 1 < n$ we have $x < n + 1$. Since $m \leq x$ and $x < n + 1$ we have $m < n + 1$, hence $m \leq n$ by the Discreteness Property of $\mathbb{Z}$. Similarly, $n \leq m$. Since $n \leq m$ and $m \leq n$, we have $n = m$. This proves uniqueness.

Next we prove existence. Let $x \in \mathbb{R}$. First let us consider the case that $x \geq 0$. Let $A = \{k \in \mathbb{Z} \,|\, k \leq x\}$. Note that $A \neq \emptyset$ (because $0 \in A$) and A is bounded above by x. By The Well-Ordering Property of $\mathbb{Z}$ in $\mathbb{R}$, A has a maximum element. Let $n = \max A$. Since $n \in A$ we have $n \in \mathbb{Z}$ and $n \leq x$. Also note that $x - 1 < n$ since $x - 1 \geq n \Longrightarrow x \geq n + 1 \Longrightarrow n + 1 \in A \Longrightarrow n \neq \max A$. Thus for $n = \max A$ we have $n \in \mathbb{Z}$ with $x - 1 < n \leq x$, as required.

Next consider the case that $x < 0$. If $x \in \mathbb{Z}$ we can take $n = x$. Suppose that $x \notin \mathbb{Z}$. We have $-x > 0$ so, by the previous paragraph, we can choose $m \in \mathbb{Z}$ with $-x - 1 < m \leq -x$. Since $m \in \mathbb{Z}$ but $x \notin \mathbb{Z}$ we have $m \neq -x$ so that $-x - 1 < m < -x$ and hence $x < -m < x + 1$. Thus we can take $n = -m - 1$ to get $x - 1 < n < x$. This completes the proof of Part 1.

1.9 Definition: Let $x \in \mathbb{R}$. The **floor** of x, denoted by $\lfloor x \rfloor$, is be the unique $n \in \mathbb{Z}$ with $x - 1 < n \leq x$. The function $f : \mathbb{R} \to \mathbb{Z}$ given by $f(x) = \lfloor x \rfloor$ is called the **floor function**. Similarly, the **ceiling** of x, denoted by $\lceil x \rceil$, is the unique $n \in \mathbb{Z}$ with $x \leq n < x + 1$. The function $f : \mathbb{R} \to \mathbb{Z}$ given by $f(n) = \lceil x \rceil$ is called the **ceiling function**.

1.10 Theorem: *(Archimedean Properties of $\mathbb{Z}$ in $\mathbb{R}$)*

(1) For every $x \in \mathbb{R}$ there exists $n \in \mathbb{Z}$ with $n > x$.
(2) For every $x \in \mathbb{R}$ there exists $m \in \mathbb{Z}$ with $m < x$.

Proof: Let $x \in \mathbb{R}$. Let $n = \lfloor x \rfloor + 1$ and $m = \lfloor x \rfloor - 1$. Since $x - 1 < \lfloor x \rfloor$ we have $x < \lfloor x \rfloor + 1 = n$ and since $\lfloor x \rfloor \leq x$ we have $m = \lfloor x \rfloor - 1 \leq x - 1 < x$.

1.11 Theorem: *(Density of $\mathbb{Q}$ in $\mathbb{R}$) For all $a, b \in \mathbb{R}$ with $a < b$ there exists $q \in \mathbb{Q}$ with $a < q < b$.*

Proof: Let $a, b \in \mathbb{R}$ with $a < b$. By the Archimedean Property, we can choose $n \in \mathbb{Z}$ with $n > \frac{1}{b-a} > 0$. Then $n(b-a) > 1$ and so $nb > na + 1$. Let $k = \lfloor na + 1 \rfloor$. Then we have $na < k \leq na + 1 < nb$ hence $a < \frac{k}{n} < b$. Thus we can take $q = \frac{k}{n}$ to get $a < q < b$.

Limits of Sequences in $\mathbb{R}$

1.12 Definition: For $p \in \mathbb{Z}$, let $\mathbb{Z}_{\geq p} = \{k \in \mathbb{Z} | k \geq p\}$. A **sequence** in a set A is a function of the form $x : \mathbb{Z}_{\geq p} \to A$ for some $p \in \mathbb{Z}$. Given a sequence $x : \mathbb{Z}_{\geq p} \to A$, the n^{th} **term** of the sequence is the element $x_n = x(n) \in A$, and we denote the sequence x by

$$(x_n)_{n \geq p} = (x_p, x_{p+1}, x_{p+2}, \cdots).$$

Note that the range of the sequence $(x_n)_{n\geq p}$ is the set $\{x_n\}_{n \geq p} = \{x_n \,|\, n \in \mathbb{Z}_p\}$.

1.13 Definition: Let $(x_n)_{n\geq p}$ be a sequence in $\mathbb{R}$. For $a \in \mathbb{R}$ we say that the sequence $(x_n)_{n\geq p}$ **converges** to a (or that the **limit** of $(x_n)_{n\geq p}$ is equal to a), and we write $x_n \to a$ (as $n \to \infty$), or we write $\lim\limits_{n\to\infty} x_n = a$, when

$$\forall\, 0 < \epsilon \in \mathbb{R}\ \exists\, m \in \mathbb{Z}\ \forall\, n \in \mathbb{Z}_{\geq p}\ \big(n \geq m \Longrightarrow |x_n - a| < \epsilon\big).$$

We say that the sequence $(x_n)_{n\geq p}$ **converges** (in $\mathbb{R}$) when there exists $a \in \mathbb{R}$ such that $(x_n)_{n\geq p}$ converges to a. We say that the sequence $(x_n)_{n\geq p}$ **diverges** (in $\mathbb{R}$) when it does not converge (to any $a \in \mathbb{R}$). We say that $(x_n)_{n\geq p}$ **diverges to infinity**, or that the limit of $(x_n)_{n\geq p}$ is equal to **infinity**, and we write $x_n \to \infty$ (as $n \to \infty$), or we write $\lim\limits_{n\to\infty} x_n = \infty$, when

$$\forall\, r \in \mathbb{R}\ \exists\, m \in \mathbb{Z}\ \forall\, n \in \mathbb{Z}_{\geq p}\ \big(n \geq m \Longrightarrow x_n > r\big).$$

Similarly we say that $(x_n)_{k\geq p}$ **diverges to** $-\infty$, or that the limit of $(x_n)_{n\geq p}$ is equal to **negative infinity**, and we write $x_n \to -\infty$ (as $n \to \infty$), or we write $\lim\limits_{n\to\infty} x_n = -\infty$ when

$$\forall\, r \in \mathbb{R}\ \exists\, m \in \mathbb{Z}\ \forall\, n \in \mathbb{Z}_{\geq p}\ \big(n \geq m \Longrightarrow x_n < r\big).$$

1.14 Note: We shall assume that students are familiar with sequences and limits of sequences from first-year calculus. For example, students should know that if the limit of a sequence exists then it is unique. Also, the limit does not depend on the first few terms (indeed the first finitely many terms) and so we often omit the starting value p from our notation and write the sequence $(x_n)_{n\geq p}$ as (x_n). Students should also be able to calculate limits using various limit rules, such as Operations on Limits, the Comparison Theorem and the Squeeze Theorem (which can all be found in the lecture notes for MATH 137).

1.15 Definition: Let $(x_n)_{n\geq p}$ be a sequence in $\mathbb{R}$. For $b \in \mathbb{R}$, we say that the sequence $(x_n)_{n\geq p}$ is **bounded above** by b when the set $\{x_n\}_{n\geq p}$ is bounded above by b, that is when $x_n \leq b$ for all $n \in \mathbb{Z}_{\geq p}$, and we say that the sequence $(x_n)_{n\geq p}$ is **bounded below** by b when the set $\{x_n\}_{n\geq p}$ is bounded below by b, that is when $b \leq x_n$ for all $n \in \mathbb{Z}_{\geq p}$. We say $(x_n)_{n\geq p}$ is **bounded above** when it is bounded above by some element $b \in \mathbb{R}$, we say that $(x_n)_{n\geq p}$ is **bounded below** when it is bounded below by some $b \in \mathbb{R}$, and we say that $(x_n)_{n\geq p}$ is **bounded** when it is bounded above and bounded below.

1.16 Definition: Let $(x_n)_{n\geq p}$ be a sequence in $\mathbb{R}$. We say that $(x_n)_{n\geq p}$ is **increasing** (or **nondecreasing**) when for all $k, l \in \mathbb{Z}_{\geq p}$, if $k \leq l$ then $x_k \leq x_l$. We say that $(x_n)_{n\geq p}$ is **strictly increasing** when for all $k, l \in \mathbb{Z}_{\geq p}$, if $k < l$ then $x_k < x_l$. Similarly, we say that $(x_n)_{n\geq p}$ is **decreasing** (or **nonincreasing**) when for all $k, l \in \mathbb{Z}_{\geq p}$, if $k \leq l$ the $x_k \geq x_l$ and we say that $(x_n)_{n\geq p}$ is **strictly decreasing** when for all $k, l \in \mathbb{Z}_{\geq p}$, if $k < l$ the $x_k > x_l$. We say that $(x_n)_{n\geq p}$ is **monotonic** when it is either increasing or decreasing.

1.17 Theorem: *(Monotonic Convergence Theorem) Let $(x_n)_{n\geq p}$ be a sequence in $\mathbb{R}$.*

(1) Suppose $(x_n)_{n\geq p}$ is increasing. If $(x_n)_{n\geq p}$ is bounded above then $x_n \to \sup\{x_n\}_{n\geq p}$, and if $(x_n)_{n\geq p}$ is not bounded above then $x_n \to \infty$.
(2) Suppose $(x_n)_{n\geq p}$ is decreasing. If $(x_n)_{n\geq p}$ is bounded below then $x_n \to \inf\{x_n\}_{n\geq p}$, and if $(x_n)_{n\geq p}$ is not bounded below then $x_n \to -\infty$.

