MATH 147 Calculus 1

Lecture Notes

by Stephen New

Chapter 1: Sets, Fields and Orders

1.1 Definition: For sets A and B, we use the following notation. We write $x \in A$ when x is an **element** of the set A. We denote the **empty set**, that is the set with no elements, by $\emptyset$. We write $A = B$ when the sets A and B are **equal**, that is when A and B have the same elements. We write $A \subseteq B$ (some books write $A \subset B$) when A is a **subset** of B, that is when every element of A is also an element of B. We write $A \subset B$, or for emphasis $A \subsetneqq B$, when A is a **proper subset** of B, they is when $A \subseteq B$ but $A \neq B$. We denote the **union** of A and B by $A \cup B$, the **intersection** of A and B by $A \cap B$, the set A **remove** B by $A \setminus B$ and the **product** of A and B by $A \times B$, that is

$$\begin{aligned} A \cup B &= \big\{x \big| x \in A \text{ or } x \in B\big\}, \\ A \cap B &= \big\{x \big| x \in A \text{ and } x \in B\big\}, \\ A \setminus B &= \big\{x \big| x \in A \big| x \notin B\big\}, \text{ and} \\ A \times B &= \big\{(a,b) \big| x \in A \text{ and } b \in B\big\}. \end{aligned}$$

We say that A and B are **disjoint** when $A \cap B = \emptyset$.

1.2 Theorem: *(Properties of Sets) Let $A, B, C \subseteq X$. Then*
(1) (Idempotence) $A \cup A = A$, $A \cap A = A$,
(2) (Identity) $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$, $A \cup X = X$, $A \cap X = A$,
(3) (Associativity) $(A \cup B) \cup C = A \cup (B \cup C)$ and $(A \cap B) \cap C) = A \cap (B \cap C)$,
(4) (Commutativity) $A \cup B = B \cup A$ and $A \cap B = B \cap A$,
(5) (Distributivity) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,
(6) (De Morgan's Laws) $X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B)$ and $X \setminus (A \cap B) = (X \setminus A) \cup (X \setminus B)$.

Proof: The proof is left as an exercise.

1.3 Definition: We write $\mathbf{N} = \{0, 1, 2, \cdots\}$ for the set of **natural numbers** (which we take to include the number 0), $\mathbf{Z} = \{0, \pm 1, \pm 2, \cdots\}$ for the set of **integers**, $\mathbf{Q}$ for the set of **rational numbers** and we write $\mathbf{R}$ for the set of **real numbers**. We assume familiarity with the algebraic operations $+$, $-$, $\cdot$, $\div$ and with the order relations $<$, $\leq$, $>$, $\geq$ on these sets. Some of the fundamental properties of these operations and order relations are discussed in this chapter.

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

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

An **interval** in $\mathbf{R}$ 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** intervals. 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.

1.5 Definition: Let A and B be sets. A **relation** on $A \times B$ is a subset $r \subseteq A \times B$. When r is a relation on $A \times B$ and $a \in A$ and $b \in B$, we say that a and b are **related** under r and we write arb when $(a,b) \in r$. The **domain** and **range** of the relation r are the sets $\text{Domain}(r) = \{x \in A \big| xry \text{ for some } y \in B\}$ and $\text{Range}(r) = \{y \in B \big| xry \text{ for some } x \in A\}$.

1.6 Definition: Let A and B be sets. A **function** from A to B is a relation f on $A \times B$ with the property that for every $x \in A$ there exists a unique element $y \in B$ such that xfy. When f is a function from A to B, we write $f : A \to B$. When $f : A \to B$ and $x \in A$ we denote the unique element $y \in B$ for which xfy by $f(x)$. Note that $\text{Domain}(f) = A$ and $\text{Range}(f) \subseteq B$. A **binary operation** on A is a function $f : A \times A \to A$

1.7 Definition: A **field** is a set F with two distinct elements $0, 1 \in F$ and two binary operations $+$ and $\cdot$ such that

(1) (Additive Associativity) for all $x, y, z \in F$ we have $(x + y) + z = x + (y + z)$,
(2) (Additive Commutativity) for all $x, y \in F$ we have $x + y = y + x$,
(3) (Additive Identity) for all $x \in F$ we have $0 + x = x$,
(4) (Additive Inverse) for all $x \in F$ there exists a unique $y \in F$ such that $x + y = 0$,
(5) (Multiplicative Associativity) for all $x, y, z \in F$ we have $(x \cdot y) \cdot z = x \cdot (y \cdot z)$,
(6) (Multiplicative Commutativity) for all $x, y \in F$ we have $x \cdot y = y \cdot x$,
(7) (Multiplicative Identity) for all $x \in F$ we have $1 \cdot x = x$,
(8) (Multiplicative Inverse) for all $0 \neq x \in F$ there exists a unique $y \in F$ such that $x \cdot y = 1$.
(9) (Distributivity) for all $x, y, z \in F$ we have $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.

1.8 Theorem: *$\mathbf{Q}$ and $\mathbf{R}$ are fields.*

Proof: We omit the proof, but we remark that $\mathbf{Z}$ is not a field because it does not satisfy Property (8).

1.9 Notation: Let F be a field and let $a, b \in F$. We denote the unique additive inverse of a by $-a$ and we write $a - b = a + (-b)$. We usually write $a \cdot b$ simply as ab, and, when $a \neq 0$, we denote the unique multiplicative inverse of a by a^{-1} and we write $b \div a = \frac{b}{a} = b\,a^{-1}$.

1.10 Theorem: *Let F be a field. Then for all $x, y, z \in F$ we have*

(1) (Additive Cancellation) if $x + y = x + z$ then $y = z$,
(2) (Uniqueness of Additive Identity) if $x + y = x$ then $y = 0$,
(3) (Multiplicative Cancellation) if $xy = xz$ then either $x = 0$ or $y = z$,
(4) (Uniqueness of Multiplicative Identity) if $xy = x$ then $y = 1$,
(5) (No Zero Divisors) if $xy = 0$ then $x = 0$ or $y = 0$.

Proof: The proof is left as an exercise.

1.11 Theorem: *(Properties of Fields) Let F be a field. Then for all $x, y \in F$ we have $0 \cdot x = 0$, $-(-x) = x$, $-(x + y) = -x - y$, $(-1)\,x = -x$, $(-x)\,y = -(xy)$, $(-x)(-y) = xy$, $(a^{-1})^{-1} = a$, $(ab)^{-1} = a^{-1}b^{-1}$ and $(-a)^{-1} = -a^{-1}$.*

Proof: The proof is left as an exercise.

1.12 Definition: An **order** on a set X is a binary relation $\leq$ on X such that

(1) (Totality) for all $x, y \in X$, either $x \leq y$ or $y \leq x$,
(2) (Antisymmetry) for all $x, y \in X$, if $x \leq y$ and $y \leq x$ then $x = y$, and
(3) (Transitivity) for all $x, y, z \in X$, if $x < y$ and $y < z$ then $x < z$.

An **ordered set** is a set X with on order $\leq$.

1.13 Theorem: *Each of* $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{Q}$ *and* $\mathbf{R}$ *is an ordered set using its standard order* $\leq$. *Under the inclusions* $\mathbf{N} \subseteq \mathbf{Z} \subseteq \mathbf{Q} \subseteq \mathbf{R}$ *the orders coincide (so that for example when* $a, b \in \mathbf{N}$ *we have* $a \leq b$ *in* $\mathbf{N}$ *if and only if* $a \leq b$ *in* $\mathbf{R}$*).*

Proof: We omit the proof.

1.14 Notation: When $\leq$ is an order on X, we write $x < y$ when $x \leq y$ and $x \neq y$, we write $x \geq y$ when $y \leq x$ and we write $x > y$ when $y < x$.

1.15 Definition: An **ordered field** is a field F with an order $\leq$ such that for all $x, y, z \in F$

(1) if $x \leq y$ then $x + z \leq y + z$, and
(2) if $0 \leq x$ and $0 \leq y$ then $0 \leq xy$.

When F is an ordered field and $x \in F$ we say that x is **positive** when $x > 0$, we say x is **negative** when $x < 0$, we say x is **nonpositive** when $x \leq 0$, and we say x is **nonnegative** when $x \geq 0$.

1.16 Theorem: $\mathbf{Q}$ *and* $\mathbf{R}$ *are ordered fields.*

Proof: We omit the proof.

1.17 Theorem: *(Properties of Ordered Fields) Let* F *be an ordered field. Then for all* $x, y, z \in F$

(1) if $x > 0$ *then* $-x < 0$*, and if* $x < 0$ *then* $-x > 0$,
(2) if $x > 0$ *and* $y < z$ *then* $xy < xz$,
(3) if $x < 0$ *and* $y < z$ *then* $xy > xz$,
(4) if $x \neq 0$ *then* $x^2 > 0$*, and in particular* $1 > 0$*, and*
(5) if $0 < x < y$ *then* $0 < \frac{1}{y} < \frac{1}{x}$.

Proof: The proof is left as an exercise.

1.18 Definition: Let F be an ordered field. For $a \in F$ we define the **absolute value** of a to be

$$|a| = \begin{cases} a \text{ if } a \geq 0, \\ -a \text{ if } a \leq 0. \end{cases}$$

1.19 Theorem: *(Properties of Absolute Value) Let* F *be an ordered field. For all* $x, y \in F$

(1) (Positive Definiteness) $|x| \geq 0$ *with* $|x| = 0 \iff x = 0$,
(2) (Symmetry) $|x - y| = |y - x|$,
(3) (Multiplicativeness) $|xy| = |x|\,|y|$
(4) (Triangle Inequality) $\big||x| - |y|\big| \leq |x + y| \leq |x| + |y|$*, and*
(5) (Approximation) for $a, b \in F$ *with* $b \geq 0$ *we have* $|x - a| \leq b \iff a - b \leq x \leq a + b$.

Proof: The proof is left as an exercise.

1.20 Theorem: *(Induction Principle) Let $m \in \mathbf{Z}$. Let $F(n)$ be a statement about n. Suppose that*

(1) $F(m)$ is true, and
(2) for all $k \in \mathbf{Z}$ with $k \geq m$, if $F(k)$ is true then $F(k+1)$ is true.

Then $F(n)$ is true for all $n \in \mathbf{Z}$ with $n \geq m$.

Proof: We omit the proof.

1.21 Theorem: *(Basic Order Properties in $\mathbf{Z}$)*

(1) for $n \in \mathbf{Z}$ we have $n \in \mathbf{N}$ if and only if $n \geq 0$,
(2) for all $k, n \in \mathbf{Z}$ we have $k \leq n$ if and only if $k < n+1$.

Proof: We omit the proof.

1.22 Theorem: *(Strong Induction Principle) Let $m \in \mathbf{Z}$. Let $F(n)$ be a statement about n. Suppose that for all $n \in \mathbf{Z}$ with $n \geq m$, if $F(k)$ is true for all $k \in \mathbf{Z}$ with $m \leq k < n$ then $F(n)$ is true. Then $F(n)$ is true for all $n \in \mathbf{Z}$ 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 $n \in \mathbf{Z}$ 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 $k \in \mathbf{Z}$ with $m \leq k \leq n$. By the Basic Order Property (2), it follows that $F(k)$ is true for all $k \in \mathbf{Z}$ 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 $n \in \mathbf{Z}$ with $n \geq m$. Let $n \in \mathbf{Z}$ with $n \geq m$. Since $G(n)$ is true, we know that $F(k)$ is true for all $k \in \mathbf{Z}$ 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 $n \in \mathbf{Z}$ with $n \geq m$.

1.23 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 $n \geq 0$.

1.24 Definition: Let X be an ordered set and let $A \subseteq X$. We say that A is **bounded above** (in X) when there exists an element $b \in X$ such that $x \leq b$ for all $x \in A$, and in this case we say that b is an **upper bound** for A (in X).

We say that A is **bounded below** (in X) when there exists an element $a \in X$ such that $a \leq x$ for all $x \in A$, and in this case we say that a is a **lower bound** for A (in X). We say that A is **bounded** (in X) when A is bounded above and bounded below.

1.25 Definition: Let X be an ordered set and let $A \subseteq X$. We say that A has a **supremum** (or a **least upper bound**) (in X) when there exists an element $b \in X$ such that b is an upper bound for A with $b \leq c$ for every upper bound $c \in X$ for A, and in this case we say that b is the **supremum** (or the **least upper bound**) of A (in X) (note that if the supremum exists then it is unique by antisymmetry) 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**) (in X) when there exists an element $a \in X$ 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 (in X) 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 $a = \min A$.

1.26 Example: Let $A = (0, \infty)$ and $B = [1, \sqrt{2})$. 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 bellow. 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.

1.27 Theorem: *(Completeness Properties of* $\mathbf{R}$*)*

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

Proof: We omit the proof.

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

(1) If $b = \sup A$ *then for all* $0 < \epsilon \in \mathbf{R}$ *there exists* $x \in A$ *with* $b - \epsilon < x \leq b$*, and*
(2) if $a = \inf A$ *then for all* $0 < \epsilon \in \mathbf{R}$ *there exists* $x \in A$ *with* $a \leq x < a + \epsilon$*.*

Proof: Let $b = \sup A$. Let $\epsilon > 0$. Suppose, for a contradiction, that there is no element $x \in A$ with $b - \epsilon < x$, or equivalently that for all $x \in A$ we have $b - \epsilon \geq x$. Let $c = b - \epsilon$. Note that c is an upper bound for A since $x \leq b - \epsilon = c$ for all $x \in A$. Since $b = \sup A$ and c is an upper bound for A we have $b \leq c$. But since $\epsilon > 0$ we have $b > b - \epsilon = c$ giving the desired contradiction. This proves that there exists $x \in A$ with $b - \epsilon < x$. Choose such an element $x \in A$. Since $b = \sup A$ we know that b is an upper bound for A and hence $b \geq x$. Thus we have $b - \epsilon < x \leq b$, as required.

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

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

Proof: We prove Part (1). Let A be a nonempty subset of $\mathbf{Z}$ which is bounded in $\mathbf{R}$. By Completeness, A has a supremum in $\mathbf{R}$. 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 $b - a \in \mathbf{Z}$ with $0 < b - a < 1$ which contradicts the Basic Order Properties of $\mathbf{Z}$ (since $b - a < 1 \to b - a \leq 0$). Thus $n \in A$ so A has a maximum element.

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

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

Proof: We prove Part (1). First we prove uniqueness. Let $x \in \mathbf{R}$ and suppose that $n, m \in \mathbf{Z}$ 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 $m \leq n$. Similarly, we have $n \leq m$. Since $n \leq m$ and $m \leq n$, we have $n = m$. This proves uniqueness.

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

Next consider the case that $x < 0$. If $x \in \mathbf{Z}$ we can take $n = x$. Suppose that $x \notin \mathbf{Z}$. We have $-x > 0$ so, by the previous paragraph, we can choose $m \in \mathbf{Z}$ with $-x - 1 < m \leq -x$. Since $m \in \mathbf{Z}$ but $x \notin \mathbf{Z}$ 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).

1.31 Definition: Given $x \in \mathbf{R}$ we define the **floor** of x to be the unique $n \in \mathbf{Z}$ with $x - 1 < n \leq x$ and we denote the floor of x by $\lfloor x \rfloor$. The function $f : \mathbf{R} \to \mathbf{Z}$ given by $f(x) = \lfloor x \rfloor$ is called the **floor function**.

1.32 Theorem: *(Archimedean Properties of* $\mathbf{Z}$ *in* $\mathbf{R}$*)*

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

Proof: Let $x \in \mathbf{R}$. 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 $m = \lfloor x \rfloor - 1 \leq x - 1 < x$.

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

Proof: Let $a, b \in \mathbf{R}$ with $a < b$. By the Archimidean Property, we can choose $n \in \mathbf{Z}$ 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 $a < q < b$.

Chapter 2: Injective and Surjective Functions and Cardinality

2.1 Definition: Let X and Y be sets and let $f : X \to Y$. Recall that the **domain** of f and the **range** of f are the sets

$$\text{Domain}(f) = X \ , \ \text{Range}(f) = f(X) = \big\{ f(x) \big| x \in X \big\} \, .$$

For $A \subseteq X$, the **image** of A under f is the set

$$f(A) = \big\{ f(x) \big| x \in A \big\}.$$

For $B \subseteq Y$, the **inverse image** of B under f is the set

$$f^{-1}(B) = \big\{ x \in X \big| f(x) \in B \big\}.$$

2.2 Definition: Let X, Y and Z be sets, let $f : X \to Y$ and let $g : Y \to Z$. We define the **composite** function $g \circ f : X \to Z$ by $(g \circ f)(x) = g\big(f(x)\big)$ for all $x \in X$.

2.3 Definition: We say that f is **injective** (or **one-to-one**, written as $1\!:\!1$) when for every $y \in Y$ there exists at most one $x \in X$ such that $f(x) = y$. Equivalently, f is injective when for all $x_1, x_2 \in X$, if $f(x_1) = f(x_2)$ then $x_1 = x_2$. We say that f is **surjective** (or **onto**) when for every $y \in Y$ there exists at least one $x \in X$ such that $f(x) = y$. Equivalently, f is surjective when $\text{Range}(f) = Y$. We say that f is **bijective** (or **invertible**) when f is both injective and surjective, that is when for every $y \in Y$ there exists exactly one $x \in X$ such that $f(x) = y$. When f is bijective, we define the **inverse** of f to be the function $f^{-1} : Y \to X$ such that for all $y \in Y$, $f^{-1}(y)$ is equal to the unique element $x \in X$ such that $f(x) = y$. Note that when f is bijective so is f^{-1}, and in this case we have $(f^{-1})^{-1} = f$.

2.4 Theorem: *Let $f : X \to Y$ and let $g : Y \to Z$. Then*
(1) if f and g are both injective then so is $g \circ f$,
(2) if f and g are both surjective then so is $g \circ f$, and
(3) if f and g are both invertible then so is $g \circ f$, and in this case $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Proof: To prove Part (1), suppose that f and g are both injective. Let $x_1, x_2 \in X$. If $g(f(x_1)) = g(f(x_2))$ then since g is injective we have $f(x_1) = f(x_2)$, and then since f is injective we have $x_1 = x_2$. Thus $g \circ f$ is injective.

To prove Part (2), suppose that f and g are surjective. Given $z \in Z$, since g is surjective we can choose $y \in Y$ so that $g(y) = z$, then since f is surjective we can choose $x \in X$ so that $f(x) = y$, and then we have $g(f(x)) = g(y) = z$. Thus $g \circ f$ is surjective.

Finally, note that Part (3) follows from Parts (1) and (2).

2.5 Definition: For a set X, we define the **identity function** on X to be the function $I_X : X \to X$ given by $I_X(x) = x$ for all $x \in X$. Note that for $f : X \to Y$ we have $f \circ I_X = f$ and $I_Y \circ f = f$.

2.6 Definition: Let X and Y be sets and let $f : X \to Y$. A **left inverse** of f is a function $g : Y \to X$ such that $g \circ f = I_X$. Equivalently, a function $g : Y \to X$ is a left inverse of f when $g\big(f(x)\big) = x$ for all $x \in X$. A **right inverse** of f is a function $h : Y \to X$ such that $f \circ h = I_Y$. Equivalently, a function $h : Y \to X$ is a right inverse of f when $f\big(h(y)\big) = y$ for all $y \in Y$.

2.7 Theorem: *Let X and Y be nonempty sets and let $f : X \to Y$. Then*
(1) f is injective if and only if f has a left inverse,
(2) f is surjective if and only if f has a right inverse, and
(3) f is bijective if and only if f has a left inverse g and a right inverse h, and in this case we have $g = h = f^{-1}$.

Proof: To prove Part (1), suppose first that f is injective. Since $X \neq \emptyset$ we can choose $a \in X$ and then define $g : Y \to X$ as follows: if $y \in \text{Range}(f)$ then (using the fact that f is 1:1) we define $g(y)$ to be the unique element $x_y \in X$ with $f(x_y) = y$, and if $y \notin \text{Range}(f)$ then we define $g(y) = a$. Then for every $x \in X$ we have $y = f(x) \in \text{Range}(f)$, so $g(y) = x_y = x$, that is $g\big(f(x)\big) = x$. Conversely, if f has a left inverse, say g, then f is 1:1 since for all $x_1, x_2 \in X$, if $f(x_1) = f(x_2)$ then $x_1 = g\big(f(x_1)\big) = g\big(f(x_2)\big) = x_2$.

To prove Part (2), suppose first that f is onto. For each $y \in Y$, choose $x_y \in X$ with $f(x_y) = y$, then define $g : X \to Y$ by $g(y) = x_y$ (we need the Axiom of Choice for this). Then g is a right inverse of f since for every $y \in Y$ we have $f\big(g(y)\big) = f(x_y) = y$. Conversely, if f has a right inverse, say g, then f is onto since given any $y \in Y$ we can choose $x = g(y)$ and then we have $f(x) = f\big(g(y)\big) = y$.

To prove Part (3), suppose first that f is bijective. The inverse function $f^{-1} : Y \to X$ is a left inverse for f because given $x \in X$ we can let $y = f(x)$ and then $f^{-1}(y) = x$ so that $f^{-1}\big(f(x)\big) = f^{-1}(y) = x$. Similarly, f^{-1} is a right inverse for f because given $y \in Y$ we can let x be the unique element in X with $y = f(x)$ and then we have $x = f^{-1}(y)$ so that $f\big(f^{-1}(y)\big) = f(x) = y$. Conversely, suppose that g is a left inverse for f and h is a right inverse for f. Since f has a left inverse, it is injective by Part (1). Since f has a right inverse, it is surjective by Part (2). Since f is injective and surjective, it is bijective. As shown above, the inverse function f^{-1} is both a left inverse and a right inverse. Finally, note that $g = f^{-1} = h$ because for all $y \in Y$ we have

$$g(y) = g\Big(f\big(f^{-1}(y)\big)\Big) = f^{-1}(y) = f^{-1}\Big(f\big(h(y)\big)\Big) = h(y)\,.$$

2.8 Corollary: *Let X and Y be sets. Then there exists an injective map $f : X \to Y$ if and only if there exists a surjective map $g : Y \to X$.*

Proof: Suppose $f : X \to Y$ is an injective map. Then f has a left inverse. Let g be a left inverse of f. Since $g \circ f = I_X$, we see that f is a right inverse of g. Since g has a right inverse, g is surjective. Thus there is a surjective map $g : Y \to X$. Similarly, if $g : Y \to X$ is surjective, then it has a right inverse $f : X \to Y$ which is injective.

2.9 Definition: Let A and B be sets. We say that A and B have the **same cardinality**, and we write $|A| = |B|$, when there exists a bijective map $f : A \to B$ (or equivalently when there exists a bijective map $g : Y \to X$). We say that the cardinality of A is **less than or equal to** the cardinality of B, and we write $|A| \leq |B|$, when there exists an injective map $f : A \to B$ (or equivalently when there exists a surjective map $g : Y \to X$). We say that the cardinality of A is **less than** the cardinality of B, and we write $|A| < |B|$, when $|A| \leq |B|$ and $|A| \neq |B|$, (that is when there exists an injective map $f : A \to B$ but there does not exist a bijective map $g : A \to B$). We also write $|A| \geq |B|$ when $|B| \leq |A|$ and $|A| > |B|$ when $|B| < |A|$.

2.10 Example: The map $f : \mathbf{N} \to 2\mathbf{N}$ given by $f(k) = 2k$ is bijective, so $|2\mathbf{N}| = |N|$. The map $g : \mathbf{N} \to \mathbf{Z}$ given by $g(2k) = k$ and $g(2k+1) = -k-1$ for $k \in \mathbf{N}$ is bijective, so we have $|\mathbf{Z}| = |\mathbf{N}|$. The map $h : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$ given by $h(k,l) = 2^k(2l+1) - 1$ is bijective, so we have $|\mathbf{N} \times \mathbf{N}| = |\mathbf{N}|$.

2.11 Theorem: *For all sets A, B and C,*

(1) $|A| = |A|$,
(2) if $|A| = |B|$ *then* $|B| = |A|$,
(3) if $|A| = |B|$ *and* $|B| = |C|$ *then* $|A| = |C|$,
(4) $|A| \leq |B|$ *if and only if* ($|A| = |B|$ *or* $|A| < |B|$), *and*
(5) if $|A| \leq |B|$ *and* $|B| \leq |C|$ *then* $|A| \leq |C|$.

Proof: Part (1) holds because the identity function $I_A : A \to A$ is bijective. Part (2) holds because if $f : A \to B$ is bijective then so is $f^{-1} : B \to A$. Part (3) holds because if $f : A \to B$ and $g : B \to C$ are bijective then so is the composite $g \circ f : A \to C$. The rest of the proof is left as an exercise.

2.12 Definition: Let A be a set. For each $n \in \mathbf{N}$, let $S_n = \{0, 1, 2, \cdots, n-1\}$. For $n \in \mathbf{N}$, we say that the cardinality of A is equal to n, or that A **has** n **elements**, and we write $|A| = n$, when $|A| = |S_n|$. We say that A is **finite** when $|A| = n$ for some $n \in \mathbf{N}$. We say that A is **infinite** when A is not finite. We say that A is **countable** when $|A| = |\mathbf{N}|$.

2.13 Note: When a set A is finite with $|A| = n$, and when $f : A \to S_n$ is a bijection, if we let $a_k = f^{-1}(k)$ for each $k \in S_n$ then we have $A = \{a_0, a_1, \cdots, a_{k-1}\}$ with the elements a_k distinct. Conversely, if $A = \{a_0, a_1, \cdots, a_{k-1}\}$ with the elements a_k all distinct, then we define a bijection $f : A \to S_n$ by $f(a_k) = k$. Thus we see that A is finite with $|A| = n$ if and only if A is of the form $A = \{a_0, a_1, \cdots, a_{n-1}\}$ with the elements a_k all distinct. Similarly, a set A is countable if and only if A is of the form $A = \{a_0, a_1, a_2, \cdots\}$ with the elements a_k all distinct.

2.14 Note: For $n \in \mathbf{N}$, if A is a finite set with $|A| = n + 1$ and $a \in A$ then $|A \setminus \{a\}| = n$. Indeed, if $A = \{a_0, a_1, \cdots, a_n\}$ with the elements a_i distinct, and if $a = a_k$ so that we have $A \setminus \{a\} = \{a_0, a_1, \cdots, a_{k-1}, a_{k+1}, \cdots, a_n\}$, then we can define a bijection $f : S_n \to A \setminus \{a\}$ by $f(i) = a_i$ for $0 \leq i < k$ and $f(i) = a_{i+1}$ for $k \leq i < n$.

2.15 Theorem: *Let A be a set. Then the following are equivalent.*

(1) A is infinite.
(2) A contains a countable subset.
(3) $|\mathbf{N}| \leq |A|$
(4) There exists a map $f : A \to A$ which is injective but not surjective.

Proof: To prove that (1) implies (2), suppose that A is infinite. Since $A \neq \emptyset$ we can choose an element $a_0 \in A$. Since $A \neq \{a_0\}$ we can choose an element $a_1 \in A \setminus \{a_0\}$. Since $A \neq \{a_0, a_1\}$ we can choose $a_3 \in A \setminus \{a_0, a_1\}$. Continue this procedure: having chosen distinct elements $a_0, a_1, \cdots, a_{n-1} \in A$, since $A \neq \{a_0, a_1, \cdots, a_{n-1}\}$ we can choose $a_n \in A \setminus \{a_0, a_1, \cdots, a_{n-1}\}$. In this way, we obtain a countable set $\{a_0, a_1, a_2, \cdots\} \subseteq A$.

Next we show that (2) is equivalent to (3). Suppose that A contains a countable subset, say $\{a_0, a_1, a_2, \cdots\} \subseteq A$ with the element a_i distinct. Since the a_i are distinct, the map $f : \mathbf{N} \to A$ given by $f(k) = a_k$ is injective, and so we have $|\mathbf{N}| \leq |A|$. Conversely, suppose that $|\mathbf{N}| \leq |A|$, and chose an injective map $f : \mathbf{N} \to A$. Considered as a map from $\mathbf{N}$ to $f(\mathbf{N})$, f is bijective, so we have $|\mathbf{N}| = |f(\mathbf{N})|$ hence $f(\mathbf{N})$ is a countable subset of A.

Next, let us show that (2) implies (4). Suppose that A has a countable subset, say $\{a_0, a_1, a_2, \cdots\} \subseteq A$ with the element a_i distinct. Define $f : A \to A$ by $f(a_k) = a_{k+1}$ for all $k \in \mathbf{N}$ and by $f(b) = b$ for all $b \in A \setminus \{a_0, a_1, a_2, \cdots\}$. Then f is injective but not surjective (the element a_0 is not in the range of f).

Finally, to prove that (4) implies (1) we shall prove that if A is finite then every injective map $f : A \to A$ is surjective. We prove this by induction on the cardinality of A. The only set A with $|A| = 0$ is the set $A = \emptyset$, and then the only function $f : A \to A$ is the empty function, which is surjective. Since that base case may appear too trivial, let us consider the next case. Let $n = 1$ and let A be a set with $|A| = 1$, say $A = \{a\}$. The only function $f : A \to A$ is the function given by $f(a) = a$, which is surjective. Let $n \geq 1$ and suppose, inductively, that for every set A with $|A| = n$, every injective map $f : A \to A$ is surjective. Let B be a set with $|B| = n + 1$ and let $g : B \to B$ be injective. Suppose, for a contradiction, that g is not surjective. Choose an element $b \in B$ which is not in the range of g so that we have $g : B \to B \setminus \{b\}$. Let $A = B \setminus \{b\}$ and let $f : A \to A$ be given by $f(x) = g(x)$ for all $x \in A$. Since $g : B \to A$ is injective and $f(x) = g(x)$ for all $x \in A$, f is also injective. Again since g is injective, there is no element $x \in B \setminus \{b\}$ with $g(x) = g(b)$, so there is no element $x \in A$ with $f(x) = g(b)$, and so f is not surjective. Since $|A| = n$ (by the above note), this contradicts the induction hypothesis. Thus f must be surjective. By the Principle of Induction, for every $n \in \mathbf{N}$ and for every set A with $|A| = n$, every injective function $f : A \to A$ is surjective.