Proof: We prove Part 1 in the case that $(x_n)_{n\geq p}$ is increasing and bounded above, say by $b \in \mathbb{R}$. Let $A = \{x_n\}_{n\geq p}$. Note that A is nonempty (since $x_p \in A$) and bounded above (indeed b is an upper bound for A). By the Supremum Property of $\mathbb{R}$, A has a supremum in $\mathbb{R}$. Let $a = \sup\{x_n\}_{n\geq p}$. Note that $a \geq x_n$ for all $n \in \mathbb{Z}_{n\geq p}$ and $a \leq b$, by the definition of the supremum. Let $\epsilon > 0$. By the Approximation Property of the supremum, we can choose an index $m \in \mathbb{Z}_{n\geq p}$ so that the element $x_m \in A$ satisfies $a - \epsilon < x_m \leq a$. Since $(x_n)_{n\geq p}$ is increasing, for all $n \geq m$ we have $x_n \geq x_m$, so we have $a - \epsilon < x_m \leq x_n \leq a$ and hence $|x_n - a| < \epsilon$. Thus $\lim\limits_{n\to\infty} x_n = a \leq b$.

1.18 Definition: For $a, b \in \mathbb{R}$ with $a \leq b$ we write

$$\begin{gathered}(a,b) = \big\{x \in \mathbb{R} \big| a < x < b\big\}\ ,\ [a,b] = \big\{x \in \mathbb{R} \big| a \leq x \leq b\big\}\,,\\ (a,b] = \big\{x \in \mathbb{R} \big| a < x \leq b\big\}\ ,\ [a,b) = \big\{x \in \mathbb{R} \big| a \leq x < b\big\}\,,\\ (a,\infty) = \big\{x \in \mathbb{R} \big| a < x\big\}\ ,\ [a,\infty) = \big\{x \in \mathbb{R} \big| a \leq x\big\}\,,\\ (-\infty,b) = \big\{x \in \mathbb{R} \big| x < b\big\}\ ,\ (-\infty,b] = \big\{x \in \mathbb{R} \big| x \leq b\big\}\,,\\ (-\infty,\infty) = \mathbb{R}\,.\end{gathered}$$

An **interval** in $\mathbb{R}$ is any set of one of the above forms. In the case that $a = b$ we have $(a,b) = [a,b) = (a,b] = \emptyset$ and $[a,b] = \{a\}$, and these intervals are called **degenerate**. The other intervals are called **nondegenerate**.. The intervals $\emptyset$, (a,b), (a,∞), $(-\infty,b)$ and $(-\infty,\infty)$ are called **open** intervals. The intervals $\emptyset$, $[a,b]$, $[a,\infty)$, $(-\infty,b]$ and $(-\infty,\infty)$ are called **closed** intervals. The intervals $\emptyset$, (a,b), $(a,b]$, $[a,b)$ and $[a,b]$ are **bounded** and the intervals (a,∞), $[a,\infty)$, $(-\infty,b)$, $(-\infty,b]$ and $(-\infty,\infty)$ are **unbounded**.

1.19 Theorem: *(Nested Interval Theorem) Let $(I_n)_{n\in\mathbb{Z}^+}$ be a sequence of nonempty, closed bounded intervals in $\mathbb{R}$ with $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$. Then $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$.*

Proof: For each $n \in \mathbb{Z}^+$, let $I_n = [a_n, b_n]$ with $a_n \leq b_n$. For each $n \in \mathbb{Z}^+$, since $I_{n+1} \subseteq I_n$ we have $a_n \leq a_{n+1} \leq b_{n+1} \leq b_n$. Since $a_n \leq a_{n+1}$ for all $n \in \mathbb{Z}^+$, the sequence $(a_n)_{n\geq 1}$ is increasing. Since $a_n \leq b_n \leq b_{n-1} \leq \cdots \leq b_1$ for all $n \in \mathbb{Z}^+$, the sequence $(a_n)_{n\in\mathbb{Z}^+}$ is bounded above by b_1. Since $(a_n)_{n\geq 1}$ is increasing and bounded above, it converges with $\lim_{n\to\infty} a_n = a$ where $a = \sup\{a_n\}_{n\geq 1}$. Similarly, $(b_n)_{n\geq 1}$ is decreasing and bounded below by a_1, and so it converges with $\lim_{n\to\infty} b_n = b$ where $b = \inf\{b_n\}_{n\geq 1}$. Since $a_n \leq b_n$ for all $n \in \mathbb{Z}^+$, by the Comparison Theorem we have $a = \lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n \leq b$. Since $a \leq b$, the interval $[a,b]$ is not empty. For all $n \in \mathbb{Z}^+$, since $a = \sup\{a_n\}_{n\geq 1}$ and $a_n \leq b_n$ and $b_n = \inf\{b_n\}_{n\geq 1}$, we have $a_n \leq a \leq b \leq b_n$ so that $[a,b] \subseteq [a_n, b_n] = I_n$. Since $[a,b] \subseteq I_n$ for all $n \in \mathbb{Z}^+$, we have $[a,b] \subseteq \bigcap_{n=1}^{\infty} I_n$ so that $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$.

1.20 Note: The above theorem does not hold for nonempty bounded open intervals. For example, for $I_n = \left(0, \frac{1}{n}\right)$ we have $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$ but $\bigcap_{n=1}^{\infty} I_n = \emptyset$. The theorem also does not hold for unbounded closed intervals. For example, consider $I_n = [n, \infty)$.

1.21 Definition: Let $(x_n)_{n\geq p}$ be a sequence in a set A. Given a strictly increasing function $f:\mathbb{Z}_{\geq q}\to\mathbb{Z}_{\geq p}$, write $n_k=f(k)$ and let $y_k=x_{n_k}$ for all $k\geq q$. Then the sequence $(y_k)_{k\geq q}$ is called a **subsequence** of the sequence $(x_n)_{n\geq p}$. In other words, a subsequence of $(x_n)_{n\geq p}$ is a sequence of the form

$$\left(x_{n_q},x_{n_{q+1}},x_{n_{q+2}},\cdots\right) \text{ with } p\leq n_q<n_{q+1}<n_{q+2}<\cdots.$$

Given a bijective function $f:\mathbb{Z}_{\geq q}\to\mathbb{Z}_{\geq p}$, write $n_k=f(k)$ and let $y_k=x_{n_k}$ for $k\geq q$. Then the sequence $(y_k)_{k\geq q}$ is called a **rearrangement** of the sequence $(x_n)_{n\geq p}$.

1.22 Theorem: *(Subsequences and Rearrangements) Let $(x_n)_{n\geq p}$ be a convergent sequence in $\mathbb{R}$ with $x_n\to a$. Then*

(1) every subsequence of $(x_n)_{n\geq p}$ converges to a, and
(2) every rearrangement of $(x_n)_{n\geq p}$ converges to a.

Proof: We shall prove Parts 1 and 2 simultaneously. Let $f:\mathbb{Z}_{\geq q}\to\mathbb{Z}_{\geq p}$ be an injective map. Write $n_k=f(k)$ and let $y_k=x_{n_k}$ for $k\geq q$. Let $\epsilon>0$. Choose $m_1\in\mathbb{Z}$ so that $n\geq m_1\implies |x_n-a|<\epsilon$. Since f is injective, there are only finitely many indices k with $p\leq f(k)<m_1$. Choose $m\in\mathbb{Z}$ with m larger than every such index k. Then for $k\geq m$ we have $n_k=f(k)\geq m_1$ and so $|y_k-a|=|x_{n_k}-a|<\epsilon$.

1.23 Theorem: *(Bolzano-Weirstrass Theorem in $\mathbb{R}$) Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Proof: Let $(x_n)_{n\geq p}$ be a bounded sequence in $\mathbb{R}$. Choose $a,b\in\mathbb{R}$ with $a\leq x_n$ for all $n\in\mathbb{Z}_{\geq p}$ and $x_n\leq b$ for all $n\in\mathbb{Z}_{\geq p}$. Then we have $x_n\in[a,b]$ for all $n\in\mathbb{Z}_{\geq p}$. We define a sequence of nonempty closed bounded intervals recursively as follows. Let $a_1=a$ and $b_1=b$, and let $I_1=[a_1,b_1]=[a,b]$. Note that there are infinitely many indices $j\in\mathbb{Z}_{\geq p}$ for which $x_j\in I_1$ (indeed $x_j\in I_1$ for every $j\in\mathbb{Z}_{\geq p}$). Let $n\in\mathbb{Z}^+$ and suppose, inductively, that we have chosen intervals $I_1,I_2,\cdots I_n$ with $I_1\supseteq I_2\supseteq\cdots\supseteq I_n$ such that for each $1\leq k\leq n$, we have $I_k=[a_k,b_k]$ with $b_k-a_k=\frac{b-a}{2^{k-1}}$, and there exist infinitely many indices $j\in\mathbb{Z}_{\geq p}$ for which $x_j\in I_k$. Note that $I_n=[a_n,b_n]=\left[a_n,\frac{a_n+b_n}{2}\right]\cup\left[\frac{a_n+b_n}{2},b_n\right]$. Let I_{n+1} be equal to one of the two intervals $\left[a_n,\frac{a_n+b_n}{2}\right]$ and $\left[\frac{a_n+b_n}{2},b_n\right]$, chosen in such a way that there are infinitely many indices j with $x_j\in I_{n+1}$. Repeat the procedure recursively to obtain a sequence $(I_n)_{n\geq 1}$ of nonempty closed bounded intervals with $I_1\supseteq I_2\supseteq I_3\supseteq\cdots$ such that for every $n\in\mathbb{Z}^+$ we have $I_n=[a_n,b_n]$ with $b_n-a_n=\frac{b-a}{2^{n-1}}$ and there exists infinitely many $j\in\mathbb{Z}_{\geq p}$ for which $x_j\in I_n$. By the Nested Interval Theorem, $\bigcap_{n=1}^{\infty}I_n\neq\emptyset$. Choose a point $c\in\bigcap_{n=1}^{\infty}$, and note that $c\in I_n$ for every $n\in\mathbb{Z}^+$.