2.16 Corollary: *Let A and B be sets.*

(1) If A is countable then A is infinite.
(2) When $|A| \leq |B|$, if B is finite then so is A (equivalently if A is infinite then so is B).
(3) If $|A| = n$ and $|B| = m$ then $|A| = |B|$ if and only if $n = m$.
(4) If $|A| = n$ and $|B| = m$ then $|A| \leq |B|$ if and only if $n \leq m$.
(5) When one of the two sets A and B is finite, if $|A| \leq |B|$ and $|B| \leq |A|$ then $|A| = |B|$.

Proof: Part (1) is immediate: if A is countable then $|\mathbf{N}| = |A|$, hence $|\mathbf{N}| \leq |A|$, and so A is infinite, by Theorem 2.15..

To prove Part (2), suppose that $|A| \leq |B$ and that $|A|$ is infinite. Since A is infinite, we have $|\mathbf{N}| \leq |A|$ (by Theorem 2.15). Since $|\mathbf{N}| \leq |A|$ and $|A| \leq |B|$ we have $|\mathbf{N}| \leq |B|$ (by Theorem 2.11). Since $|\mathbf{N}| \leq |B|$, B is infinite (by Theorem 2.15 again).

To Prove Part (3), suppose that $|A| = n$ and $|B| = m$. If $n = m$ then we have $S_n = S_m$ and so $|A| = |S_n| = |S_m| = |B|$. Conversely, suppose that $|A| = |B|$. Suppose, for a contradiction, that $n \neq m$, say $n > m$, and note that $S_m \subsetneqq S_n$. Since $|A| = |B|$ we have $|S_n| = |A| = |B| = |S_m|$ so we can choose a bijection $f : S_n \to S_m$. Since $S_m \subsetneqq S_n$, we can consider f as a function $f : S_n \to S_n$ which is injective but not surjective. This contradicts Theorem 2.16, and so we must have $n = m$. This proves Part (3).

To prove Part (4), we again suppose that $|A| = n$ and $|B| = m$. If $n \leq m$ then $S_n \subseteq S_m$ so the inclusion map $I : S_n \to S_m$ is injective and we have $|A| = |S_n| \leq |S_m| = |B|$. Conversely, suppose that $|A| \leq |B|$ and suppose, for a contradiction, that $n > m$. Since $|A| \leq |B|$ we have $|S_n| = |A| \leq |B| = |S_m|$ so we can choose an injective map $f : S_n \to S_m$. Since $n > m$ we have $S_m \subsetneqq S_n$ so we can consider f as a map $f : S_n \to S_n$, and this map is injective but not surjective. This contradicts Theorem 2.16, and so $n \leq m$.

Finally, to prove Part (5) we suppose that one of the two sets A and B is finite, and that $|A| \leq |B|$ and $|B| \leq |A|$. If A is finite then, since $|B| \leq |A|$, Part (2) implies that B is finite. If B is finite then, since $|A| \leq |B|$, Part (2) implies that A is finite. Thus, in either case, we see that A and B are both finite. Since A and B are both finite with $|A| \leq |B|$ and $|B| \leq |A|$, we must have $|A| = |B|$ by Parts (3) and (4).

2.17 Theorem: *Let A be a set. Then $|A| \leq |\mathbf{N}|$ if and only if A is finite or countable.*

Proof: First we claim that every subset of $\mathbf{N}$ is either finite or countable. Let $A \subseteq \mathbf{N}$ and suppose that A is not finite. Since $A \neq \emptyset$, we can set $a_0 = \min A$ (using the Well-Ordering Property of $\mathbf{N}$). Note that $\{0, 1, \cdots, a_0\} \cap A = \{a_0\}$. Since $A \neq \{a_0\}$ (so the set $A \setminus \{a_0\}$ is nonempty) we can set $a_1 = \min A \setminus \{a_0\}$. Then we have $a_0 < a_1$ and $\{0, 1, 2, \cdots, a_1\} \cap A = \{a_0, a_1\}$. Since $A \neq \{a_0, a_1\}$ we can set $a_2 = \min A \setminus \{a_0, a_1\}$. Then we have $a_0 < a_1 < a_2$ and $\{0, 1, 2, \cdots, a_3\} \cap A = \{a_0, a_1, a_2\}$. We continue the procedure: having chosen $a_0, a_1, \cdots, a_{n-1} \in A$ with $a_0 < a_1 < \cdots < a_{n-1}$ such that $A \cap \{0, 1, \cdots, a_{n-1}\} = \{a_0, a_1, \cdots, a_{n-1}\}$, since $A \neq \{a_0, a_1, \cdots, a_{n-1}\}$ we can set $a_n = \min A \setminus \{a_0, a_1, \cdots, a_{n-1}\}$, and then we have $a_0 < a_1 < \cdots < a_{n-1} < a_n$ and $A\{0, 1, 2, \cdots, a_n\} \cap A = \{a_0, a_1, \cdots, a_n\}$. In this way, we obtain a countable set $\{a_0, a_1, a_2, \cdots\} \subseteq A$ with $a_0 < a_1 < a_2 < \cdots$ with the property that for all $m \in \mathbf{N}$, $\{0, 1, 2, \cdots, a_m\} \cap A = \{a_0, a_1, \cdots, a_m\}$. Since $0 \leq a_0 < a_1 < a_2 < \cdots$, it follows (by induction) that $a_k \geq k$ for all $k \in \mathbf{N}$. It follows in turn that $A \subseteq \{a_0, a_1, a_2 \cdots\}$ because given $m \in A$, since $m \leq a_m$ we have

$$m \in \{0, 1, 2, \cdots, m\} \cap A \subseteq \{0, 1, 2, \cdots, a_m\} \cap A = \{a_0, a_1, \cdots, a_m\}.$$

Thus $A = \{a_0, a_1, a_2, \cdots\}$ and the elements a_i are distinct, so A is countable. This proves our claim that every subset of $\mathbf{N}$ is either finite or countable.

Now suppose that $|A| \leq |N|$ and choose an injective map $f : A \to \mathbf{N}$. Since f is injective, when we consider it as a map $f : A \to f(A)$, it is bijective, and so $|A| = |f(A)|$. Since $f(A) \subseteq \mathbf{N}$, the previous paragraph shows that $f(A)$ is either finite or countable. If $f(A)$ is finite with $|f(A)| = n$ then $|A| = |f(A)| = |S_n|$, and if $f(A)$ is countable then we have $|A| = |f(A)| = |\mathbf{N}|$. Thus A is finite or countable.

2.18 Theorem: *Let A be a set. Then*

(1) $|A| < |\mathbf{N}|$ if and only if A is finite,
(2) $|\mathbf{N}| < |A|$ if and only if A is neither finite nor countable, and
(3) if $|A| \leq |\mathbf{N}|$ and $|\mathbf{N}| \leq |A|$ the $|A| = |\mathbf{N}|$.

Proof: Part (1) follows from Theorem 2.15 because

$$\begin{aligned} |A| < |\mathbf{N}| &\iff (|A| \leq |\mathbf{N}| \text{ and } |A| \neq |\mathbf{N}|) \\ &\iff (A \text{ is finite or countable and } A \text{ is not countable}) \\ &\iff A \text{ is finite} \end{aligned}$$

and Part (2) follows from Theorem 2.17 because

$$\begin{aligned} |\mathbf{N}| < |A| &\iff (|\mathbf{N}| \leq |A| \text{ and } |\mathbf{N}| \neq |A|) \\ &\iff (A \text{ is not finite and } A \text{ is not countable.}) \end{aligned}$$

To prove Part (3), suppose that $|A| \leq |\mathbf{N}|$ and $|\mathbf{N}| \leq |A|$. Since $|A| \leq |\mathbf{N}|$, we know that A is finite or countable by Theorem 2.17. Since $|\mathbf{N}| \leq |A|$, we know that that A is infinite by Theorem 2.15. Since A is finite or countable and A is not finite, it follows that A is countable. Thus $|A| = |\mathbf{N}|$.

2.19 Definition: Let A be a set. When A is countable we write $|A| = \aleph_0$. When A is finite we write $|A| < \aleph_0$. When A is infinite we write $|A| \geq \aleph_0$. When A is either finite or countable we write $|A| \leq \aleph_0$ and we say that A is **at most countable**. when A is neither finite nor countable we write $|A| > \aleph_0$ and we say that A is **uncountable**.

2.20 Theorem:

(1) If A and B are countable sets, then so is $A \times B$.
(2) If A and B are countable sets, then so is $A \cup B$.
(3) If $A_0, A_1, A_2, \cdots$ are countable sets, then so is $\bigcup_{k=0}^{\infty} A_k$.
(4) $\mathbf{Q}$ is countable.

Proof: To prove Parts (1) and (2), let $A = \{a_0, a_1, a_2, \cdots\}$ with the a_i distinct and let $B = \{b_0, b_1, b_2, \cdots\}$ with the b_i distinct. Since every positive integer can be written uniquely in the form $2^k(2l+1)$ with $k, l \in \mathbf{N}$, the map $f : A \times B \to \mathbf{N}$ given by $f(a_k, b_l) = 2^k(2l+1)-1$ is bijective, and so $|A \times B| = |\mathbf{N}|$. This proves Part (1). Since the map $g : \mathbf{N} \to A \cup B$ given by $g(k) = a_k$ is injective, we have $|\mathbf{N}| \leq |A \cup B|$. Since the map $h : \mathbf{N} \to A \cup B$ given by $h(2k) = a_k$ and $h(2k+1) = b_k$ is surjective, we have $|A \cup B| \leq |N|$. Since $|\mathbf{N}| \leq |A \cup B|$ and $|A \cup B| \leq |\mathbf{N}|$, we have $|A \cup B| = |\mathbf{N}|$ by Part (3) of Theorem 2.18. This proves (2).

To prove Part (3), for each $k \in \mathbf{N}$, let $A_k = \{a_{k0}, a_{k1}, a_{k2}, \cdots\}$ with the a_{ki} distinct. Since the map $f : \mathbf{N} \to \bigcup_{k=0}^{\infty} A_k$ given by $f(k) = a_{0,k}$ is injective, $|\mathbf{N}| \leq \left|\bigcup_{k=0}^{\infty} A_k\right|$. Since $\mathbf{N} \times \mathbf{N}$ is countable by Part (1), and since the map $g : \mathbf{N} \times \mathbf{N} \to \bigcup_{k=0}^{\infty} A_k$ given by $g(k, l) = a_{k,l}$ is surjective, we have $\left|\bigcup_{k=0}^{\infty} A_k\right| \leq |\mathbf{N} \times \mathbf{N}| = |\mathbf{N}|$. By Part (3) of Theorem 2.18, we have $\left|\bigcup_{k=0}^{\infty} A_k\right| = |\mathbf{N}|$, as required.

Finally, we prove Part (4). Since the map $f : \mathbf{N} \to \mathbf{Q}$ given by $f(k) = k$ is injective, we have $|\mathbf{N}| \leq |\mathbf{Q}|$. Since the map $g : \mathbf{Q} \to \mathbf{Z} \times \mathbf{Z}$, given by $g\left(\frac{a}{b}\right) = (a, b)$ for all $a, b \in \mathbf{Z}$ with $b > 0$ and $\gcd(a, b) = 1$, is injective, and since $\mathbf{Z} \times \mathbf{Z}$ is countable, we have $|\mathbf{Q}| \leq |\mathbf{Z} \times \mathbf{Z}| = |\mathbf{N}|$. Since $|\mathbf{N}| \leq |\mathbf{Q}|$ and $|\mathbf{Q}| \leq |\mathbf{N}|$, we have $|\mathbf{Q}| = |\mathbf{N}|$, as required.

2.21 Definition: For a set A, let $\mathcal{P}(A)$ denote the **power set** of A, that is the set of all subsets of A, and let 2^A denote the set of all functions from A to $S_2 = \{0, 1\}$.

2.22 Theorem:

(1) For every set A, $\left|\mathcal{P}(A)\right| = \left|2^A\right|$.
(2) For every set A, $|A| < \left|\mathcal{P}(A)\right|$.
(3) $\mathbf{R}$ is uncountable.

Proof: Let A be any set.Define a map $g : \mathcal{P}(A) \to 2^A$ as follows. Given $S \in \mathcal{P}(A)$, that is given $S \subseteq A$, we define $g(S) \in 2^A$ to be the map $g(S) : A \to \{0, 1\}$ given by

$$g(S)(a) = \begin{cases} 1 \text{ if } a \in S, \\ 0 \text{ if } a \notin S. \end{cases}$$

Define a map $h : 2^A \to \mathcal{P}(A)$ as follows. Given $f \in 2^A$, that is given a map $f : A \to \{0, 1\}$, we define $h(f) \in \mathcal{P}(A)$ to be the subset

$$h(f) = \left\{a \in A \middle| f(a) = 1\right\} \subseteq A.$$

The maps g and h are the inverses of each other because for every $S \subseteq A$ and every $f : A \to \{0, 1\}$ we have

$$f = g(S) \iff \forall a \in A \;\; f(a) = g(S)(a) \iff \forall a \in A \;\; f(a) = \begin{cases} 1 \text{ if } a \in S, \\ 0 \text{ if } a \notin S, \end{cases}$$

$$\iff \forall a \in A \;\big(f(a) = 1 \iff a \in S\big) \iff \left\{a \in A \middle| f(a) = 1\right\} = S \iff h(f) = S.$$

This completes the proof of Part (1).

Let us prove Part (2). Again we let A be any set. Since the the map $f : A \to \mathcal{P}(A)$ given by $f(a) = \{a\}$ is injective, we have $|A| \leq \big|\mathcal{P}(A)\big|$. We need to show that $|A| \neq \big|\mathcal{P}(A)\big|$. Let $g : A \to \mathcal{P}(A)$ be any map. Let $S = \big\{a \in A \big| a \notin g(a)\big\}$. Note that S cannot be in the range of g because if we could choose $a \in A$ so that $g(a) = S$ then, by the definition of S, we would have $a \in S \iff a \notin g(a) \iff a \notin S$ which is not possible. Since S is not in the range of g, the map g is not surjective. Since g was an arbitrary map from A to $\mathcal{P}(A)$, it follows that there is no surjective map from A to $\mathcal{P}(A)$. Thus there is no bijective map from A to $\mathcal{P}(A)$ and so we have $|A| \neq \big|\mathcal{P}(A)\big|$, as desired.

Finally, we shall prove that $\mathbf{R}$ is uncountable using the fact (which we did not prove) that every real number has a unique decimal expansion which does not end with an infinite string of 9's. We define a map $g : 2^{\mathbf{N}} \to \mathbf{R}$ as follows. Given $f \in 2^{\mathbf{N}}$, that is given a map $f : \mathbf{N} \to \{0, 1\}$, we define $g(f)$ to be the real number $g(f) \in [0, 1)$ with the decimal expansion $g(f) = .f(0)f(1)f(2)f(3)\cdots$ (for those who have seen infinite series, this is the number $g(f) = \sum_{k=0}^{\infty} f(k)10^{-k-1}$). By the uniqueness of decimal expansions, the map g is injective, so we have $\big|2^{\mathbf{N}}\big| \leq |\mathbf{R}|$. Thus $|\mathbf{N}| < \big|\mathcal{P}(\mathbf{N})\big| = \big|2^{\mathbf{N}}\big| \leq |\mathbf{R}|$, and so $\mathbf{R}$ is uncountable, by Part (2) of Theorem 2.18.

2.23 Theorem: *(Cantor - Schroeder - Bernstein) Let A and B be sets. Suppose that $|A| \leq |B|$ and $|B| \leq |A|$. Then $|A| = |B|$*

Proof: We sketch a proof. Choose injective functions $f : A \to B$ and $g : B \to A$. Since the functions $f : A \to f(A)$, $g : B \to g(B)$ and $f : g(B) \to f\big(g(B)\big)$ are bijective we have $|A| = |f(A)|$ and $|B| = |g(B)| = \big|f(g(B))\big|$. Also note that $f\big(g(B)\big) \subseteq f(A) \subseteq B$. Let $X = f\big(g(B)\big)$, $Y = f(A)$ and $Z = B$. Then we have $X \subseteq Y \subseteq Z$ and we have $|X| = |Z|$ and we need to show that $|Y| = |Z|$. The composite $h = f \circ g : Z \to X$ is a bijection. Define sets Z_n and Y_n for $n \in \mathbf{N}$ recursively by

$$Z_0 = Z,\ Z_n = h(Z_{n-1}) \ \text{ and } \ Y_0 = Y\ ,\ Y_n = h(Y_{n-1}).$$

Since $Y_0 = Y$, $Z_0 = Z$, $Z_1 = h(Z_0) = h(Z) = X$ and $X \subseteq Y \subseteq Z$, we have

$$Z_1 \subseteq Y_0 \subseteq Z_0.$$

Also note that for $1 \leq n \in \mathbf{N}$,

$$Z_n \subseteq Y_{n-1} \subseteq Z_{n-1} \to h(Z_n) \subseteq h(Y_{n-1}) \subseteq h(Z_{n-1}) \to Z_{n+1} \subseteq Y_n \subseteq Z_n.$$

By the Induction Principle, it follows that $Z_n \subseteq Y_{n-1} \subseteq Z_{n-1}$ for all $n \geq 1$, so we have

$$Z_0 \supseteq Y_0 \supseteq Z_1 \supseteq Y_1 \supseteq Z_2 \supseteq Y_2 \supseteq \cdots$$

Let $U_n = Z_n \setminus Y_n$, $U = \bigcup\limits_{n=1}^{\infty} U_n$ and $V = Z \setminus U$. Define $H : Z \to Y$ by

$$H(x) = \begin{cases} h(x) & \text{if } x \in U, \\ x & \text{if } x \in V. \end{cases}$$

Verify that H is bijective.

2.24 Exercise: Show that $|\mathbf{R}| = \left|2^{\mathbf{N}}\right|$.

Solution: $g : 2^{\mathbf{N}} \to \mathbf{R}$ as follows: for $f \in 2^{\mathbf{N}}$ we let $g(f)$ be the real number $g(f) \in [0,1)$ with decimal expansion $g(f) = 0.f(0)f(1)f(2)\cdots.$. Then g is injective so $\left|s^{\mathbf{N}}\right| \leq |\mathbf{R}|$. Define $h : 2^{\mathbf{N}} \to [0,1)$ as follows: for $f \in 2^{\mathbf{N}}$ let $h(f)$ be the real number $h(f) \in [0,1]$ with binary expansion $h(f) = 0.f(0)f(1)f(2)\cdots$. Then h is surjective so we have $\left|[0,1]\right| \leq \left|2^{\mathbf{N}}\right|$. The map $k : \mathbf{R} \to [0,1]$ given by $k(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1} x$ is injective so we have $|\mathbf{R}| \leq \left|[0,1]\right|$. Since $|\mathbf{R}| \leq \left|[0,1]\right| \leq \left|2^{\mathbf{N}}\right|$ and $\left|2^{\mathbf{N}}\right| \leq |\mathbf{R}|$, we have $|\mathbf{R}| = \left|2^{\mathbf{N}}\right|$ by the Cantor-Schroeder-Bernstein Theorem.

Chapter 3: Sequences

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

$$\langle x_k \rangle_{k \geq p} = \langle x_k | k \geq p \rangle = \langle x_p, x_{p+1}, x_{p+2}, \cdots \rangle.$$

Note that the range of the sequence $\langle x_k \rangle_{k \geq p}$ is the set $\{x_k\}_{k \geq p} = \{x_k | k \geq p\}$.

3.2 Definition: Let F be an ordered field and let $\langle x_k \rangle_{k \geq p}$ be a sequence in F. For $a \in F$ we say that the sequence $\langle x_k \rangle_{k \geq p}$ **converges** to a (or that the **limit** of $\langle x_k \rangle_{k \geq p}$ is equal to a), and we write $x_k \to a$ (as $k \to \infty$), or we write $\lim\limits_{k \to \infty} x_k = a$, when

$$\forall\, 0 < \epsilon \in F\ \exists m \in \mathbf{Z}\ \forall k \in \mathbf{Z}_{\geq p}\ \big(k \geq m \to |x_k - a| \leq \epsilon\big).$$

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

$$\forall r \in F\ \exists m \in \mathbf{Z}\ \forall k \in \mathbf{Z}_{\geq p}\ \big(k \geq m \to x_k \geq r\big).$$

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

$$\forall r \in \mathbf{R}\ \exists m \in \mathbf{Z}\ \forall k \in \mathbf{Z}_{\geq p}\ \big(k \geq m \to x_k \leq r\big).$$

3.3 Example: Let $\langle x_k \rangle_{k \geq 0}$ be the sequence in $\mathbf{R}$ given by $x_k = \frac{(-2)^k}{k!}$ for $k \geq 0$. Show that $\lim\limits_{k \to \infty} x_k = 0$.

Solution: Note that for $k \geq 2$ we have

$$|x_k| = \tfrac{2^k}{k!} = \left(\tfrac{2}{1}\right)\left(\tfrac{2}{2}\right)\left(\tfrac{2}{3}\right)\cdots\left(\tfrac{2}{k-1}\right)\left(\tfrac{2}{k}\right) \leq \tfrac{2}{1} \cdot \tfrac{2}{n} = \tfrac{4}{n}.$$

Given $\epsilon \in \mathbf{R}$ with $\epsilon > 0$, we can choose $m \in \mathbf{Z}_{\geq 2}$ with $m \geq \frac{4}{\epsilon}$ and then for all $k \geq m$ we have $|x_k - 0| = |x_k| \leq \frac{4}{k} \leq \frac{4}{m} \leq \epsilon$. Thus $\lim\limits_{k \to \infty} x_k = 0$, by the definition of the limit.

3.4 Example: Let $\langle a_k \rangle_{k \geq 0}$ be the **Fibonacci sequence** in $\mathbf{R}$, which is defined recursively by $a_0 = 0$, $a_1 = 1$ and by $a_k = a_{k-1} + a_{k-2}$ for $k \geq 2$. Show that $\lim\limits_{k \to \infty} a_k = \infty$.

Solution: We have $a_0 = 0$, $a_1 = 1$, $a_2 = 1$ and $a_3 = 2$. Note that $a_k \geq k - 1$ when $k \in \{0, 1, 2, 3\}$. Let $n \geq 4$ and suppose, inductively, that $a_k \geq k - 1$ for all $k \in \mathbf{Z}$ with $0 \leq k < n$. Then $a_n = a_{n-1} + a_{n-2} \geq (n-2) + (n-3) = n + n - 5 \geq n + 4 - 5 = n - 1$. By the Strong Principle of Induction, we have $a_n \geq n - 1$ for all $n \geq 0$. Given $r \in \mathbf{R}$ we can choose $m \in \mathbf{Z}_{\geq 0}$ with $m \geq r + 1$, and then for all $k \geq m$ we have $a_k \geq k - 1 \geq m - 1 \geq r$. Thus $\lim\limits_{k \to \infty} a_k = \infty$ by the definition of the limit.

3.5 Example: Let $x_k = (-1)^k$ for $k \geq 0$. Show that $\langle x_k \rangle_{k \geq 0}$ diverges.

Solution: Suppose, for a contradiction, that $\langle x_k\rangle_{k\geq 0}$ converges and let $a=\lim\limits_{k\to\infty} x_k$. By taking $\epsilon=\frac{1}{2}$ in the definition of the limit, we can choose $m\in\mathbf{Z}$ so the for all $k\in\mathbf{N}$, if $k\geq m$ then $|x_k-a|\leq\frac{1}{2}$. Choose $k\in\mathbf{N}$ with $2k\geq m$. Since $|x_{2k}-a|\leq\frac{1}{2}$ and $x_{2k}=(-1)^{2k}=1$, we have $|1-a|\leq\frac{1}{2}$ so that $\frac{1}{2}\leq a\leq\frac{3}{2}$. Since $|x_{2k+1}-a|\leq\frac{1}{2}$ and $x_{2k+1}=(-1)^{2k+1}=-1$, we also have $|-1-a|\leq\frac{1}{2}$ which implies that $-\frac{3}{2}\leq a\leq-\frac{1}{2}$. But then we have $a\leq-\frac{1}{2}$ and $a\geq\frac{1}{2}$, which is not possible.

3.6 Theorem: *(Independence of the Limit on the Initial Terms) Let $\langle x_k\rangle_{k\geq p}$ be a sequence in an ordered field F.*

(1) If $q\geq p$ and $y_k=x_k$ for all $k\geq q$, then $\langle x_k\rangle_{k\geq p}$ converges if and only if $\langle y_k\rangle_{k\geq q}$ converges, and in this case $\lim\limits_{k\to\infty} x_k=\lim\limits_{k\to\infty} y_k$.
(2) If $l\geq 0$ and $y_k=x_{k+l}$ for all $k\geq p$, then $\langle x_k\rangle_{k\geq p}$ converges if and only if $\langle y_k\rangle_{k\geq p}$ converges, and in this case $\lim\limits_{k\to\infty} x_k=\lim\limits_{k\to\infty} y_k$.

Proof: We prove Part (1) and leave the proof of Part (2) as an exercise. Let $q\geq p$ and let $y_k=x_k$ for $k\geq q$. Suppose $\langle x_k\rangle_{k\geq p}$ converges and let $a=\lim\limits_{k\to\infty} x_k$. Let $\epsilon>0$. Choose $m\in\mathbf{Z}$ so that for all $k\in\mathbf{Z}_{\geq p}$, if $k\geq m$ then $|x_k-a|\leq\epsilon$. Let $k\in\mathbf{Z}_{\geq q}$ with $k\geq m$. Since $q\geq p$ we also have $k\in\mathbf{Z}_{\geq p}$ and so $|y_k-a|=|x_k-a|\leq\epsilon$. Thus $\langle y_k\rangle_{k\geq q}$ converges with $\lim\limits_{k\to\infty} y_k=a$. Conversely, suppose that $\langle y_k\rangle_{k\geq q}$ converges and let $a=\lim\limits_{k\to\infty} y_k$. Let $\epsilon>0$. Choose $m_1\in\mathbf{Z}$ so that for all $k\in\mathbf{Z}_{\geq q}$, if $k\geq m_1$ then $|y_k-a|\leq\epsilon$. Choose $m=\max\{m_1,q\}$. Let $k\in\mathbf{Z}_{\geq p}$ with $k\geq m$. Since $k\geq m$, we have $k\geq q$ and $k\geq m_1$ and so $|x_k-a|=|y_k-a|\leq\epsilon$. Thus $\langle x_k\rangle_{k\geq p}$ converges with $\lim\limits_{k\to\infty} x_k=a$.

3.7 Remark: Because of the above theorem, we often denote the sequence $\langle x_k\rangle_{k\geq p}$ simply as $\langle x_k\rangle$ (omitting the initial index p from our notation).

3.8 Theorem: *(Uniqueness of the Limit) Let $\langle x_k\rangle$ be a sequence in an ordered field F. If $\langle x_k\rangle$ has a limit (finite or infinite) then the limit is unique.*

Proof: Suppose, for a contradiction, that $x_k\to\infty$ and $x_k\to-\infty$. Since $x_k\to\infty$ we can choose $m_1\in\mathbf{Z}$ so that $k\geq m_1\to x_k\geq 1$. Since $x_k\to-\infty$ we can choose $m_2\in\mathbf{Z}$ so that $k\geq m_2\to x_k\leq -1$. Choose any $k\in\mathbf{Z}_{\geq p}$ with $k\geq m_1$ and $k\geq m_2$. Then $x_k\geq 1$ and $x_k\leq -1$, which is not possible.

Suppose, for a contradiction, that $x_k\to\infty$ and $x_k\to a\in F$. Since $x_k\to a$ we can choose $m_1\in\mathbf{Z}$ so that $k\geq m_1\to|x_k-a|\leq 1$. Since $x_k\to\infty$ we can choose $m_2\in\mathbf{Z}$ so that $k\geq m_2\to x_k\geq a+2$. Choose any $k\in\mathbf{Z}_{\geq p}$ with $k\geq m_1$ and $k\geq m_2$. Then we have $|x_k-a|\leq 1$ so that $x\leq a+1$ and we have $x_k\geq a+2$, which is not possible. Similarly, it is not possible to have $x_k\to-\infty$ and $x_k\to a\in F$.

Finally suppose, for a contradiction, that $x_k\to a$ and $x_k\to b$ where $a,b\in F$ with $a\neq b$. Since $x_k\to a$ we can choose $m_1\in\mathbf{Z}$ so that $k\geq m_1\to|x_k-a|\leq\frac{|a-b|}{3}$. Since $x_k\to b$ we can choose $m_2\in\mathbf{Z}$ so that $k\geq m_2\to|x_k-b|\leq\frac{|a-b|}{3}$. Choose any $k\in\mathbf{Z}_{\geq p}$ with $k\geq m_1$ and $k\geq m_2$. Then we have $|x_k-a|\leq\frac{b-a}{3}$ and $|x_k-b|\leq\frac{b-a}{3}$ and so, using the Triangle Inequality, we have

$$|a-b|=|a-x_k+x_k-b|\leq|x_k-a|+|x_k-b|\leq\tfrac{|a-b|}{3}+\tfrac{|a-b|}{3}<|a-b|,$$

which is not possible.

3.9 Theorem: *(Basic Limits) In any ordered field F, for $a \in F$ we have*

$$\lim_{k\to\infty} a = a \ , \ \lim_{k\to\infty} k = \infty \ \text{ and } \ \lim_{k\to\infty} \frac{1}{k} = 0.$$

Proof: The proof is left as an exercise.

3.10 Theorem: *(Operations on Limits) Let $\langle x_k \rangle$ and $\langle y_k \rangle$ be sequences in an ordered field F and let $c \in F$. Suppose that $\langle x_k \rangle$ and $\langle y_k \rangle$ both converge with $x_k \to a$ and $y_k \to b$. Then*

(1) $\langle c\, x_k \rangle$ converges with $c\, x_k \to ca$,
(2) $\langle x_k + y_k \rangle$ converges with $(x_k + y_k) \to a + b$,
(3) $\langle x_k - y_k \rangle$ converges with $(x_k - y_k) \to a - b$,
(4) $\langle x_k y_k \rangle$ converges with $x_k y_k \to ab$, and
(5) if $b \neq 0$ then $\langle x_k / y_k \rangle$ converges with $x_k / y_k \to a/b$.