We shall now construct a subsequence of $(x_n)_{n\geq p}$ which converges to c. Since for each $n\in\mathbb{Z}^+$ there exist infinitely many indices $j\in\mathbb{Z}^+$ with $x_j\in I_n$, we can construct a subsequence of $(x_n)_{n\geq p}$ as follows. Choose $n_1\in\mathbb{Z}_{\geq p}$ and note that $x_{n_1}\in I_1$, then choose $n_2>n_1$ so that $x_{n_2}\in I_2$, then choose $n_3>n_2$ with $x_{n_3}\in I_3$, and so on. In this way, we obtain a subsequence $(x_{n_k})_{k\geq 1}$ of $(x_n)_{n\geq p}$ with $x_{n_k}\in I_k$ for all $k\in\mathbb{Z}^+$. We claim that $\lim_{k\to\infty}x_{n_k}=c$. Let $\epsilon>0$. Since $\frac{b-a}{2^{m-1}}\to 0$ as $m\to\infty$, we can choose $m\in\mathbb{Z}^+$ so that $\frac{b-a}{2^{m-1}}<\epsilon$. For $k\geq m$, since $c\in I_k=[a_k,b_k]$ and $x_{n_k}\in I_k=[a_k,b_k]$, it follows that

$$|x_{n_k}-c|\leq b_k-a_k=\tfrac{b-a}{2^{k-1}}\leq\tfrac{b-a}{2^{m-1}}<\epsilon.$$

Thus $\lim_{k\to\infty}x_{n_k}=c$, as claimed.

1.24 Definition: Let $(x_n)_{n\geq p}$ be a sequence in $\mathbb{R}$. We say that $(x_n)_{n\geq p}$ is **Cauchy** when

$$\forall\,\epsilon>0\ \exists\,m\in\mathbb{Z}_{\geq p}\ \forall\,k,l\in\mathbb{Z}_{\geq p}\ \big(k,l\geq m \Longrightarrow |x_k-x_l|<\epsilon\big).$$

1.25 Theorem: *Every convergent sequence in $\mathbb{R}$ is Cauchy.*

Proof: Let $(x_n)_{n\geq p}$ be a convergent sequence in $\mathbb{R}$ and let $a=\lim_{n\to\infty}x_n$. Let $\epsilon>0$ and choose $m\in\mathbb{Z}_{\geq p}$ so that $n\geq m \Longrightarrow |x_n-a|<\frac{\epsilon}{2}$. Then for $k,l\geq m$, using the Triangle Inequality, we have

$$|x_k-x_l|=\big|(x_k-a)+(a-x_l)\big|\leq|x_k-a|+|a-x_l|<\tfrac{\epsilon}{2}+\tfrac{\epsilon}{2}=\epsilon.$$

Thus $(x_n)_{n\geq p}$ is Cauchy.

1.26 Theorem: *(The Cauchy Criterion for Convergence, or The Completeness of $\mathbb{R}$) Every Cauchy sequence in $\mathbb{R}$ converges in $\mathbb{R}$.*

Proof: Let $(x_n)_{n\geq p}$ be a Cauchy sequence in $\mathbb{R}$. We claim that $(x_n)_{n\geq p}$ is bounded. Since $(x_n)_{n\geq p}$ is Cauchy, we can choose $m\in\mathbb{Z}_{\geq p}$ so that $k,l\geq m\Longrightarrow|x_k-x_l|<1$. In particular, for all $k\geq m$ we have $|x_k-x_m|<1$ so that $x_m-1<x_k<x_m+1$ It follows that $(x_n)_{n\geq p}$ is bounded above by $\max\big\{x_p,x_{p+1},\cdots,x_{m-1},x_m+1\big\}$ and bounded below by $\min\big\{x_p,x_{p+1},\cdots,x_{m-1},x_m-1\big\}$.

Because $(x_n)_{n\geq p}$ is bounded, by the Bolzano-Weierstrass Theorem, we can choose a convergent subsequence $(x_{n_k})_{k\geq q}$ of $(x_n)_{n\geq p}$ and let $a=\lim_{k\to\infty}x_{n_k}$. We claim that the original sequence $(x_n)_{n\geq p}$ converges with $\lim_{n\to\infty}x_n=a$. Let $\epsilon>0$. Since $(x_n)_{n\geq p}$ is Cauchy, we can choose $M\in\mathbb{Z}_{\geq p}$ so that $n,m\geq M\Longrightarrow|x_n-x_m|<\frac{\epsilon}{2}$. Since $x_{n_k}\to a$ we can choose $K\in\mathbb{Z}_{\geq q}$ so that $k\geq K\Longrightarrow|x_{n_k}-a|<\frac{\epsilon}{2}$. Choose an index $k\geq K$ so that $n_k\geq M$. Then for all $n\geq M$ we have

$$|x_n-a|=\big|(x_n-x_{n_k})+(x_{n_k}-a)\big|\leq|x_n-x_{n_k}|+|x_{n_k}-a|<\tfrac{\epsilon}{2}+\tfrac{\epsilon}{2}=\epsilon.$$

Thus $x_n\to a$, as claimed.

1.27 Definition: Let $(x_n)_{n\geq p}$ be a sequence in $\mathbb{R}$. The **series** $\sum_{n\geq p}x_n$ is defined to be the sequence $(S_\ell)_{\ell\geq p}$ where $S_\ell=\sum_{n=p}^{\ell}x_n=x_p+x_{p+1}+\cdots+x_\ell$. The term S_ℓ is called the ℓ^{th} **partial sum** of the series $\sum_{n\geq p}x_n$. The **sum** of the series, denoted by $S=\sum_{n=p}^{\infty}x_n=x_p+x_{p+1}+x_{p+2}+\cdots$, is defined to be the limit $\lim_{\ell\to\infty}S_\ell$, if it exists, and we say the series **converges** when the sum exists and is finite.

1.28 Note: As with sequences, we assume students are familiar with series and various tests for convergence.

1.29 Theorem: *(Cauchy Criterion for Series) Let $(x_n)_{n\geq p}$ be a sequence. Then the series $\sum_{n\geq p}x_n$ converges if and only if*

$$\forall\,\epsilon>0\ \exists\,N\in\mathbb{Z}_{\geq p}\ \forall\,\ell,m\in\mathbb{Z}_{\geq p}\left(m>\ell\geq N\Longrightarrow\Big|\sum_{n=\ell+1}^{m}x_n\Big|<\epsilon\right).$$

Proof: This follows from the Cauchy Criterion for the convergence of the sequence of partial sums. Indeed $(S_\ell)_{\ell\geq p}$ converges if and only if for all $\epsilon>0$ there exists $N\geq p$ such that $m>\ell\geq N\Longrightarrow|S_m-S_\ell|<\epsilon$, and we have $|S_m-S_\ell|=\Big|\sum\limits_{n=p}^{m}x_n-\sum\limits_{n=p}^{\ell}x_n\Big|=\Big|\sum\limits_{n=\ell+1}^{m}x_n\Big|$.

Limits of Functions and Continuity in $\mathbb{R}$

1.30 Definition: Let $A \subseteq \mathbb{R}$. For $a \in \mathbb{R}$, we say that a is a **limit point** of A when

$$\forall \delta > 0\ \exists x \in A\ 0 < |x - a| < \delta.$$

When $a \in A$ and a is not a limit point of A we say that a is an **isolated point** of A.

1.31 Definition: Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$. When a is a limit point of A, we make the following definitions.

(1) For $b \in \mathbb{R}$, we say that the **limit** of $f(x)$ as x tends to a is equal to b, and we write $\lim\limits_{x \to a} f(x) = b$ or we write $f(x) \to b$ as $x \to a$, when

$$\forall \epsilon > 0\ \exists \delta > 0\ \forall x \in A\ \big(0 < |x - a| < \delta \Longrightarrow |f(x) - b| < \epsilon\big).$$

(2) We say the limit of $f(x)$ as x tends to a is equal to **infinity**, and we write $\lim\limits_{x \to a} f(x) = \infty$, or we write $f(x) \to \infty$ as $x \to a$, when

$$\forall r \in \mathbb{R}\ \exists \delta > 0\ \forall x \in A\ \big(0 < |x - a| < \delta \Longrightarrow f(x) > r\big).$$

(3) We say that the limit of $f(x)$ as x tends to a is equal to **negative infinity**, and we write $\lim\limits_{x \to a} f(x) = -\infty$, or we write $f(x) \to -\infty$ as $x \to a$, when

$$\forall r \in \mathbb{R}\ \exists \delta > 0\ \forall x \in A\ \big(0 < |x - a| < \delta \Longrightarrow f(x) < r\big).$$

1.32 Note: We assume that students are familiar with limits of functions and are able to calculate limits using various limit rules , such as Operations on Limits, and the Comparison and the Squeeze Theorems. We also assume familiarity with one-sided limits $\lim\limits_{x \to a^+} f(x)$ and $\lim\limits_{x \to a^-} f(x)$ as well as asymptotic limits $\lim\limits_{x \to -\infty} f(x)$ and $\lim\limits_{x \to \infty} f(x)$. Here is one theorem that relates limits of functions and limits of sequences which students may not have seen.

1.33 Theorem: *(Sequential Characterization of Limits of Functions) Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$, let $a \in \mathbb{R}$ be a limit point of A, and let $b \in \mathbb{R}$. Then $\lim\limits_{x \to a} f(x) = b$ if and only if for every sequence (x_k) in $A \setminus \{a\}$ with $x_k \to a$ we have $f(x_k) \to b$.*

Proof: Suppose that $\lim\limits_{x \to a} f(x) = b$. Let (x_k) be a sequence in $A \setminus \{a\}$ with $x_k \to a$. Let $\epsilon > 0$. Since $\lim\limits_{x \to a} f(x) = b$, we can choose $\delta > 0$ so that $0 < |x - a| < \delta \Longrightarrow |f(x) - b| < \epsilon$. Since $x_k \to a$ we can choose $m \in \mathbb{Z}$ so that $k \geq m \Longrightarrow |x_k - a| < \delta$. Then for $k \geq m$, we have $|x_k - a| < \delta$ and we have $x_k \neq a$ (since the sequence (x_k) is in the set $A \setminus \{a\}$) so that $0 < |x_k - a| < \delta$ and hence $|f(x_k) - b| < \epsilon$. This shows that $f(x_k) \to b$.

Conversely, suppose that $\lim\limits_{x \to a} f(x) \neq b$. Choose $\epsilon_0 > 0$ so that for all $\delta > 0$ there exists $x \in A$ with $0 < |x - a| < \delta$ and $|f(x) - b| \geq \epsilon_0$. For each $k \in \mathbb{Z}^+$, choose $x_k \in A$ with $0 < |x_k - a| \leq \frac{1}{k}$ and $|f(x_k) - b| \geq \epsilon_0$. In this way we obtain a sequence $(x_k)_{k \geq 1}$ in $A \setminus \{a\}$. Since $|x_k - a| \leq \frac{1}{k}$ for all $k \in \mathbb{Z}^+$, it follows that $x_k \to a$ (indeed, given $\epsilon > 0$ we can choose $m \in \mathbb{Z}$ with $m > \frac{1}{\epsilon}$ and then $k \geq m \Longrightarrow |x_k - a| \leq \frac{1}{k} \leq \frac{1}{m} < \epsilon$). Since $|f(x_k) - b| \geq \epsilon_0$ for all k, it follows that $f(x_k) \not\to b$ (indeed if we had $f(x_k) \to b$ we could choose $m \in \mathbb{Z}$ so that $k \geq m \Longrightarrow |f(x_k) - b| < \epsilon_0$ and then we could choose $k = m$ to get $|f(x_k) - b| < \epsilon_0$).

1.34 Definition: Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$. For $a \in A$, we say that f is **continuous** at a when

$$\forall \epsilon > 0 \;\; \exists \delta > 0 \;\; \forall x \in A \;\; \big(|x - a| < \delta \Longrightarrow |f(x) - f(a)| < \epsilon \big).$$

We say that f is **continuous** (on A) when f is continuous at every point $a \in A$.

1.35 Note: Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and let $a \in A$. Verify, as an exercise, that
(1) if a is an isolated point of A then f is continuous at a, and
(2) if a is a limit point of A then f is continuous at a if and only if $\lim\limits_{x \to a} f(x) = f(a)$.

1.36 Note: We assume the reader is familiar with continuity. In particular, we assume the reader knows that every elementary function is continuous in its domain (an **elementary** function is any function which can be obtained from the basic elementary functions x, $\sqrt[n]{x}$, e^x, $\ln x$, $\sin x$ and $\sin^{-1} x$ using addition, subtraction, multiplication, division, and composition of functions).

1.37 Theorem: *(The Sequential Characterization of Continuity) Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and let $a \in A$. Then f is continuous at a if and only if for every sequence (x_k) in A with $x_k \to a$ we have $f(x_k) \to f(a)$.*

Proof: Suppose that f is continuous at a. Let (x_k) be a sequence in A with $x_k \to a$. Let $\epsilon > 0$. Choose $\delta > 0$ so that for all $x \in A$ we have $|x - a| < \delta \Longrightarrow |f(x) - f(a)| < \epsilon$. Choose $m \in \mathbb{Z}$ so that for all indices k we have $k \geq m \Longrightarrow |x_k - a| < \delta$. Then when $k \geq m$ we have $|x_k - a| < \delta$ and hence $|f(x_k) - f(a)| < \epsilon$. Thus we have $f(x_k) \to f(a)$.

Conversely, suppose that f is not continuous at a. Choose $\epsilon_0 > 0$ so that for all $\delta > 0$ there exists $x \in A$ with $|x - a| < \delta$ and $|f(x) - f(a)| \geq \epsilon_0$. For each $k \in \mathbb{Z}^+$, choose $x_k \in A$ with $|x_k - a| \leq \frac{1}{k}$ and $|f(x_k) - f(a)| \geq \epsilon_0$. Consider the sequence (x_k) in A (we remark that the Axiom of Choice is being used here). Since $|x_k - a| \leq \frac{1}{k}$ for all $k \in \mathbb{Z}^+$, it follows that $x_k \to a$. Since $|f(x_k) - f(a)| \geq \epsilon_0$ for all $k \in \mathbb{Z}^+$, it follows that $f(x_k) \not\to f(a)$.

1.38 Theorem: *(Intermediate Value Theorem) Let I be an interval in $\mathbb{R}$ and let $f : I \to \mathbb{R}$ be continuous. Let $a, b \in I$ with $a \leq b$ and let $y \in \mathbb{R}$. Suppose that either $f(a) \leq y \leq f(b)$ or $f(b) \leq y \leq f(a)$. Then there exists $x \in [a, b]$ with $f(x) = y$.*

Proof: We prove the theorem in the case that $f(a) \leq y \leq f(b)$. If $y = f(a)$ then we can take $x = a$ and if $y = f(b)$ then we can take $x = b$. Suppose that $f(a) < y < f(b)$. Let $A = \big\{ t \in [a, b] \big| f(t) \leq y \big\}$. Note that $A \neq \emptyset$ (since $a \in A$) and A is bounded above (by b) and so A has a supremum in $\mathbb{R}$. Let $x = \sup A$. Since $a \in A$ and $x = \sup A$ we have $x \geq a$. Since b is an upper bound for A and $x = \sup A$ we have $x \leq b$. Thus $x \in [a, b]$.

We claim that $f(x) = y$. Suppose, for a contradiction, that $f(x) > y$. Since $x \neq a$ (because $f(a) < y$ but $f(x) > y$) we can choose $\delta_1 > 0$ so that $(x - \delta_1, x] \subseteq [a, b]$. Since f is continuous at x with $f(x) > y$, we can choose δ_2 so that for all $t \in [a, b]$ we have $|t - x| < \delta_2 \Longrightarrow f(t) > y$. Let $\delta = \min\{\delta_1, \delta_2\}$. Since $x = \sup A$, by the Approximation Property we can choose $t \in A$ with $x - \delta < t \leq x$. Since $t \in A$ we have $f(t) \leq y$, but since $t \in (x - \delta, x]$ we have $f(t) > y$, so we have obtained the desired contradiction. Now suppose, for a contradiction, that $f(x) < y$. Since $x \neq b$ (because $f(b) > y$ but $f(x) < y$) we can choose $\delta_1 > 0$ so that $[x, x + \delta_1) \subseteq [a, b]$. Since f is continuous at x with $f(x) < y$ we can choose $\delta_2 > 0$ so that for all $t \in [a, b]$ we have $|t - x| < \delta_2 \Longrightarrow f(t) < y$. Let $\delta = \min\{\delta_1, \delta_2\}$ so that $[x, x + \delta) \subseteq [a, b]$ and for all $t \in [x, x + \delta)$ we have $f(t) < y$. But then $x + \delta \in A$ so we cannot have $x = \sup A$, and we have obtained the desired contradiction.

1.39 Definition: Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$. For $a \in A$, if $f(a) \geq f(x)$ for every $x \in A$, then we say that $f(a)$ is the **maximum value** of f and that f attains its maximum value at a. Similarly for $b \in A$, if $f(b) \leq f(x)$ for every $x \in A$ then we say that $f(b)$ is the **minimum value** of f (in A) and that f attains its minimum value at b. We say that f attains its **extreme values** in A when f attains its maximum value at some point $a \in A$ and f attains its minimum value at some point $b \in A$.

1.40 Theorem: *(Extreme Value Theorem) Let $a, b \in \mathbb{R}$ with $a < b$, and let $f : [a, b] \to \mathbb{R}$ be continuous. Then f attains its extreme values in $[a, b]$.*

Proof: We prove that f attains its maximum. First we claim that f is bounded above. Suppose, for a contradiction, that it is not. For each $k \in \mathbb{Z}^+$, choose $x_k \in [a, b]$ such that $f(x_k) \geq k$. By the Bolzano Weierstrass Theorem, we can choose a convergent subsequence (x_{k_j}). Let $p = \lim\limits_{j\to\infty} x_{k_j}$. Note that $p \in [a, b]$ by Comparison (since $x_{k_j} \geq a$ for all j we have $p \geq a$, and since $x_{k_j} \leq b$ for all j we have $p \leq b$). Since $f(x_{k_j}) \geq k_j$ and $k_j \to \infty$ we must have $f(x_{k_j}) \to \infty$ as $j \to \infty$. But by the Sequential Characterization of Continuity, we should have $f(x_{k_j}) \to f(p) \in \mathbb{R}$, so we have obtained the desired contradiction. Thus f is bounded above, as claimed.