Proof: We prove Parts (4) and (5) leaving the proofs of the other parts as an exercise. First we prove Part (4). Note that for all k we have

$$|x_k y_k - ab| = |x_k y_k - x_k b + x_k b - ab| \leq |x_k y_k - x_k b| + |x_k b - ab| = |x_k|\,|y_k - b| + |b||x_k - a|.$$

Since $x_k \to a$ we can choose $m_1 \in \mathbf{Z}$ so that $k \geq m_1 \to |x_k - a| \leq 1$ and we can choose $m_2 \in \mathbf{Z}$ so that $k \geq m_2 \to |x_k - a| \leq \frac{\epsilon}{2(1+|b|)}$. Since $y_k \to b$ we can choose $m_3 \in \mathbf{Z}$ so that $k \geq m_3 \to |y_k - b| \leq \frac{\epsilon}{2(1+|a|)}$. Let $m = \max\{m_1, m_2, m_3\}$ and let $k \geq m$. Then we have $|x_k - a| \leq 1$, $|x_k - a| \leq \frac{\epsilon}{2(1+|b|)}$ and $|x_k - b| \leq \frac{\epsilon}{2(1+|a|)}$. Since $|x_k - a| \leq 1$, we have $|x_k| = |x_k - a + a| \leq |x_k - a| + |a| \leq 1 + |a|$. By our above calculation (where we found a bound for $|x_k y_k - ab|$) we have

$$\begin{aligned} |x_k y_k - ab| &\leq |x_k||y_k - b| + |b||x_k - a| \leq (1 + |a|)|y_k - b| + (1 + |b|)|x_k - a| \\ &\leq (1 + |a|)\tfrac{\epsilon}{2(1+|a|)} + (1 + |b|)\tfrac{\epsilon}{2(1+|b|)} = \epsilon. \end{aligned}$$

Thus we have $x_k y_k \to ab$, by the definition of the limit.

To prove Part (5), suppose that $b \neq 0$. Since $y_k \to b \neq 0$, we can choose $m_1 \in \mathbf{Z}$ so that that $k \geq m_1 \to |y_k - b| \leq \frac{|b|}{2}$. Then for $k \geq m_1$ we have

$$|b| = |b - y_k + y_k| \leq |b - y_k| + |y_k| \leq \tfrac{|b|}{2} + |y_k|$$

so that

$$|y_k| \geq |b| - \tfrac{|b|}{2} = \tfrac{|b|}{2} > 0.$$

In particular, we remark that when $k \geq m_1$ we have $y_k \neq 0$ so that $\frac{1}{y_k}$ is defined. Note that for all $k \geq m_1$ we have

$$\left| \frac{1}{y_k} - \frac{1}{b} \right| = \frac{|b - y_k|}{|y_k|\,|b|} \leq \frac{|b - y_k|}{\frac{|b|}{2} \cdot |b|} = \frac{2}{|b|^2} \cdot |y_k - b|.$$

Let $\epsilon > 0$. Choose $m_2 \in \mathbf{Z}$ so that $k \geq m_2 \to |y_k - b| \leq \frac{|b|^2 \epsilon}{2}$. Let $m = \max\{m_1, m_2\}$. For $k \geq m$ we have $k \geq m_1$ and $k \geq m_2$ and so $|y_k| \geq \frac{|b|^2}{2}$ and $|y_k - b| \leq \frac{|b|^2 \epsilon}{2}$ and so

$$\left| \frac{1}{y_k} - \frac{1}{b} \right| \leq \tfrac{2}{|b|^2} \cdot |y_k - b| \leq \tfrac{2}{|b|^2} \cdot \tfrac{|b|^2 \epsilon}{2} = \epsilon.$$

This proves that $\lim\limits_{k\to\infty} \frac{1}{y_k} = \frac{1}{b}$. Using Part (4), we have $\lim\limits_{k\to\infty} \frac{x_k}{y_k} = \lim\limits_{k\to\infty} \left(x_k \cdot \frac{1}{y_k} \right) = a \cdot \frac{1}{b} = \frac{a}{b}$.

3.11 Example: Let $x_k = \frac{k^2+1}{2k^2+k+3}$ for $k \geq 0$. Find $\lim\limits_{k\to\infty} x_k$.

Solution: We have $x_k = \frac{k^2+1}{2k^2+k+2} = \frac{1+\left(\frac{1}{k}\right)^2}{2+\frac{1}{k}+3\cdot\left(\frac{1}{k}\right)^2} \longrightarrow \frac{1+0^2}{2+0+3\cdot 0^2} = \frac{1}{2}$ where we used the Basic Limits $1 \to 1$, $2 \to 2$ and $\frac{1}{k} \to 0$ together with Operations on Limits.

3.12 Definition: The above theorem can be extended to include many situations involving infinite limits. To deal with these cases, given an ordered field F, we define the **extended ordered field** $\hat{F}$ to be the set

$$\hat{F} = F \cup \{-\infty, \infty\}.$$

We extend the order relation $<$ on F to an order relation on $\hat{F}$ by defining $-\infty < \infty$ and $-\infty < a$ and $a < \infty$ for all $a \in F$. We partially extend the operations $+$ and $\cdot$ to $\hat{F}$; for $a \in F$ we define

$$\infty + \infty = \infty \ , \ \infty + a = \infty \ , \ (-\infty) + (-\infty) = -\infty \ , \ (-\infty) + a \ ,$$
$$\infty \cdot \infty = \infty \ , (\infty)(-\infty) = -\infty \ , \ (-\infty)(-\infty) = \infty \ ,$$
$$\infty \cdot a = \begin{cases} \infty \text{ if } a > 0 \\ -\infty \text{ if } a < 0 \end{cases} \text{ and } (-\infty)(a) = \begin{cases} -\infty \text{ if } a > 0, \\ \infty \text{ if } a < 0, \end{cases}$$

but other values, including $\infty + (-\infty)$, $\infty \cdot 0$ and $-\infty \cdot 0$ are left undefined in $\hat{F}$. In a similar way, we partially extend the inverse operations $-$ and $\div$ to $\hat{F}$. For example, for $a \in F$ we define

$$\infty - (-\infty) = \infty \, , -\infty - \infty = -\infty \, , \infty - a = \infty \, , -\infty - a = -\infty \, , a - \infty = -\infty \, , a - (-\infty) = \infty \, ,$$
$$\frac{a}{\infty} = 0 \ , \ \frac{\infty}{a} = \begin{cases} \infty \text{ if } a > 0 \\ -\infty \text{ if } a < 0 \end{cases} \text{ and } \frac{-\infty}{a} = \begin{cases} -\infty \text{ if } a > 0 \\ \infty \text{ if } a < 0 \end{cases}$$

with other values, including $\infty - \infty$, $\frac{\infty}{\infty}$ and $\frac{\infty}{0}$, left undefined. The expressions which are left undefined in $\hat{F}$, including

$$\infty - \infty \ , \ \infty \cdot 0 \ , \ \frac{\infty}{\infty} \ , \ \frac{\infty}{0} \ , \ \frac{a}{0}$$

are known as **indeterminate forms**.

3.13 Theorem: *(Extended Operations on Limits) Let $\langle x_k \rangle$ and $\langle y_k \rangle$ be sequences in F. Suppose that $\lim\limits_{k\to\infty} x_k = u$ and $\lim\limits_{k\to\infty} y_k = v$ where $u, v \in \hat{F}$.*

(1) if $u + v$ is defined in $\hat{F}$ then $\lim\limits_{k\to\infty} (x_k + y_k) = u + v$,

(2) if $u - v$ is defined in $\hat{F}$ then $\lim\limits_{k\to\infty} (x_k - y_k) = u - v$,

(3) if $u \cdot v$ is defined in $\hat{F}$ then $\lim\limits_{k\to\infty} (x_k \cdot y_k) = u \cdot v$, and

(4) if u/v is defined in $\hat{F}$ then $\lim\limits_{k\to\infty} (x_k / y_k) = u/v$.

Proof: The proof is left as an exercise.

3.14 Theorem: *(Monotonic Surjective Functions) Let I and J be intervals in an ordered field F. Suppose $f : I \to J$ is increasing and surjective. Let $\langle x_k \rangle$ be a sequence in I. Then*
(1) If $x_k \to a \in I$ then $f(x_k) \to f(a) \in J$,
(2) if $x_k \to u$ where $u \in F \cup \{\infty\}$ is the right endpoint of I, then $f(x_k) \to v$ where $v \in F \cup \{\infty\}$ is the right endpoint of J, and
(3) if $x_k \to u$ where $u \in F \cup \{-\infty\}$ is the left endpoint of I then $f(x_k) \to v$ where $v \in F \cup \{-\infty\}$ is the left endpoint of J.

Analogous results hold when $f : I \to J$ is decreasing and surjective.

Proof: We prove Part (1). Let $a \in I$, suppose $x_k \to a$, and let $b = f(a) \in J$. Note that since f is surjective, it has a right inverse. Let $g : J \to I$ be a right inverse of f. Let $\epsilon > 0$. We consider several cases, depending on whether or not b is an endpoint of J. Suppose first that b is not an endpoint of J. Choose ϵ_0 with $0 < \epsilon_0 \le \epsilon$ so that $[b - \epsilon_0, b + \epsilon_0] \subseteq J$. Note that since f is increasing we have $g(b - \epsilon_0) < a < g(b + \epsilon_0)$ (since $g(b - \epsilon_0) \ge a \to b - \epsilon = f\big(g(b - \epsilon_0)\big) \le f(a) = b$ which is impossible, and $a \ge g(b + \epsilon_0) \to b = f(a) \ge f\big(g(b + \epsilon_0)\big) = b + \epsilon_0$ which is impossible). Since $x_k \to a$ we can choose $m \in \mathbf{Z}$ so that $k \ge m \to g(b - \epsilon_0) \le x_k \le g(b + \epsilon_0)$. Then for $k \ge m$ we have $b - \epsilon_0 = f\big(g(b - \epsilon_0)\big) \le f(x_k) \le f\big(g(b + \epsilon_0)\big) = g + \epsilon_0$. Thus $f(x_k) \to b = f(a)$.

Next consider the case that b is equal to one (but not both) of the endpoints of J, say b is the right endpoint of J, and say the left endpoint of J is smaller than b. In this case, we choose ϵ_0 with $0 < \epsilon_0 \le \epsilon$ so that $[b - \epsilon_0, b] \subseteq J$. Note that since f is increasing we have $g(b - \epsilon_0) < a$. Choose $m \in \mathbf{Z}$ so that $k \ge m \to g(b - \epsilon_0) \le x_k$. Then for $k \ge m$, since f is increasing we have $b - \epsilon_0 \le f(x_k)$. Since b is the right endpoint of J, it follows that $b - \epsilon_0 \le f(x_k) \le b$ for all $k \ge m$, and so $f(x_k) \to b = f(a)$.

Finally, note that if b is equal to both the left and right endpoints of J, then we have $J = \{b\}$ and so $f(x_k) = b$ for all k, and hence $f(x_k) \to b$.

3.15 Corollary: *(Basic Elementary Functions Acting on Limits) Let $\langle x_k \rangle$ be a sequence in $\mathbf{R}$ and let $b \in \mathbf{R}$. Then*

(1) if $x_k \to a > 0$ then ${x_k}^b \to a^b$,

$$\text{if } x_k \to \infty \text{ then } \lim_{k\to\infty} {x_k}^b = \begin{cases} \infty \text{ if } b > 0 \\ 0 \text{ if } b < 0, \end{cases}$$

(2) if $x_k \to a$ and $b > 0$ then $b^{x_k} \to b^a$,

$$\text{if } x_k \to \infty \text{ and } b > 0 \text{ then } \lim_{k\to\infty} b^{x_k} = \begin{cases} \infty \text{ if } b > 1 \\ 0 \text{ if } 0 < b < 1, \end{cases}$$

(3) if $x_k \to a > 0$ and $b > 0$ then $\log_b x_k \to \log_b a$,

$$\text{if } x_k \to \infty \text{ and } b > 0 \text{ then } \lim_{k\to\infty} x_k = \begin{cases} \infty \text{ if } b > 1 \\ -\infty \text{ if } 0 < b < 1 \end{cases}$$

(4) if $x_k \to a$ then $\sin x_k \to \sin a$ and $\cos x_k \to \cos a$
if $x_k \to a$, where $a \ne \frac{\pi}{2} + 2\pi\, t$ with $t \in \mathbf{Z}$, then $\tan x_k \to \tan a$
(5) if $x_k \to a \in [-1, 1]$ then $\sin^{-1} x_k \to \sin^{-1} a$ and $\cos^{-1} x_k \to \cos^{-1} a$
if $x_k \to a$ then $\tan^{-1} x_k \to \tan^{-1} a$
if $x_k \to \infty$ then $\tan^{-1} x_k \to \frac{\pi}{2}$,
if $x_k \to -\infty$ then $\tan^{-1} x_k \to -\frac{\pi}{2}$.

Proof: All of these follow immediately from the previous theorem, except for the first statement in Part (4) (some care is needed when $\sin a = \pm 1$ or $\cos a = \pm 1$).

3.16 Example: Let $x_k = \frac{\sqrt{3k^2+1}}{k+2}$ for $k \geq 0$. Find $\lim\limits_{k\to\infty} x_k$.

Solution: We have $x_k = \frac{\sqrt{3k^2+1}}{k+2} = \frac{\sqrt{3+\frac{1}{k^2}}}{1+2\cdot\frac{1}{k}} \longrightarrow \frac{\sqrt{3+0}}{1+2\cdot 0} = \sqrt{3}$ where we used Basic Limits, Operations on Limits, and Functions Acting on Limits (specifically, we used Part (1) of Corollary 3.15 with $b = \frac{1}{2}$).

3.17 Example: Let $x_k = \frac{1+3k}{\sqrt[3]{2+k-k^2}}$ for $k \geq 0$. Find $\lim\limits_{k\to\infty} x_k$.

Solution: We have $x_k = \frac{1+3k}{\sqrt[3]{2+k-k^2}} = \frac{\frac{1}{k}+3}{\sqrt[3]{\frac{2}{k^2}+\frac{1}{k}-1}} \cdot k^{1/3} \longrightarrow \frac{0+3}{\sqrt[3]{0+0-1}} \cdot \infty = -1 \cdot \infty = -\infty$ where we used Basic Limits, Extended Operations, and Functions Acting on Limits.

3.18 Example: Let $x_k = \sin^{-1}\left(k - \sqrt{k^2+k}\right)$ for $k \geq 0$. Find $\lim\limits_{k\to\infty} x_k$.

Solution: Note that $k - \sqrt{k^2+k} = \frac{k^2-(k^2+k)}{k+\sqrt{k^2+k}} = \frac{-k}{k+\sqrt{k^2+k}} = \frac{-1}{1+\sqrt{1+\frac{1}{k}}} \longrightarrow \frac{-1}{1+\sqrt{1+0}} = -\frac{1}{2}$, and so $x_k = \sin^{-1}\left(k - \sqrt{k^2-k}\right) \longrightarrow \sin^{-1}\left(-\frac{1}{2}\right) = -\frac{\pi}{6}$.

3.19 Theorem: *(Comparison) Let $\langle x_k \rangle$ and $\langle y_k \rangle$ be sequences in an ordered field F. Suppose that $x_k \leq y_k$ for all k. Then*
(1) if $x_k \to a$ and $y_k \to b$ then $a \leq b$,
(2) if $x_k \to \infty$ then $y_k \to \infty$, and
(3) if $y_k \to -\infty$ then $x_k \to -\infty$.

Proof: We prove Part (1). Suppose that $x_k \to a$ and $y_k \to b$. Suppose, for a contradiction, that $a > b$. Choose $m_1 \in \mathbf{Z}$ so that $k \geq m_1 \to |x_k - a| \leq \frac{a-b}{3}$. Choose $m_2 \in \mathbf{Z}$ so that $k \geq m_2 \to |y_k - b| \leq \frac{a-b}{3}$. Let $k = \max\{m_1, m_2\}$. Since $|x_k - a| \leq \frac{a-b}{3} < \frac{a-b}{2}$, we have $x_k > a - \frac{a-b}{2} = \frac{a+b}{2}$. Since $|y_k - b| \leq \frac{a-b}{3} < \frac{a-b}{2}$, we have $y_k < b + \frac{a-b}{2} = \frac{a+b}{2}$. This is not possible since $x_k \leq y_k$.

3.20 Example: Let $x_k = (\frac{3}{2} + \sin k)\ln k$ for $k \geq 1$. Find $\lim\limits_{k\to\infty} x_k$.

Solution: For all $k \geq 1$ we have $\sin k \geq -1$ so $(\frac{3}{2} + \sin k) \geq \frac{1}{2}$ and hence $x_k \geq \frac{1}{2}\ln k$. Since $x_k \geq \frac{1}{2}\ln k$ for all $k \geq 1$ and $\frac{1}{2}\ln k \longrightarrow \frac{1}{2}\cdot\infty = \infty$, it follows that $x_k \to \infty$ by the Comparison Theorem.

3.21 Theorem: *(Squeeze) Let $\langle x_k \rangle$, $\langle y_k \rangle$ and $\langle z_k \rangle$ be sequences in an ordered field F.*
(1) If $x_k \leq y_k \leq z_k$ for all k and $x_k \to a$ and $z_k \to a$ then $y_k \to a$.
(2) If $|x_k| \leq y_k$ for all k and $y_k \to 0$ then $x_k \to 0$.

Proof: We prove Part (1). Suppose that $x_k \leq y_k \leq z_k$ for all k, and suppose that $x_k \to a$ and $z_k \to a$. Let $\epsilon > 0$. Choose $m_1 \in \mathbf{Z}$ so that $k \geq m_1 \to |x_k - a| \leq \epsilon$, choose $m_2 \in \mathbf{Z}$ so that $k \geq m_2 \to |z_k - a| \leq \epsilon$ and let $m = \max\{m_1, m_2\}$. Then for $k \geq m$ we have $a - \epsilon \leq x_k \leq y_k \leq z_k \leq a + \epsilon$ and so $|y_k - a| \leq \epsilon$. Thus $y_k \to a$, as required.

3.22 Example: Let $x_k = \frac{k+\tan^{-1} k}{2k+\sin k}$ for $k \geq 1$. Find $\lim\limits_{k\to\infty} x_k$.

Solution: For all $k \geq 1$ we have $-\frac{\pi}{2} < \tan^{-1} k < \frac{\pi}{2}$ and $-1 \leq \sin k \leq 1$ and so

$$\frac{k - \frac{\pi}{2}}{2k+1} \leq \frac{k + \tan^{-1} k}{2k + \sin k} \leq \frac{k + \frac{\pi}{2}}{2k-1}.$$

As in previous examples, we have $\frac{k-\frac{\pi}{2}}{2k+1} \to \frac{1}{2}$ and $\frac{k+\frac{\pi}{2}}{2k-1} \to \frac{1}{2}$, and so $x_k = \frac{k+\tan^{-1} k}{2k+\sin k} \to \frac{1}{2}$ by the Squeeze Theorem.

3.23 Definition: Let $\langle x_k \rangle$ be a sequence in an ordered set X. We say that the sequence $\langle x_k \rangle$ is **bounded above** by $b \in X$ when $x_k \leq b$ for all k. We say that the sequence $\langle x_k \rangle$ is **bounded below** by $b \in X$ when $b \leq x_k$ for all k. We say $\langle x_k \rangle$ is **bounded above** when it is bounded above by some element $b \in X$, we say that $\langle x_k \rangle$ is **bounded below** when it is bounded below by some $b \in X$, and we say that $\langle x_k \rangle$ is **bounded** when it is bounded above and bounded below.

3.24 Definition: Let $\langle x_k \rangle$ be a sequence in an ordered field F. We say that $\langle x_k \rangle$ is **increasing** (for $k \geq p$) when for all $k, l \in \mathbf{Z}_{\geq p}$, if $k \leq l$ then $x_k \leq x_l$. We say that $\langle x_k \rangle$ is **strictly increasing** (for $k \geq p$) when for all $k, l \in \mathbf{Z}_{\geq p}$, if $k < l$ then $x_k < x_l$. Similarly, we say that $\langle x_k \rangle$ is **decreasing** when for all $k, l \in \mathbf{Z}_{\geq p}$, if $k \leq l$ the $x_k \geq x_l$ and we say that $\langle x_k \rangle$ is **strictly decreasing** when for all $k, l \in \mathbf{Z}_{\geq p}$, if $k < l$ the $x_k > x_l$. We say that $\langle x_k \rangle$ is **monotonic** when it is either increasing or decreasing.

3.25 Theorem: *(Monotonic Convergence) Let $\langle x_k \rangle$ be a sequence in* $\mathbf{R}$.
(1) Suppose $\langle x_k \rangle$ is increasing. If $\langle x_k \rangle$ is bounded above then $x_k \to \sup\{x_k\}$, and if $\langle x_k \rangle$ is not bounded above then $x_k \to \infty$.
(2) Suppose $\langle x_k \rangle$ is decreasing. If $\langle x_k \rangle$ is bounded below then $x_k \to \inf\{x_k\}$, and if $\langle x_k \rangle$ is not bounded below then $x_k \to -\infty$.

Proof: We prove Part (1) in the case that $\langle x_k \rangle_{k \geq p}$ is increasing and bounded above, say by $b \in \mathbf{R}$. Let $A = \{x_k | k \geq p\}$ (so A is the range of the sequence $\langle x_k \rangle$). Note that A is nonempty and bounded above (indeed b is an upper bound for A). By the Completeness Property of $\mathbf{R}$, A has a supremum in $\mathbf{R}$. Let $a = \sup\{x_k | k \geq p\}$. Note that $a \geq x_k$ for all $k \geq p$ 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 $m \geq p$ so that the element $x_m \in A$ satisfies $a - \epsilon < x_m \leq a$. Since $\langle x_k \rangle$ is increasing, for all $k \geq m$ we have $x_k \geq x_m$, so we have $a - \epsilon \leq x_m \leq x_k \leq a$ and hence $|x_k - a| < \epsilon$. Thus $\lim\limits_{k \to \infty} x_k = a \leq b$.

3.26 Example: Let $x_1 = \frac{4}{3}$ and let $x_{k+1} = 5 - \frac{4}{x_n}$ for $k \geq 1$. Determine whether $\langle x_k \rangle$ converges, and if so then find the limit.

Solution: Suppose, for now, that $\langle x_k \rangle$ does converge, say $x_k \to a$. By Independence of Converge on Initial Terms, we also have $x_{k+1} \to a$. Using Operations on Limits, we have $a = \lim\limits_{k \to \infty} x_{k+1} = \lim\limits_{k \to \infty} \left(5 - \frac{4}{x_k}\right) = 5 - \frac{4}{a}$. Since $a = 5 - \frac{4}{a}$, we have $a^2 = 5a - 4$ or equivalently $(a-1)(a-4) = 0$. We have proven that if the sequence converges then its limit must be equal to 1 or 4.

The first few terms of the sequence are $x_1 = \frac{4}{3}$, $x_2 = 2$ and $x_3 = 3$. Since the terms appear to be increasing, we shall try to prove that $1 \leq x_n \leq x_{n+1} \leq 4$ for all $n \geq 1$. This is true when $n = 1$. Suppose it is true when $n = k$. Then we have

$$1 \leq x_k \leq x_{k+1} \leq 4 \to 1 \geq \tfrac{1}{x_k} \geq \tfrac{1}{x_{k+1}} \geq \tfrac{1}{4} \to -4 \leq -\tfrac{4}{x_k} \leq -\tfrac{4}{x_{k+1}} \leq -1$$
$$\to 1 \leq 5 - \tfrac{4}{x_k} \leq 5 - \tfrac{4}{x_{k+1}} \leq 4 \to 1 \leq x_{k+1} \leq x_{k+2} \leq 4.$$

Thus, by the Principle of Induction, we have $1 \leq x_n \leq x_{n+1} \leq 4$ for all $n \geq 1$.

Since $x_n \leq x_{n+1}$ for all $n \geq 1$, the sequence is increasing, and since $x_n \leq 4$ for all $n \geq 1$, the sequence is bounded above by 4. By the Monotone Convergence Theorem, the sequence does converge. By the first paragraph, we know the limit must be either 1 or 4, and since the sequence starts at $x_1 = 2$ and increases, the limit must be 4.

3.27 Theorem: *(The Nested Interval Theorem) Let $I_0, I_1, I_2, \cdots$ be nonempty, closed bounded intervals in $\mathbf{R}$. Suppose that $I_0 \supseteq I_1 \supset I_2 \supset \cdots$. Then $\bigcap_{k=0}^{\infty} I_k \neq \emptyset$.*

Proof: For each $k \geq 1$, let $I_k = [a_k, b_k]$ with $a_k < b_k$. For each k, since $I_k \subseteq I_{k+1}$ we have $a_{k+1} \leq a_k < b_k \leq b_{k+1}$. Since $a_k \geq a_{k+1}$ for all k, the sequence $\langle a_k \rangle$ is increasing. Since $a_k < b_k \leq b_{k-1} \leq \cdots \leq b_1$ for all k, the sequence $\langle a_k \rangle$ is bounded above by b_1. Since $\langle a_k \rangle$ is increasing and bounded above, it converges. Let $a = \sup\{a_k\} = \lim_{k\to\infty} a_k$. Similarly, $\langle b_k \rangle$ is decreasing and bounded below by a_1, and so it converges. Let $b = \inf\{b_k\} = \lim_{k\to\infty} b_k$. Fix $m \geq 1$. For all $k \geq m$ we have $a_m < b_m \leq b_{m+1} \leq \cdots \leq b_k$. Since $a_k \leq b_k$ for all k, by the Comparison Theorem we have $a \leq b$, and so the interval $[a, b]$ is not empty. Since $\langle a_k \rangle$ is increasing with $a_k \to a$, it follows (we leave the proof as an exercise) that $a_k \leq a$ for all $k \geq 1$. Similarly, we have $b_k \geq b$ for all $k \geq 1$ and so $[a, b] \subseteq [a_k, b_k] = I_k$. Thus $[a, b] \subseteq \bigcap_{k=1}^{\infty} I_k$, and so $\bigcap_{k=1}^{\infty} I_k \neq \emptyset$.

3.28 Definition: Let $\langle x_k \rangle_{k\geq p}$ be a sequence in a set X Given a strictly increasing function $f : \mathbf{Z}_{\geq q} \to \mathbf{Z}_{\geq p}$, write $k_l = f(l)$ and let $y_l = x_{k_l}$ for all $l \geq q$. Then the sequence $\langle y_l \rangle_{l\geq q}$ is called a **subsequence** of the sequence $\langle x_k \rangle_{k\geq p}$. In other words, a subsequence of $\langle x_k \rangle_{k\geq p}$ is a sequence of the form

$$\left\langle x_{k_q}, x_{k_{q+1}}, x_{k_{q+2}}, \cdots \right\rangle \text{ with } p \leq k_q < k_{q+1} < k_{q+2} < \cdots.$$

Given a bijective function $f : \mathbf{Z}_{\geq q} \to \mathbf{Z}_{\geq p}$, write $k_l = f(l)$ and let $y_l = x_{k_l}$ for $l \geq 1$. Then the sequence $\langle y_l \rangle_{l\geq q}$ is called a **rearrangement** of the sequence $\langle x_k \rangle$.

3.29 Theorem: *Let $\langle x_k \rangle$ be a sequence in an ordered field F. Suppose that $x_k \to a$. Then*

(1) every subsequence of $\langle x_k \rangle$ converges to a, and
(2) every rearrangement of $\langle x_k \rangle$ converges to a.

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

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

Proof: Let $\langle x_k \rangle$ be a bounded sequence in $\mathbf{R}$. Choose $a, b \in \mathbf{R}$ with $a \leq x_k$ for all k and $x_k \leq b$ for all k. Then we have $x_k \in [a, b]$ for all k. We define a sequence of nonempty closed intervals recursively as follows. Let $I_0 = [a_0, b_0] = [a, b]$. Note that $I_0 = \left[a, \frac{a+b}{2}\right] \cup \left[\frac{a+b}{2}, b\right]$. Let $I_1 = [a_1, b_1]$ be equal to one of the two intervals $\left[a, \frac{a+b}{2}\right]$ and $\left[\frac{a+b}{2}, b\right]$, chosen in such a way that there are infinitely many indices k with $x_k \in I_1$. Suppose we have chosen intervals $I_j = [a_j, b_j]$ with $b_j - a_j = \frac{1}{2^j}(b - a)$ for $1 \leq j \leq n$, such that $I_0 \supseteq I_1 \supseteq I_2 \supseteq \cdots \supseteq I_n$ and such that for each index j, there are infinitely many indices k with $x_k \in I_j$. 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 k with $x_k \in I_{n+1}$. In this way, we obtain a sequence $\langle I_j \rangle_{j\geq 0}$ of nonempty closed intervals.

By the Nested Interval Theorem, $\bigcap_{j=0}^{\infty} I_j$ is not empty. Choose a point c with $c \in I_n$ for every $n \geq 0$.

We shall now construct a subsequence of $\langle x_k \rangle$ which converges to c. Since for each $j \geq 0$ there exist infinitely many indices k with $x_k \in I_j$, we can construct a subsequence of $\langle x_k \rangle$ as follows. Choose k_0 so that $x_{k_0} \in I_0$, then choose $k_1 > k_0$ so that $x_{k_1} \in I_1$, then choose $k_2 > k_1$ with $x_{k_2} \in I_2$, and so on. In this way, we obtain a subsequence $\langle x_{k_j} \rangle_{j \geq 0}$ of $\langle x_k \rangle$ with $x_{k_j} \in I_j$ for all $j \geq 0$. We claim that $x_{k_j} \to c$ as $j \to \infty$. Let $\epsilon > 0$ Choose $m \in \mathbf{Z}$ so that $\frac{1}{2^m}(b-a) \leq \epsilon$. For $j \geq m$, since $c \in [a,b] \subseteq [a_j, b_j]$ and $x_{k_j} \in [a_j, b_j]$, it follows that

$$|x_{k_j} - c| = \max\{x_{k_j}, c\} - \min\{x_{k_j}, c\} \leq b_j - a_j = \tfrac{1}{2^j}(b-a) \leq \tfrac{1}{2^m}(b-a) \leq \epsilon.$$

Thus $x_{k_j} \to c$ as $j \to \infty$, as claimed.