Since the range $f([a, b])$ is nonempty and bounded above, it has a supremum. Let $m = \sup f([a, b])$. By the Approximation Property of the supremum, for each $k \in \mathbb{Z}^+$ we can choose $y_k \in [a, b]$ such that $m - \frac{1}{k} \leq f(y_k) \leq m$. By the Bolzano Weierstrass Theorem, we can choose a convergent subsequence (y_{k_j}). Let $c = \lim\limits_{j\to\infty} y_{k_j}$. Since we have $m - \frac{1}{k_j} \leq f(y_{k_j}) \leq m$ and $\frac{1}{k_j} \to 0$, we have $f(y_{k_j}) \to m$ as $j \to \infty$ by the Squeeze Theorem. Since f is continuous at c, by the Sequential Characterization of Continuity we have $f(y_{k_j}) \to f(c)$, and so by the Uniqueness of Limits, we have $f(c) = m$. Thus f attains its maximum value at c.

1.41 Definition: Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$. We say that f is **uniformly continuous** (on A) when

$$\forall \epsilon > 0 \;\; \exists \delta > 0 \;\; \forall a \in A \;\; \forall x \in A \big(|x - a| < \delta \implies |f(x) - f(a)| < \epsilon \big).$$

1.42 Example: Define $f : (0, \infty) \to (0, \infty)$ by $f(x) = \frac{1}{x}$. Let $\epsilon = 1$. Let $\delta > 0$. If $\delta \geq 1$ then for $x = \frac{1}{3}$ and $a = 1$ we have $|x - a| = \frac{2}{3} < \delta$ but $|f(x) - f(a)| = 2 > \epsilon$. If $0 < \delta < 1$ then for $x = \frac{\delta}{3}$ and $a = \delta$ we have $|x - a| = \frac{2}{3}\delta < \delta$ but $|f(x) - f(a)| = \frac{2}{\delta} > 2 > \epsilon$. This proves that f is not uniformly continuous (but f is continuous because it is elementary).

1.43 Theorem: *(Closed Bounded Intervals and Uniform Continuity) Let $a, b \in \mathbb{R}$ with $a < b$ and let $f : [a, b] \to \mathbb{R}$. If f is continuous then f is uniformly continuous (on $[a, b]$).*

Proof: Suppose, for a contradiction, that $f : [a, b] \to \mathbb{R}$ is continuous but not uniformly continuous on $[a, b]$. Choose $\epsilon > 0$ so that for all $\delta > 0$ there exist $x, y \in [a, b]$ such that $|x - y| < \delta$ but $|f(x) - f(y)| \geq \epsilon$. For each $k \in \mathbb{Z}^+$ choose x_k and y_k in $[a, b]$ with $|x_k - y_k| \leq \frac{1}{k}$ and $|f(x_k) - f(y_k)| \geq \epsilon$. By the Bolzano Weierstrass Theorem, we can choose a convergent subsequence (y_{k_j}) of (y_k). Let $c = \lim\limits_{j\to\infty} y_{k_j}$. For all j we have $|x_{k_j} - y_{k_j}| \leq \frac{1}{k_j}$ hence $y_{k_j} - \frac{1}{k_j} \leq x_{k_j} \leq y_{k_j} + \frac{1}{k_j}$. Since $y_{k_j} \to c$ and $\frac{1}{k_j} \to 0$ we have $y_{k_j} \pm \frac{1}{k_j} \to c$ and hence $x_{k_j} \to c$ by the Squeeze Theorem. Since f is continuous at c and $x_{k_j} \to c$ and $y_{k_j} \to c$, we have $f(x_{k_j}) \to f(c)$ and $f(y_{k_j}) \to f(c)$ by the Sequential Characterization of Continuity. Since $f(x_{k_j}) \to c$ and $f(y_{k_j}) \to c$ we have $f(x_{k_j}) - f(y_{k_j}) \to 0$. But this implies that we can choose j so that $|f(x_{k_j}) - f(y_{k_j})| < \epsilon$, giving the desired contradiction.

Differentiation in $\mathbb{R}$

1.44 Definition: For a subset $A \subseteq \mathbb{R}$, we say that A is **open** when it is a union of open intervals. Let $A \subseteq \mathbb{R}$ be open, let $f : A \to \mathbb{R}$. For $a \in A$, we say that f is **differentiable** at a when the limit

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists in $\mathbb{R}$. In this case we call the limit the **derivative** of f at a, and we denote to by $f'(a)$, so we have

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

We say that f is **differentiable** (on A) when f is differentiable at every point $a \in A$. In this case, the **derivative** of f is the function $f' : A \to \mathbb{R}$ defined by

$$f'(x) = \lim_{u \to x} \frac{f(u) - f(x)}{u - x}.$$

When f' is differentiable at a, denote the derivative of f' at a by $f''(a)$, and we call $f''(a)$ the **second derivative** of f at a. When $f''(a)$ exists for every $a \in A$, we say that f is **twice differentiable** (on A), and the function $f'' : A \to \mathbb{R}$ is called the **second derivative** of f. Similarly, $f'''(a)$ is the derivative of f'' at a and so on.

1.45 Remark: Note that

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

To be precise, the limit on the left exists in $\mathbb{R}$ if and only if the limit on the right exists in $\mathbb{R}$, and in this case the two limits are equal.

1.46 Note: The student should be familiar with derivatives from first year calculus, and should be able to calculate the derivatives of elementary functions using differentiation rules including the Product Rule, the Quotient Rule and the Chain Rule. We shall provide proofs of some of the theorems whose proofs are often omitted in calculus courses.

1.47 Exercise: Let $A \subseteq \mathbb{R}$ be open, let $f : A \to \mathbb{R}$, and let $a \in A$. Show that f is differentiable at a with derivative $f'(a)$ if and only if

$$\forall \epsilon > 0\ \exists \delta > 0\ \forall x \in A\ \Big(|x - a| \leq \delta \Longrightarrow \big| f(x) - f(a) - f'(a)(x - a) \big| \leq \epsilon\, |x - a| \Big).$$

1.48 Theorem: *(Differentiability Implies Continuity) Let $A \subseteq \mathbb{R}$ be open, let $f : A \to \mathbb{R}$ and let $a \in A$. If f is differentiable at a then f is continuous at a.*

Proof: The proof is left as an exercise (the proof is often given in first year calculus).

1.49 Theorem: *(Chain Rule) Let $A, B \subseteq \mathbb{R}$ be open, let $f : A \to B$, let $g : B \to \mathbb{R}$ and let $h = g \circ f : A \to \mathbb{R}$. Let $a \in A$ and let $b = f(a) \in B$. Suppose that f is differentiable at a and g is differentiable at b. Then h is differentiable at a with*

$$h'(a) = g'\big(f(a)\big)\, f'(a).$$

Proof: We shall use the ϵ-δ formulation of the derivative given in Exercise 1.48. Note first that for $x \in A$ and $y = f(x) \in B$ we have

$$\begin{aligned}
&\big|h(x) - h(a) - g'(f(a))f'(a)(x-a)\big| \\
&\quad = \big|g(f(x)) - g(f(a)) - g'(f(a))f'(a)(x-a)\big| \\
&\quad = \big|g(y) - g(b) - g'(b)f'(a)(x-a)\big| \\
&\quad = \big|g(y) - g(b) - g'(b)(y-b) + g'(b)(y-b) - g'(b)f'(a)(x-a)\big| \\
&\quad \leq \big|g(y) - g(b) - g'(b)(y-b)\big| + |g'(b)|\,|y - b - f'(a)(x-a)| \\
&\quad = \big|g(y) - g(b) - g'(b)(y-b)\big| + |g'(b)|\,\big|f(x) - f(a) - f'(a)(x-a)\big|
\end{aligned}$$

and also

$$\begin{aligned}
|y-b| = \big|f(x)-f(a)\big| &= \big|f(x) - f(a) - f'(a)(x-a) + f'(a)(x-a)\big| \\
&\leq \big|f(x) - f(a) - f'(a)(x-a)\big| + |f'(a)|\,|x-a|.
\end{aligned}$$

Let $\epsilon > 0$. Since g is differentiable at b, we can choose $\delta_0 > 0$ so that

$$|y-b| \leq \delta_0 \Longrightarrow \big|g(y) - g(b) - g'(b)(y-b)\big| \leq \tfrac{\epsilon}{2(1+|f'(a)|)}\,|y-b|.$$

Since f is continuous at a (by Theorem 1.49), we can choose δ_1 so that

$$|x-a| \leq \delta_1 \Longrightarrow |f(x) - f(a)| \leq \delta_0 \Longrightarrow |y-b| \leq \delta_0.$$

Since f is differentiable at a we can choose $\delta_2 > 0$ and $\delta_3 > 0$ so that

$$\begin{aligned}
&|x-a| \leq \delta_2 \Longrightarrow \big|f(x) - f(a) - f'(a)(x-a)\big| \leq |x-a| \text{ and} \\
&|x-a| \leq \delta_3 \Longrightarrow \big|f(x) - f(a) - f'(a)(x-a)\big| \leq \tfrac{\epsilon}{2(1+g'(b)|)}\,|x-a|.
\end{aligned}$$

Let $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Let $x \in A$ and let $y = f(x) \in B$. Then when $|x-a| \leq \delta$ we have

$$\begin{aligned}
&\big|h(x) - h(a) - g'(f(a))f'(a)(x-a)\big| \\
&\quad \leq \big|g(y) - g(b) - g'(b)(y-b)\big| + |g'(b)|\,\big|f(x) - f(a) - f'(a)(x-a)\big| \\
&\quad \leq \tfrac{\epsilon}{2(1+|f'(a)|)}\,|y-b| + (1+|g'(b)|)\cdot \tfrac{\epsilon}{2(1+|g'(b)|)}\,|x-a| \\
&\quad \leq \tfrac{\epsilon}{2(1+|f'(a)|)}\Big(\big|f(x) - f(a) - f'(a)(x-a)\big| + |f'(a)|\,|x-a|\Big) + \tfrac{\epsilon}{2}\,|x-a| \\
&\quad \leq \tfrac{\epsilon}{2(1+|f'(a)|)}\Big(|x-a| + |f'(a)|\,|x-a|\Big) + \tfrac{\epsilon}{2}\,|x-a| \\
&\quad = \tfrac{\epsilon}{2}\,|x-a| + \tfrac{\epsilon}{2}\,|x-a| = \epsilon|x-a|.
\end{aligned}$$

Thus h is differentiable at a with $h'(a) = g'(f(a))f'(a)$, as required.