3.31 Definition: Let $\langle x_k \rangle_{k \geq p}$ be a sequence in an ordered field F. We say that $\langle x_k \rangle$ is **Cauchy** when

$$\forall \epsilon > 0 \; \exists m \in \mathbf{Z} \; \forall k, l \in \mathbf{Z}_{\geq p} \; \big(k, l \geq m \to |x_k - x_l| \leq \epsilon\big).$$

3.32 Theorem: *(Cauchy Criterion for Convergence)*

(1) For a sequence $\langle x_k \rangle$ in an ordered field F, if $\langle x_k \rangle$ converges then it is Cauchy.
(2) For a sequence $\langle x_k \rangle$ in $\mathbf{R}$, if $\langle x_k \rangle$ is Cauchy then it converges.

Proof: To prove Part (1), let $\langle x_k \rangle$ be a sequence in an ordered field F and suppose that $x_k \to a$. Let $\epsilon > 0$ and choose $m \in \mathbf{Z}$ so that $k \geq m \to |x_k - a| \leq \frac{\epsilon}{2}$. Then for $k, l \geq m$ we h ave

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

Thus $\langle x_k \rangle$ is Cauchy.

To prove Part (2), let $\langle x_k \rangle_{k \geq p}$ be a sequence in $\mathbf{R}$ and suppose that $\langle x_k \rangle$ is Cauchy. We claim that $\langle x_k \rangle$ is bounded. Since $\langle x_k \rangle$ is Caucy, we can choose $m \in \mathbf{Z}$ so that $k, l \geq m \to |x_k - x_l| \leq 1$. In particular, for all $k \geq m$ we have $|x_k - x_m| \leq 1$ and so $|x_k| = |x_k - x_m + x_m| \leq |x_k - x_m| + |x_m| \leq 1 + |x_m|$. It follows that $\langle x_k \rangle$ is bounded by $b = \max\big\{|x_p|, |x_{p+1}|, \cdots, |x_{m-1}|, 1 + |x_m|\big\}$.

Because $\langle x_k \rangle$ is bounded, it has a convergent subsequence, by the Bolzano Weierstrass Theorem. Let $\langle x_{k_j} \rangle$ be a convergent subsequence of $\langle x_k \rangle$ and let $a = \lim_{j \to \infty} x_{k_j}$. We claim that $x_k \to a$. Let $\epsilon > 0$. Since $\langle x_k \rangle$ is Cauchy, we can choose $m \in \mathbf{Z}$ so that $k, l \geq m \to |x_k - x_l| \leq \frac{\epsilon}{2}$. Since $x_{k_j} \to a$ we can choose $m_0 \in \mathbf{Z}$ so that $j \geq m_0 \to |x_{k_j} - a| \leq \frac{\epsilon}{2}$. Choose an index $j \geq m_0$ so that $k_j \geq m$. Then for all $k \geq m$ we have

$$|x_k - a| = |x_k - x_{k_j} + x_{k_j} - a| \leq |x_k - x_{k_j}| + |x_{k_j} - a| \leq \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon.$$

Thus $x_k \to a$, as claimed.

Chapter 4: Limits of Functions

4.1 Definition: Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, and let $f : A \to F$. For $a \in F$, we say that a is a **limit point** of A when

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

When a is a limit point of A, we make the following definitions.

(1) For $b \in F$, 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

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

(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

$$\forall r \in F\ \exists \delta > 0\ \forall x \in A\ \big(0 < |x - a| \leq \delta \to f(x) \geq r\big).$$

(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

$$\forall r \in F\ \exists \delta > 0\ \forall x \in A\ \big(0 < |x - a| \leq \delta \to f(x) \leq r\big).$$

For $a \in F$, we say that a is a **limit point of A from below** when

$$\forall \delta > 0\ \exists x \in A \quad a - \delta \leq x < a\,.$$

When a is a limit point of A from below and $b \in F$, we make the following definitions.

(4) $\lim\limits_{x \to a^-} f(x) = b \iff \forall \epsilon > 0\ \exists \delta > 0\ \forall x \in A\ \big(a - \delta \leq x < a \to |f(x) - b| \leq \epsilon\big).$

(5) $\lim\limits_{x \to a^-} f(x) = \infty \iff \forall r \in F\ \exists \delta > 0\ \forall x \in A\ \big(a - \delta \leq x < a \to f(x) \geq r\big).$

(6) $\lim\limits_{x \to a^-} f(x) = -\infty \iff \forall r \in F\ \exists \delta > 0\ \forall x \in A\ \big(a - \delta \leq x < a \to f(x) \leq r\big).$

For $a \in F$, we say that a is a **limit point of A from above** when

$$\forall \delta > 0\ \exists x \in A \quad a < x \leq a + \delta\,.$$

When a is a limit point of A from above and $b \in F$, we make the following definitions.

(7) $\lim\limits_{x \to a^+} f(x) = b \iff \forall \epsilon > 0\ \exists \delta > 0\ \forall x \in A\ \big(a < x \leq a + \delta \to |f(x) - b| \leq \epsilon\big).$

(8) $\lim\limits_{x \to a^+} f(x) = \infty \iff \forall r \in F\ \exists \delta > 0\ \forall x \in A\ \big(a < x \leq a + \delta \to f(x) \geq r\big).$

(9) $\lim\limits_{x \to a^+} f(x) = -\infty \iff \forall r \in F\ \exists \delta > 0\ \forall x \in A\ \big(a < x \leq a + \delta \to f(x) \leq r\big).$

We say that infinity is a limit point of A (from below) when A is not bounded above, that is when $\forall m\in F\ \exists x\in A\ \ x\geq m$. When A is not bounded above and $b\in F$, we make the following definitions.

(10) $\lim\limits_{x\to\infty} f(x)=b \iff \forall \epsilon>0\ \exists m\in F\ \forall x\in A\ \big(x\geq m\to |f(x)-b|\leq\epsilon\big)$.

(11) $\lim\limits_{x\to\infty} f(x)=\infty \iff \forall r\in F\ \exists m\in F\ \forall x\in A\ \big(x\geq m\to f(x)\geq r\big)$.

(12) $\lim\limits_{x\to\infty} f(x)=-\infty \iff \forall r\in F\ \exists m\in F\ \forall x\in A\ \big(x\geq m\to f(x)\leq r\big)$.

We say that negative infinity is a limit point of A (from above) when A is not bounded below, that is when $\forall m\in F\ \exists x\in A\ \ x\leq m$. When A is not bounded below and $b\in F$, we make the following definitions.

(13) $\lim\limits_{x\to-\infty} f(x)=b \iff \forall \epsilon>0\ \exists m\in F\ \forall x\in A\ \big(x\leq m\to |f(x)-b|\leq\epsilon\big)$.

(14) $\lim\limits_{x\to-\infty} f(x)=\infty \iff \forall r\in F\ \exists m\in F\ \forall x\in A\ \big(x\leq m\to f(x)\geq r\big)$.

(15) $\lim\limits_{x\to-\infty} f(x)=-\infty \iff \forall r\in F\ \exists m\in F\ \forall x\in A\ \big(x\leq m\to f(x)\leq r\big)$.

4.2 Example: Let $f(x)=\dfrac{x^2+2x-3}{x^2-1}$. Show that $\lim\limits_{x\to1} f(x)=2$.

Solution: Note that for $x\neq1$ we have

$$|f(x)-2|=\left|\tfrac{x^2+2x-3}{x^2-1}-2\right|=\left|\tfrac{(x+3)(x-1)}{(x+1)(x-1)}-2\right|=\left|\tfrac{x+3}{x+1}-2\right|=\left|\tfrac{x+3-2x-2}{x+1}\right|=\left|\tfrac{-x+1}{x+1}\right|=\tfrac{|x-1|}{|x+1|}.$$

Let $\epsilon>0$. Choose $\delta=\min\{1,\epsilon\}$. Let $0<|x-1|\leq\delta$. Since $0<|x-1|$ we have $x\neq1$ so, as shown above, $|f(x)-2|=\frac{|x-1|}{|x+1|}$. Since $|x-1|\leq\delta\leq1$ we have $0\leq x\leq3$ so that $1\leq x+1$, and hence $|f(x)-2|=\frac{|x-1|}{|x+1|}\leq|x-1|$. Finally, since $|x-a|\leq\delta\leq\epsilon$ we have $|f(x)-2|\leq|x-1|\leq\epsilon$. Thus $\lim\limits_{x\to1} f(x)=2$.

4.3 Theorem: *(Two Sided Limits) Let F be a subfield of $\mathbf{R}$, let $A\subseteq F$, let $f:A\to F$ and let $a\in F$. Suppose that a is a limit point of A both from the left and from the right. Then for $u\in\hat{F}$ we have $\lim\limits_{x\to a} f(x)=a$ if and only if $\lim\limits_{x\to a^-} f(x)=u=\lim\limits_{x\to a^+} f(x)$.*

Proof: We prove the theorem in the case that $u=b\in F$. Suppose that $\lim\limits_{x\to a} f(x)=b\in F$. Let $\epsilon>0$. Choose $\delta>0$ so that for all $x\in A$, if $0<|x-a|\leq\delta$ then $|f(x)-b|\leq\epsilon$. For $x\in A$ with $a-\delta\leq x<a$ we have $0<|x-a|\leq\delta$ and so $|f(x)-b|\leq\epsilon$. This shows that $\lim\limits_{x\to a^-} f(x)=b$. For $x\in A$ with $a<x\leq x+\delta$ we have $0<|x-a|\leq\delta$ and so $|f(x)-b|\leq\epsilon$. This show that $\lim\limits_{x\to a^+} f(x)=b$.

Conversely, suppose that $\lim\limits_{x\to a^-} f(x)=b=\lim\limits_{x\to a^+} f(x)$. Let $\epsilon>0$. Since $f(x)\to b$ as $x\to a^-$, we can choose $\delta_1>0$ so that for all $x\in A$ with $a-\delta\leq a<a$ we have $|f(x)-b|\leq\epsilon$. Since $f(x)\to b$ as $x\to a^+$ we can choose $\delta_2>0$ so that for all $x\in A$ with $a<x\leq a+\delta_2$ we have $|f(x)-b|\leq\epsilon$. Let $\delta=\min\{\delta_1,\delta_2\}$. Let $x\in A$ with $0<|x-a|\leq\delta$. Either we have $x<a$ or we have $x>a$. In the case that $x<a$ we have $a-\delta_1\leq a-\delta\leq x<a$ and so $|f(x)-b|\leq\epsilon$ (by the choice of δ_1). In the case that $x>a$ we have $a<x\leq a+\delta\leq a+\delta_2$ and so $|f(x)-b|\leq\epsilon$ (by the choice of δ_2. In either case we have $|f(x)-b|\leq\epsilon$, and so $\lim\limits_{x\to a} f(x)=b$, as required.

4.4 Remark: For the sequence $\langle x_k \rangle_{k \geq p}$ in F given by $x_k = f(k)$ where $f : \mathbf{Z}_{\geq p} \to F$, the definitions (10), (11) and (12) agree with our definitions for limits of sequences. Thus limits of sequences are a special case of limits of functions. The following theorem shows that limits of functions are determined by limits of sequences.

4.5 Theorem: *(The Sequential Characterization of Limits of Functions) Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f : A \to F$, and let $u \in \hat{F}$.*

(1) When $a \in F$ is a limit point of A, $\lim\limits_{x \to a} f(x) = u$ if and only if for every sequence $\langle x_k \rangle$ in $A \setminus \{a\}$ with $x_k \to a$ we have $f(x_k) \to u$.

(2) When a is a limit point of A from below, $\lim\limits_{x \to a^-} f(x) = u$ if and only if for every sequence $\langle x_k \rangle$ in $\{x \in A | x < a\}$ with $x_k \to a$ we have $f(x_k) \to u$.

(3) When a is a limit point of A from above, $\lim\limits_{x \to a^+} f(x) = u$ if and only if for every sequence $\langle x_k \rangle$ in $\{x \in A | x > a\}$ with $x_k \to a$ we have $f(x_k) \to u$.

(4) When A is not bounded above, $\lim\limits_{x \to \infty} f(x) = u$ if and only if for every sequence $\langle x_k \rangle$ in A with $x_k \to \infty$ we have $f(x_k) \to u$.

(5) When A is not bounded below, $\lim\limits_{x \to -\infty} f(x) = u$ if and only if for every sequence $\langle x_k \rangle$ in A with $x_k \to -\infty$ we have $f(x_k) \to u$.

Proof: We prove Part (1) in the case that $u = b \in F$. Let $a \in F$ be a limit point of A. Suppose that $\lim\limits_{x \to a} f(x) = b \in F$. Let $\langle x_k \rangle$ 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 $0 < |x - a| \leq \delta \to |f(x) = b| \leq \epsilon$. Since $x_k \to a$ we can choose $m \in \mathbf{Z}$ so that $k \geq m \to |x_k - a| \leq \delta$. Then for $k \geq m$, we have $|x_k - a| \leq \delta$ and we have $x_k \neq a$ (since the sequence $\langle x_k \rangle$ is in the set $A \setminus \{a\}$) so that $0 < |x - a| \leq \delta$ and hence $|f(x_k) - b| \leq \epsilon$. 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 $0 < |x - a| \leq \delta$ and $|f(x) - b| > \epsilon_0$. For each $k \in \mathbf{Z}^+$, choose $x_k \in A$ with $0 < |x_k - a| \leq \frac{1}{k}$ and $|f(x_k) - b| > \epsilon_0$. In this way we obtain a sequence $\langle x_k \rangle_{k \geq 1}$ in $A \setminus \{a\}$ (we remark that the Axiom of Choice is required to construct this sequence $\langle x_k \rangle$). Since $|x_k - a| \leq \frac{1}{k}$ for all $k \in \mathbf{Z}^+$, it follows that $x_k \to a$ (indeed, given $\epsilon > 0$ we can choose $m \in \mathbf{Z}$ with $m \geq \frac{1}{\epsilon}$ and then $k \geq m \to |x_k - a| \leq \frac{1}{k} \leq \frac{1}{m} \leq \epsilon$). Since $|f(x_k) - b| > \epsilon_0$ for all k, it follows that $f(x_k) \not\to b$ (indeed if we had $f(x_k) \to b$ we could choose $m \in \mathbf{Z}$ so that $k \geq m \to |f(x_k) - b| \leq \epsilon_0$ and then we could choose $k = m$ to get $|f(x_k) - b| \leq \epsilon_0$).

4.6 Remark: It follows from the Sequential Characterization of Limits of Functions that all of our theorems about limits of sequences imply analogous theorems in the more general setting of limits of functions. We list several of those theorems and give one sample proof.

4.7 Theorem: *(Local Determination of Limits) Let F be a subfield of $\mathbf{R}$, let $A, B \subseteq F$, let $f : A \to F$ and let $g : B \to F$. Suppose that $a \in F$ is a limit point of both sets A and B, and that for some $\delta > 0$ we have $C = \{x \in A \big| 0 < |x - a| \leq \delta\} \subseteq \{x \in B \big| 0 < |x - a| \leq \delta\}$ and that $f(x) = g(x)$ for all $x \in C$. Then if $\lim\limits_{x \to a} g(x) = u \in \hat{F}$ then $\lim\limits_{x \to a} f(x) = u$.*
Similar results holds for limits $x \to a^{\pm}$ and $x \to \pm\infty$.