1.50 Exercise: Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$. Show that f is strictly monotonic if and only if f has the property that for all $a, b, c \in A$, if b lies strictly between a and c then $f(b)$ lies strictly between $f(a)$ and $f(c)$.

1.51 Theorem: *(The Inverse Function Theorem) Let A be an interval in $\mathbb{R}$, let $f : A \to \mathbb{R}$ be injective and continuous, let $B = f(A)$ (and note that $f : A \to B$ is bijective) and let $g = f^{-1} : B \to A$ be the inverse function. Then*

(1) the functions f and g are strictly monotonic and g is continuous, and

(2) if A is an open interval then so is B, and if f is differentiable at $a \in A$ with $f'(a) \neq 0$, then g is differentiable at $b = f(a)$ with $g'(b) = \frac{1}{f'(a)}$.

Proof: To prove Part 1, suppose that f is injective and continuous. Let $a, b, c \in A$ with $a < b < c$. Since f is injective and $a \neq c$, we have $f(a) \neq f(c)$, so either $f(a) < f(c)$ or $f(a) > f(c)$. Consider the case that $f(a) < f(c)$. Suppose, for a contradiction, that $f(b) \geq f(c)$. Note that since f is injective and $b \neq c$ we have $f(b) \neq f(c)$ and so $f(b) > f(c)$. Choose y with $f(c) < y < f(b)$. Since f is continuous on $[a, b]$ and on $[b, c]$, by the Intermediate Value Theorem, we can choose $x_1 \in [a, b]$ and $x_2 \in [b, c]$ with $f(x_1) = y = f(x_2)$. Since $y \neq f(b)$ we cannot have $x_1 = b$ or $x_2 = b$ so we have $x_1 < b < x_2$ with $f(x_1) = f(x_2)$, which contradicts the fact that f is injective. Thus we cannot have $f(b) \geq f(c)$ so we have $f(b) < f(c)$. A similar argument shows that we cannot have $f(b) \leq f(a)$ so we must have $f(b) > f(a)$. This proves that in the case that $f(a) < f(c)$ we have $f(a) < f(b) < f(c)$. A similar argument shows that in the case that $f(a) > f(c)$ we have $f(a) > f(b) > f(c)$. It follows that f is strictly monotonic, by Exercise 1.51. It is easy to see that if f is strictly increasing then g is strictly increasing (indeed when $u, v \in B$ with $u < v$ and $a = g(u)$ and $b = g(v)$, we must have $a < b$ because if $a = b$ then $u = v$ and if $a > b$ then $u > v$ since f is strictly increasing) and if f is strictly decreasing then g is strictly decreasing.

To complete the proof of Part 1, it remains to show that g is continuous. Suppose that f and g are strictly increasing (the case that f and g are strictly decreasing is similar). Let $b \in B$ and let $a = g(b)$ so that $f(a) = b$. Since f and g are strictly increasing, it follows that b is the left (or right) endpoint of B if and only if a is the left (or right) endpoint of A. To show that g is continuous at b, it suffices to show that if b is not the right endpoint of B then $\lim_{y \to b^+} g(y) = g(b)$ and that if b is not the left endpoint then $\lim_{y \to b^-} g(y) = g(b)$. We shall prove the first of these two statements (the proof of the second is similar). Suppose that b is not the right endpoint of B and hence a is not the right endpoint of A. Let $\epsilon > 0$ be small enough that $a + \epsilon \in A$. Choose $\delta = f(a + \epsilon) - b = f(a + \epsilon) - f(a)$ and note that $\delta > 0$ since f is strictly increasing. Then for all $y \in B$, if $b < y < b + \delta$ then $a = g(b) < g(y) < g(b + \delta) = g(f(a + \epsilon)) = a + \epsilon$. Thus $\lim_{y \to b^+} g(y) = g(b)$, as required.

To prove Part 2, suppose that A is an open interval and that f is differentiable at $a \in A$ with $f'(a) \neq 0$. Note that B is an interval by Theorem 1.39 and B is open because, as mentioned above, if $u \in B$ was a right or left endpoint of B then $g(u)$ would be a right or left endpoint of A. By Part 1, we know that g is continuous at $b = f(a)$, and so as $y \to b$ in B we have $g(y) \to g(b)$ in A, and so for $x = g(y)$ we have

$$\frac{g(y) - g(b)}{y - b} = \frac{x - a}{f(x) - f(a)} = \frac{1}{\frac{f(x) - f(a)}{x - a}} \longrightarrow \frac{1}{f'(a)} \quad \text{as } y \to b.$$

1.52 Definition: Let $f : A \subseteq \mathbb{R} \to \mathbb{R}$ and let $a \in A$. We say that f has a **local maximum** value at a when $\exists \delta > 0\ \forall x \in A\ \Big(|x-a| \le \delta \Longrightarrow f(x) \le f(a)\Big)$. Similarly, we say that f has a **local minimum** value at a when $\exists \delta > 0\ \forall x \in A\ \Big(|x-a| \le \delta \Longrightarrow f(x) \ge f(a)\Big)$.

1.53 Theorem: *(Fermat's Theorem) Let $A \subseteq \mathbb{R}$ be open, let $f : A \to \mathbb{R}$, and let $a \in A$. Suppose that f is differentiable at a and that f has a local maximum or minimum value at a. Then $f'(a) = 0$.*

Proof: The proof is left as an exercise (you probably saw the proof in first year calculus).

1.54 Theorem: *(Mean Value Theorems) Let $a, b \in \mathbb{R}$ with $a < b$.*

(1) (Rolle's Theorem) If $f : [a,b] \to \mathbb{R}$ is differentiable in (a,b) and continuous at a and b with $f(a) = 0 = f(b)$ then there exists a point $c \in (a,b)$ such that $f'(c) = 0$.

(2) (The Mean Value Theorem) If $f : [a,b] \to \mathbb{R}$ is differentiable in (a,b) and continuous at a and b then there exists a point $c \in (a,b)$ with

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

(3) (Cauchy's Mean Value Theorem) If $f, g : [a,b] \to \mathbb{R}$ are differentiable in (a,b) and continuous at a and b, then there exists a point $c \in (a,b)$ such that

$$f'(c)\big(g(b) - g(a)\big) = g'(c)\big(f(b) - f(a)\big).$$

Proof: To Prove Rolle's Theorem, let $f : [a,b] \to \mathbb{R}$ be differentiable in (a,b) and continuous at a and b with $f(a) = 0 = f(b)$. If f is constant, then $f'(x) = 0$ for all $x \in [a,b]$, so we can choose any $c \in (a,b)$ and we have $f'(c) = 0$. Suppose that f is not constant. Either $f(x) > 0$ for some $x \in (a,b)$ or $f(x) < 0$ for some $x \in (a,b)$. Suppose that $f(x) > 0$ for some $x \in (a,b)$ (the case that $f(x) < 0$ for some $x \in (a,b)$ is similar). By the Extreme Value Theorem, f attains its maximum value at some point, say $c \in [a,b]$. Since $f(x) > 0$ for some $x \in (a,b)$, we must have $f(c) > 0$. Since $f(a) = f(b) = 0$ and $f(c) > 0$, we have $c \in (a,b)$. By Fermat's Theorem, we have $f'(c) = 0$. This completes the proof of Rolle's Theorem.

To prove the Mean Value Theorem, suppose that $f : [a,b] \to \mathbb{R}$ is differentiable in (a,b) and continuous at a and b. Let $g(x) = f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x-a)$. Then g is differentiable in (a,b) with $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$ and g is continuous at a and b with $g(a) = 0 = g(b)$. By Rolle's Theorem, we can choose $c \in (a,b)$ so that $g'(c) = 0$, and then $g'(c) = \frac{f(b)-f(a)}{b-a}$, as required.

Finally, we use the Mean Value Theorem to Prove Cauchy's Mean Value Theorem. Suppose $f, g : [a,b] \to \mathbb{R}$ are both differentiable in (a,b) and continuous at a and b. Let $h(x) = f(x)\big(g(b) - g(a)\big) - g(x)\big(f(b) - f(a)\big)$. Then h is differentiable in (a,b) and continuous at a and b with $h(a) = f(a)g(b) - g(a)f(b) = h(b)$. By the Mean Value Theorem, we can choose $c \in (a,b)$ so that $h'(c) = \frac{h(b)-h(a)}{b-a} = 0$, and then we have $f'(c)\big(g(b) - g(a)\big) - g'(c)\big(f(b) - f(a)\big) = 0$, as required.

1.55 Corollary: *Let $a, b \in \mathbb{R}$ with $a < b$. Let $f : [a,b] \to \mathbb{R}$. Suppose that f is differentiable in (a,b) and continuous at a and b. If $f'(x) > 0$ for all $x \in (a,b)$ then f is strictly increasing on $[a,b]$.*

Proof: The proof is left as an exercise (the proof is often given in first year calculus).

1.56 Theorem: *(l'Hôpital's Rule) Let A be a nonempty open interval in $\mathbb{R}$. Let $a \in A$, or let a be an endpoint of A. Let $f, g : A \setminus \{a\} \to \mathbb{R}$. Suppose that f and g are differentiable in $A \setminus \{a\}$ with $g'(x) \neq 0$ for all $x \in A \setminus \{a\}$. Suppose either that $\lim\limits_{x \to a} f(x) = 0 = \lim\limits_{x \to a} g(x)$ or that $\lim\limits_{x \to a} g(x) = \pm\infty$. Suppose that $\lim\limits_{x \to a} \dfrac{f'(x)}{g'(x)} = u \in \mathbb{R}$. Then $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = u$.*

Similar results hold for limits $x \to a^+$, $x \to a^-$, $x \to \infty$ and $x \to -\infty$ and also when the limit is $u = \pm\infty$.

Proof: We give the proof for $x \to a^+$ (assuming that $a \in A$ or a is the left endpoint of A) with $u \in \mathbb{R}$. Suppose first that $\lim\limits_{x \to a^+} f(x) = 0 = \lim\limits_{x \to a^+} g(x)$. Choose $b \in A$ with $a < b$. Extend the maps f and g to obtain maps $f, g : [a, b] \to \mathbb{R}$ by defining $f(a) = 0 = g(a)$. Note that f and g are continuous at a since $\lim\limits_{x \to a^+} f(x) = 0$ and $\lim\limits_{x \to a^+} g(x) = 0$. Let (x_k) be a sequence in $(a, b]$ with $x_k \to a$. For each index k, by Cauchy's Mean Value Theorem we can choose $c_k \in (a, x_k)$ so that $f'(c_k)\big(g(x_k) - g(a)\big) = g'(c_k)\big(f(x_k) - f(a)\big)$. Since $f(a) = 0 = g(a)$, this simplifies to $f'(c_k)g(x_k) = g'(c_k)f(x_k)$ and so we have $\frac{f(x_k)}{g(x_k)} = \frac{f'(c_k)}{g'(c_k)}$. Since $a < c_k < x_k$ and $x_k \to a$, we have $c_k \to a$ by the Squeeze Theorem. Since $\lim\limits_{x \to a^+} \dfrac{f'(x)}{g'(x)} = u$ and $c_k \to a$, we have $\frac{f(x_k)}{g(x_k)} = \frac{f'(c_k)}{g'(c_k)} \to u$ by the Sequential Characterization of Limits. We have shown that for every sequence (x_k) in $(a, b]$ with $x_k \to a$ we have $\frac{f(x_k)}{g(x_k)} \to u$. By the Sequential Characterization of limits, it follows that $\lim\limits_{x \to a^+} \dfrac{f(x)}{g(x)} = u$.

Now suppose that $\lim\limits_{x \to a^+} g(x) = \infty$. Since $\lim\limits_{x \to a^+} g(x) = \infty$ we can choose $b \in A$ with $b > a$ so that $g(x) > 0$ for all $x \in (a, b]$. Let (x_k) be a sequence in $(a, b]$ with $x_k \to a$. For each pair of indices k, l, by Cauchy's Mean Value Theorem we can choose $c_{kl} \in (a, x_k)$ so that $f'(c_{kl})\big(g(x_k) - g(x_l)\big) = g'(c_{kl})\big(f(x_k) - f(x_l)\big)$. Divide both sides by $g'(c_{kl})g(x_l)$ to get

$$\frac{f'(c_{kl})}{g'(c_{kl})} \frac{g(x_k)}{g(x_l)} - \frac{f'(c_{kl})}{g'(c_{kl})} = \frac{f(x_k)}{g(x_l)} - \frac{f(x_l)}{g(x_l)}.$$

so we have

$$\frac{f(x_l)}{g(x_l)} = \frac{f'(c_{kl})}{g'(c_{kl})} + \frac{f(x_k)}{g(x_l)} - \frac{f'(c_{kl})}{g'(c_{kl})} \frac{g(x_k)}{g(x_l)}.$$

Let $\epsilon > 0$. Since $\lim\limits_{x \to a} \dfrac{f'(x)}{g'(x)} = u$ we can choose $\delta > 0$ so that $|x - a| \leq \delta \Longrightarrow \left| \frac{f'(x)}{g'(x)} - u \right| \leq \frac{\epsilon}{3}$. Since $x_k \to a$ we can choose $m \in \mathbb{Z}^+$ so $k \geq m \Longrightarrow |x_k - a| \leq \delta$. Note that when $k, l \geq m$, since c_{kl} lies between x_k and x_l we also have $|c_{kl} - a| \leq \delta$ so $\left| \frac{f'(c_{kl})}{g'(c_{kl})} - u \right| \leq \min\left\{1, \frac{\epsilon}{3}\right\}$. Fix $k \geq m$. Choose l large enough so that $\left| \frac{f(x_k)}{g(x_l)} \right| \leq \frac{\epsilon}{3}$ and $\left| \frac{f'(c_{kl})}{g'(c_{kl})} \frac{g(x_k)}{g(x_l)} \right| \leq \frac{\epsilon}{3}$. Then we have

$$\left| \frac{f(x_l)}{g(x_l)} - u \right| \leq \left| \frac{f'(c_{kl})}{g'(c_{kl})} - u \right| + \left| \frac{f(x_k)}{g(x_l)} \right| + \left| \frac{f'(c_{kl})}{g'(c_{kl})} \frac{g(x_k)}{g(x_l)} \right| \leq \epsilon.$$

Appendix 1. Mathematical Induction

1.57 Theorem: *(Induction Principle) Let $m \in \mathbb{Z}$. Let $F(n)$ be a statement about n. Suppose that $F(m)$ is true, and suppose that for all $k \in \mathbb{Z}$ with $k > m$, if $F(k-1)$ is true then $F(k)$ is true. Then $F(n)$ is true for all $n \in \mathbb{Z}$ with $n \geq m$.*

Proof: Let $S = \big\{k\in\mathbb{Z}\big|k\geq m \text{ and } F(k) \text{ is false}\big\}$. To prove that $F(n)$ is true for all $n \geq m$, we shall prove that $S = \emptyset$. Suppose, for a contradiction, that $S \neq \emptyset$. Since $S \neq \emptyset$ and S is bounded below by m, it follows from the Well-Ordering Property of $\mathbb{Z}$ that S has a minimum element. Let $k = \min(S)$. Since $k \in S$ it follows that $k \geq m$ and $F(k)$ is false. Since $F(m)$ is true and $F(k)$ is false, it follows that $k \neq m$, so we have $k > m$. We claim that $F(k-1)$ is true. Suppose, for a contradiction, that $F(k-1)$ is false. Since $k-1 \geq m$ and $F(k-1)$ is false, it follows that $k-1 \in S$. Since $k = \min(S)$ and $k-1 \in S$, we have $k \leq k-1$ giving the desired contradiction (to the assumption that $F(k-1)$ is false). Thus $F(k-1)$ is true, as claimed. Since $k > m$ and $F(k-1)$ is true, it follows by the hypothesis in the statement of the theorem that $F(k)$ is true. But, as mentioned earlier, since $k \in S$ we know that $F(k)$ is false, so we have obtained the desired contradiction (to the assumption that $S \neq \emptyset$). Thus $S = \emptyset$, as required.

1.58 Note: It follows, from the above theorem, that in order to prove that $F(n)$ is true for all $n \geq m$, we can do the following:

1. Prove that $F(m)$ is true (this is called proving the **base case**).
2. Let $n > m$ and suppose that $F(n-1)$ is true (this is called the **induction hypothesis**).
3. Prove that $F(n)$ is true.

1.59 Theorem: *(Strong Induction Principle) Let $m\in\mathbb{Z}$. Let $F(n)$ be a statement about n. Suppose that for all $n \in \mathbb{Z}$ with $n \geq m$, if $F(k)$ is true for all $k \in \mathbb{Z}$ with $m \leq k < n$ then $F(n)$ is true. Then $F(n)$ is true for all $n \in \mathbb{Z}$ with $n \geq m$.*

Proof: Let $G(n)$ be the statement "$F(k)$ is true for all $m \leq k < n$". Note that $G(m)$ is true vacuously since there are no elements k with $m \leq k < m$. Let $n \in \mathbb{Z}$ with $n \geq m$ and suppose, inductively, that $G(n)$ is true, in other words that $F(k)$ is true for all $m \leq k < n$. It follows from the hypothesis of the theorem that $F(n)$ is true, and so we have $F(k)$ true for all $k \in \mathbb{Z}$ with $m \leq k \leq n$. By the Discreteness Property of $\mathbb{Z}$, it follows that $F(k)$ is true for all $k \in \mathbb{Z}$ with $m \leq k < n+1$, or equivalently that $G(n+1)$ is true. By the Induction Principle, it follows that $G(n)$ is true for all $n \in \mathbb{Z}$ with $n \geq m$. Let $n \in \mathbb{Z}$ with $n \geq m$. Since $G(n)$ is true, we know that $F(k)$ is true for all $k \in \mathbb{Z}$ with $m \leq k < n$. By the hypothesis of the theorem, it follows that $F(n)$ is true. Thus $F(n)$ is true for all $n \in \mathbb{Z}$ with $n \geq m$.

1.60 Note: In order to prove that $F(n)$ is true for all $n \geq m$, we can do the following:

1. Let $n \geq m$ and suppose that $F(k)$ is true for all $k \in \mathbb{Z}$ with $m \leq k < n$.
2. Prove that $F(n)$ is true.

Although strong induction, used as above, does not require the proof that $F(m)$ is true (the base case), there are situations in which one ore more base cases must be verified to make this method of proof valid. For example, if a sequence $(x_n)_{n\geq 1}$ is defined by specifying the values of x_1 and x_2 and by giving a recursion formula for x_n in terms of x_{n-1} and x_{n-2} for all $n \geq 3$, then in order to prove that x_n satisfies the closed-form formula $x_n = f(n)$ for all $n \geq 1$ it suffices to prove that $x_1 = f(1)$ and $x_2 = f(2)$ (two base cases) and to prove that for all $n \geq 3$, if $x_{n-1} = f(n-1)$ and $x_{n-2} = f(x_{n-2})$ then $x_n = f(n)$.

1.61 Example: Let $a_0 = 0$ and $a_1 = 1$ and for $n \geq 2$ let $a_n = a_{n-1} + 6a_{n-2}$. Show that $a_n = \frac{1}{5}\big(3^n - (-2)^n\big)$ for all $n \geq 0$.

Solution: We claim that $a_n = \frac{1}{5}\big(3^n - (-2)^n\big)$ for all $n \geq 0$. When $n = 0$ we have $a_n = a_0 = 0$ and $\frac{1}{5}\big(3^n - (-2)^n\big) = \frac{1}{5}\big(3^0 - (-2)^0\big) = 0$, so the claim is true when $n = 0$. When $n = 1$ we have $a_n = a_1 = 1$ and $\frac{1}{5}\big(3^n - (-2)^n\big) = \frac{1}{5}\big(3 - (-2)\big) = 1$, so the claim is true when $n = 1$. Let $n \geq 2$ and suppose the claim is true for all $k < n$. In particular we suppose the claim is true for $n-1$ and $n-2$, that is we suppose $a_{n-1} = \frac{1}{5}\big(3^{n-1} - (-2)^{n-1}\big)$ and $a_{n-2} = \frac{1}{5}\big(3^{n-2} - (-2)^{n-2}\big)$. Then

$$\begin{aligned}
a_n &= a_{n-1} + 6a_{n-2} \\
&= \tfrac{1}{5}\big(3^{n-1} - (-2)^{n-1}\big) + \tfrac{6}{5}\big(3^{n-2} - (-2)^{n-2}\big) \\
&= \big(\tfrac{1}{5} \cdot 3^{n-1} + \tfrac{6}{5} \cdot 3^{n-2}\big) - \big(\tfrac{1}{5}(-2)^{n-1} + \tfrac{6}{5}(-2)^{n-2}\big) \\
&= \big(\tfrac{3}{5} \cdot 3^{n-2} + \tfrac{6}{5} \cdot 3^{n-2}\big) - \big(-\tfrac{2}{5}(-2)^{n-2} + \tfrac{6}{5}(-2)^{n-2}\big) \\
&= \tfrac{9}{5} \cdot 3^{n-2} - \tfrac{4}{5}(-2)^{n-2} = \tfrac{1}{5} \cdot 3^n - \tfrac{1}{5}(-2)^n \\
&= \tfrac{1}{5}\big(3^n - (-2)^n\big) = \tfrac{1}{5}\big(3^n - (-2)^n\big)\,.
\end{aligned}$$

By Strong Induction, we have $a_n = \frac{1}{5}\big(3^n - (-2)^n\big)$ for all $n \geq 0$.

1.62 Note: Suppose that we choose k of n objects, When the objects are chosen with replacement (so that repetition is allowed) and the order of the chosen objects matters (so the chosen objects form an ordered k-tuple), the number of ways to choose k of n objects is equal to n^k (since we have n choices for each of the k objects). For example, the number of ways to roll 3 six-sided dice is equal to $6^3 = 216$.

When the objects are chosen without replacement (so that the k chosen objects are distinct) and the order matters, the number of ways to choose k of n objects is equal to $n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$ (since we have n choices for the first object and $n-1$ choices for the second object and so on). In particular, the number of ways to arrange n objects in order (to form an ordered n-tuple) is equal to $n!$.

When the objects are chosen without replacement and the order does not matter (so the chosen objects form a k-element set), the number of ways to choose k of n objects is equal to $\frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$ (since each k-element set can be ordered in $k!$ ways to form $k!$ ordered k-tuples, and there are $\frac{n!}{(n-k)!}$ such ordered k-tuples). For example, the number of 4-element subsets of the set $\{1,2,3,4,5,6,7\}$ is equal to $\frac{7!}{4!3!} = \frac{7\cdot6\cdot5\cdot4}{4\cdot3\cdot2\cdot1} = 7\cdot5 = 35$.

1.63 Definition: For $n, k \in \mathbb{N}$ with $0 \leq k \leq n$, we define the **binomial coefficient** $\binom{n}{k}$, read as "n choose k", by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}.$$

1.64 Theorem: *(Pascal's Triangle) For $k, n \in \mathbb{N}$ with $0 \leq k \leq n$ we have*

$$\binom{n}{0} = \binom{n}{n} = 1\,,\ \binom{n}{k} = \binom{n}{n-k} \text{ and } \binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

Proof: The formulas $\binom{n}{0} = \binom{n}{n} = 1$ and $\binom{n}{k} = \binom{n}{n-k}$ are immediate from the definition of $\binom{n}{k}$ (since $0! = 1$) and we have

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-k-1)!} = \frac{(k+1)n!}{(k+1)!(n-k)!} + \frac{(n-k)n!}{(k+1)!(n-k)!} \\
&= \frac{(k+1+n-k)n!}{(k+1)!(n-k)!} = \frac{(n+1)!}{(k+1)!((n+1)-(k+1))!} = \binom{n+1}{k+1}.
\end{aligned}$$

1.65 Exercise: Make a table displaying the values $\binom{n}{k}$ for $0 \leq k \leq n \leq 10$. The table forms a triangle of positive integers in which each entry is obtained by adding two of the entries above.

1.66 Notation: Let R be a ring and let $a \in R$. For $k \in \mathbb{Z}^+$ we write $ka = a + a + \cdots + a$ with k terms in the sum, and we write $(-k)a = k(-a)$, and we write $a^k = a \cdot a \cdot \ldots \cdot a$ with k terms in the product. For $0 \in \mathbb{Z}$ we write $0a = 0$ and $a^0 = 1$. When $a \in R$ is a unit, for $k \in \mathbb{Z}^+$ we write $a^{-k} = (a^{-1})^k$.

1.67 Exercise: Let R be a ring and let $a, b \in R$. Show that for all $k, l \in \mathbb{Z}$ we have $(-k)a = -(ka)$, $(k + l)a = ka + la$ and $(ka)(lb) = (kl)(ab)$. Show that for all $k, l \in \mathbb{Z}^+$ we have $a^{k+l} = a^k a^l$. Show that if $ab = ba$ then for all $k, l \in \mathbb{Z}^+$ we have $(ab)^k = a^k b^k$. Show that if a is a unit, then for all $k, l \in \mathbb{Z}$ we have $a^{-k} = (a^k)^{-1}$ and $a^{k+l} = a^k a^l$.

1.68 Theorem: *(Binomial Theorem) Let R be a ring, let $a, b \in R$ with $ab = ba$, and let $n \in \mathbb{N}$. Then*

$$\begin{aligned}(a+b)^n &= \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \\ &= \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \cdots + \binom{n}{n-1} a\, b^{n-1} + \binom{n}{n} b^n .\end{aligned}$$

Proof: We shall prove, by induction, that $(a + b)^n = \sum\limits_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$ for all $n \geq 0$.

When $n = 0$ we have $\sum\limits_{k=0}^{n} \binom{n}{k} a^{n-k} b^k = \binom{0}{0} a^0 b^0 = 1 = (a + b)^0 = (a + b)^n$.

When $n = 1$ we have $\sum\limits_{k=0}^{n} \binom{n}{k} a^{n-k} b^k = \binom{1}{0} a^1 b^0 + \binom{1}{1} a^0 b^1 = a + b = (a + b)^1 = (a + b)^n$.

Let $n \geq 1$ and suppose, inductively that $(a + b)^n = \sum\limits_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$. Then

$$\begin{aligned}(a+b)^{n+1} &= (a+b)(a+b)^n = (a+b) \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \\ &= (a+b)\Big(\binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \binom{n}{2} a^{n-2} b^2 + \cdots + \binom{n}{n-1} a\, b^{n-1} + \binom{n}{n} b^n \Big) \\ &= \binom{n}{0} a^{n+1} + \binom{n}{1} a^n b + \binom{n}{2} a^{n-1} b^2 + \cdots + \binom{n}{n-1} a^2\, b^{n-1} + \binom{n}{n} a\, b^n \\ &\qquad + \binom{n}{0} a^n b + \binom{n}{1} a^{n-1} b^2 + \cdots + \binom{n}{n-2} a^2 b^{n-1} + \binom{n}{n-1} a\, b^n + \binom{n}{n} b^{n+1} \\ &= a^{n+1} + \left(\binom{n}{0} + \binom{n}{1}\right) a^n b + \left(\binom{n}{1} + \binom{n}{2}\right) a^{n-1} b + \cdots \\ &\qquad + \left(\binom{n}{n-2} + \binom{n}{n-1}\right) a^2 b^{n-1} + \left(\binom{n}{n-1} + \binom{n}{n}\right) a\, b^n + b^{n+1} \\ &= \binom{n+1}{0} a^{n+1} + \binom{n+1}{1} a^n b + \binom{n+1}{2} a^{n-1} b^2 + \cdots + \binom{n+1}{n-1} a^2\, b^n + \binom{n+1}{n+1} a\, b^n \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k+1} a^{n+1-k} b^k\end{aligned}$$

as required, since $\binom{n}{0} = 1 = \binom{n+1}{0}$ and $\binom{n}{n} = 1 = \binom{n+1}{n+1}$ and $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$ for all k with $0 \leq k \leq n$. By induction, we have $(a + b)^n = \sum\limits_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$ for all $n \geq 0$.

Finally note that, by interchanging a and b, we also have $(a + b)^n = \sum\limits_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$.