4.8 Theorem: *(Uniqueness of Limits) Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f : A \to F$, and let a be a limit point of A. For $u, v \in \hat{F}$, if $\lim\limits_{x \to a} f(x) = u$ and $\lim\limits_{x \to a} f(x) = v$ then $u = v$.*
Similar results hold for limits $x \to a^{\pm}$ and $x \to \pm\infty$.

4.9 Theorem: *(Extended Operations on Limits) Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f, g : A \to F$ and let a be a limit point of A. Let $u, v \in \hat{F}$ and suppose that $\lim\limits_{x\to a} f(x) = u$ and $\lim\limits_{x\to a} g(x) = v$. Then*

(1) if $u+v$ is defined in $\hat{F}$ then $\lim\limits_{x\to a}(f+g)(x) = u+v$,

(2) if $u-v$ is defined in $\hat{F}$ then $\lim\limits_{x\to a}(f-g)(x) = u-v$,

(3) if $u\cdot v$ is defined in $\hat{F}$ then $\lim\limits_{x\to a}(fg)(x) = u\cdot v$, and

(4) if u/v is defined in $\hat{F}$ then $\lim\limits_{x\to a}(f/g)(x) = u/v$.

Similar results hold for limits $x \to a^{\pm}$ and $x \to \pm\infty$.

Proof: We prove Part (4). Suppose that u/v is defined in $\hat{F}$. Let $\langle x_k\rangle$ be any sequence in $A\setminus\{a\}$ with $x_k \to a$. By the Sequential Characterization of Limits, since $\lim\limits_{x\to a} f(x) = u$ we have $f(x_k) \to u$, and since $\lim\limits_{x\to a} g(x) = v$ we have $f(x_k) \to v$. By Extended Operations on Limits of Sequences (Theorem 3.13), since $f(x_k) \to u$ and $g(x_k) \to v$ and u/v is defined in $\hat{F}$, we have $(f/g)(x_k) = \frac{f(x_k)}{g(x_k)} \to u/v$. Thus $(f/g)(x_k) \to u/v$ for every sequence $\langle x_k\rangle$ in $A \setminus \{a\}$ with $x_k \to a$. By the Sequential Characterization of Limits, it follows that $\lim\limits_{x\to a}(f/g)(x) = u/v$.

4.10 Theorem: *(Basic Limits) Let F be a subfield of $\mathbf{R}$, and let $A \subseteq F$. For the constant function $f : A \to F$ given by $f(x) = b$ for all $x \in A$, we have*

$$\lim_{x\to a} f(x) = b\ ,\ \lim_{x\to a^+} f(x) = b\ ,\ \lim_{x\to a^-} f(x) = b\ ,\ \lim_{x\to\infty} f(x) = b \ \text{ and } \lim_{x\to-\infty} f(x) = b,$$

and for the identity function $f : A \to F$ given by $f(x) = x$ for all $x \in A$ we have

$$\lim_{x\to a} f(x) = a\ ,\ \lim_{x\to a^+} f(x) = a\ ,\ \lim_{x\to a^-} f(x) = a\ ,\ \lim_{x\to\infty} f(x) = \infty \ \text{ and } \lim_{x\to-\infty} f(x) = -\infty$$

whenever the limits are defined.

4.11 Theorem: *(Basic Elementary Functions Acting on Limits) For $f : A \subseteq \mathbf{R} \to \mathbf{R}$ and for $a, b, c \in \mathbf{R}$ with a a limit point of A, we have the following.*

(1) If $\lim\limits_{x\to a} f(x) = b > 0$ then $\lim\limits_{x\to a} f(x)^c = b^c$,

$$\text{if } \lim_{x\to a} f(x) = \infty \text{ then } \lim_{x\to a} f(x)^c = \begin{cases} \infty \text{ if } c > 0 \\ \ 1 \text{ if } c = 0 \\ \ 0 \text{ if } c < 0, \end{cases}$$

$$\text{if } f(x) > 0 \text{ for all } x \in A \text{ and } \lim_{x\to a} f(x) = 0 \text{ then } \lim_{x\to a} f(x)^c = \begin{cases} \ 0 \text{ if } c > 0 \\ \ 1 \text{ if } c = 0 \\ \infty \text{ if } c < 0. \end{cases}$$

(2) If $\lim\limits_{x\to a} f(x) = b$ and $c > 0$ then $\lim\limits_{x\to a} c^{f(x)} = c^b$,

$$\text{if } \lim_{x\to a} f(x) = \infty \text{ and } c > 0 \text{ then } \lim_{x\to\infty} c^{f(x)} = \begin{cases} \infty \text{ if } c > 1 \\ \ 1 \text{ if } c = 1 \\ \ 0 \text{ if } 0 < c < 1, \end{cases}$$

$$\text{if } \lim_{x\to a} f(x) = -\infty \text{ and } c > 0 \text{ then } \lim_{x\to a} c^{f(x)} = \begin{cases} 0 \text{ if } c > 1 \\ 1 \text{ if } c = 1 \\ 0 \text{ if } 0 < c < 1. \end{cases}$$

(3) If $\lim_{x\to a} f(x) = b > 0$ *and* $c > 0$ *then* $\lim_{x\to a} \log_c f(x) = \log_c b$,

if $\lim_{x\to a} f(x) = \infty$ *and* $c > 0$ *then* $\lim_{x\to a} \log_c f(x) = \begin{cases} \infty \text{ if } c > 1 \\ -\infty \text{ if } 0 < c < 1, \end{cases}$

if $f(x) > 0$ *for all* $x \in A$, $\lim_{x\to a} f(x) = 0$ *and* $c > 0$ *then* $\lim_{x\to a} \log_c f(x) = \begin{cases} -\infty \text{ if } c > 1 \\ \infty \text{ if } 0 < c < 1. \end{cases}$

(4) If $\lim_{x\to a} f(x) = b$ *then* $\lim_{x\to a} \sin f(x) = \sin b$ *and* $\lim_{x\to a} \cos f(x) = \cos b$,

the limits $\lim_{x\to\pm\infty} \sin x$, $\lim_{\pm\infty} \cos x$ *and* $\lim_{x\to\pm\infty} \tan x$ *do not exist.*

(5) If $f(x) \in [-1, 1]$ *for all* $x \in A$ *and* $\lim_{x\to a} f(x) = b$ *then* $\lim_{x\to a} \sin^{-1} f(x) = \sin^{-1} b$,

if $\lim_{x\to a} f(x) = b \in \mathbf{R}$ *then* $\lim_{x\to a} \tan^{-1} f(x) = \tan^{-1} b$,

if $\lim_{x\to a} f(x) = \infty$ *then* $\lim_{x\to a} \tan^{-1} f(x) = \frac{\pi}{2}$, *and*

if $\lim_{x\to a} f(x) = -\infty$ *then* $\lim_{x\to a} \tan^{-1} f(x) = -\frac{\pi}{2}$.

Similar results hold for limits $x \to a^\pm$ *and* $x \to \pm\infty$.

4.12 Example: Evaluate each of the following limits, if they exist.

(a) $\lim_{x\to 3} \dfrac{\sqrt{x+1} - 2}{3 - x}$

(b) $\lim_{x\to 1} \sin^{-1}\left(\dfrac{2}{x-1} - \dfrac{x+3}{x^2-1}\right)$

(c) $\lim_{x\to 0} e^{-1/x^2}$

(d) $\lim_{x\to\infty} \dfrac{(3x+1)\sqrt{x}}{\sqrt{4x^3 - 2x + 1}}$

(e) $\lim_{x\to 1^-} \dfrac{\sqrt{x^3 - 2x^2 + x}}{x^2 + 2x - 3}$

(f) $\lim_{x\to -1^+} \dfrac{x^2 - 2x - 3}{x^3 + 4x^2 + 5x + 2}$

Solution: I may include solutions later.

4.13 Theorem: *(The Comparison Theorem) Let* F *be a subfield of* $\mathbf{R}$, *let* $A \subseteq F$, *let* f *and* g *be two functions* $f, g : A \to F$ *and let* $a \in F$ *be a limit point of* A. *Suppose that* $f(x) \le g(x)$ *for all* $x \in A$. *Then*

(1) if $\lim_{x\to a} f(x) = u$ *and* $\lim_{x\to a} f(x) = v$ *with* $u, v \in \hat{F}$, *then* $u \le v$,

(2) if $\lim_{x\to a} f(x) = \infty$ *then* $\lim_{x\to a} g(x) = \infty$, *and*

(3) if $\lim_{x\to a} g(x) = -\infty$ *then* $\lim_{x\to a} g(x) = -\infty$.

Similar results hold for limits $x \to a^\pm$ *and* $x \to \pm\infty$.

4.14 Theorem: *(The Squeeze Theorem) Let* F *be a subfield of* $\mathbf{R}$, *let* $A \subseteq F$, *let* $f, g, h : A \to F$, *and let* a *be a limit point of* A.

(1) If $f(x) \le g(x) \le h(x)$ *for all* $x \in A$ *and* $\lim_{x\to a} f(x) = b = \lim_{x\to a} h(x)$, *then* $\lim_{x\to a} g(x) = b$.

(2) If $|f(x)| \le g(x)$ *for all* $x \in A$ *and* $\lim_{x\to a} g(x) = 0$ *then* $\lim_{x\to a} f(x) = 0$.

Similar results hold for limits $x \to a^\pm$ *and* $x \to \pm\infty$.

4.15 Definition: Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, and let $f : A \to F$. For $a \in A$, we say that f is **continuous** at a when

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

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

4.16 Theorem: *Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f : A \to F$ and let $a \in A$. Then*
(1) if a is not a limit 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 $\lim\limits_{x \to a} f(x) = f(a)$.

Proof: The proof is left as an exercise.

4.17 Theorem: *(The Sequential Characterization of Continuity) Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f : A \to F$ and let $a \in A$. Then f is continuos at a if and only if for every sequence $\langle x_k \rangle$ in A with $x_k \to a$ we have $f(x_k) \to f(a)$.*

Proof: Suppose that f is continuous at a. Let $\langle x_k \rangle$ 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 $|x - a| \leq \delta \to |f(x) - f(a)| \leq \epsilon$. Choose $m \in \mathbf{Z}$ so that for all indices k we have $k \geq m \to |x_k - a| \leq \delta$. Then when $k \geq m$ we have $|x_k - a| \leq \delta$ and hence $|f(x_k) - f(a)| \leq \epsilon$. 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 $|x - a| \leq \delta$ and $|f(x) - f(a)| > \epsilon_0$. For each $k \in \mathbf{Z}^+$, choose $x_k \in A$ with $|x_k - a| \leq \frac{1}{k}$ and $|f(x_k) - f(a)| > \epsilon_0$. Consider the sequence $\langle x_k \rangle$ in A (we remark that the Axiom of Choice is being used here). Since $|x_k - a| \leq \frac{1}{k}$ for all $k \in \mathbf{Z}^+$, it follows that $x_k \to a$. Since $|f(x_k) - f(a)| > \epsilon_0$ for all $k \in \mathbf{Z}^+$, it follows that $f(x_k) \not\to f(a)$.

4.18 Theorem: *(Operations on Continuous Functions) Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f, g : A \to F$, let $a \in A$ and let $c \in F$. Suppose that f and g are continuous at a. Then the functions cf, $f + g$, $f - g$ and fg are all continuous at a, and if $g(a) \neq 0$ then the function f/g is continuous at a.*

Proof: The proof is left as an exercise.

4.19 Theorem: *(Composition of Continuous Functions) Let F be a subfield of $\mathbf{R}$, let $A, B \subseteq \mathbf{R}$, let $f : A \to \mathbf{R}$ and $g : B \to \mathbf{R}$, let $h = g \circ f : C \to \mathbf{R}$ where $C = A \cap f^{-1}(B)$.*
(1) If f is continuous at $a \in C$ and g is continuous at $f(a)$, then h is continuous at a.
(2) If f is continuous (in A) and g is continuous (in B) then h is continuous (in C).

Proof: Note that Part (2) follows immediately from Part (1), so it suffices to prove Part (1). Suppose that f is continuous at $a \in A$ and g is continuous at $b = f(a) \in B$. Let $\langle x_k \rangle$ be a sequence in C with $x_k \to a$. Since f is continuous at a, we have $f(x_k) \to f(a) = b$ by the Sequential Characterization of Continuity. Since $\big\langle f(x_k) \big\rangle$ is a sequence in B with $f(x_k) \to b$ and since g is continuous at b, we have $g(f(x_k)) \to g(b)$ by the Sequential Characterization of Continuity. Thus we have $h(x_k) = g\big(f(x_k)\big) \to g(b) = g\big(f(a)\big) = h(a)$. We have shown that for every sequence $\langle x_k \rangle$ in C with $x_k \to a$ we have $h(x_k) \to h(a)$. Thus h is continuous at a by the Sequential Characterization of Continuity.

4.20 Corollary: *Every elementary function is continuous (in its domain).*

Proof: The basic elementary functions are all continuous in their domains by the Basic Elementary Functions Acting on Limits Theorem. If follows that every elementary function is continuous by Theorems 4.18 and 4.19.

4.21 Theorem: *(Functions Acting on Limits) Let F be a subfield of $\mathbf{R}$, let $A, B \subseteq F$, let $f : A \to F$, let $g : B \to F$ and let $h = g \circ f : C \to F$ where $C = A \cap f^{-1}(B)$. Let a be a limit point of C (hence also of A) and let b be a limit point of B. Suppose that $\lim\limits_{x \to a} f(x) = a$ and $\lim\limits_{y \to b} g(y) = c$. Suppose either that $f(x) \neq b$ for all $x \in C \setminus \{a\}$ or that g is continuous at $b \in B$. Then $\lim\limits_{x \to a} h(x) = c$.*

Analogous results hold, dealing with limits $x \to a^{\pm}$, $x \to \pm\infty$, $y \to b^{\pm}$ and $y \to \pm\infty$.

Proof: The proof is similar to the proof of the Composition of Continuous Functions Theorem.

4.22 Theorem: *(Intermediate Value Theorem) Let I be an interval in $\mathbf{R}$ and let $f : I \to \mathbf{R}$ be continuous. Let $a, b \in I$ with $a \leq b$ and let $y \in \mathbf{R}$. 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 $\mathbf{R}$. 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 $|t - x| \leq \delta_2 \to f(t) > y$. Let $\delta = \min\{\delta_1, \delta_2\}$. Since $x = \sup A$, by the Approximation Property we can choose $t \in A$ with $x - \delta \leq t \leq x$. 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 $|t - x| \leq \delta_2 \to f(t) < y$. 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.

4.23 Example: Define $f : \mathbf{Q} \to \mathbf{Q}$ be $f(x) = x^2$. For $a = 0$ and $b = 2$ and $y = 2$ we have $f(a) < y < f(b)$ but there is no point x in the rational interval $[a, b] = \{t \in \mathbf{Q} | a \leq t \leq b\}$ for which $f(x) = y$. So the conclusion of the Intermediate Value Theorem does not hold in this case.

4.24 Definition: Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, and let $f : A \to F$. 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 $b \in A$.

4.25 Theorem: *(Extreme Value Theorem) Let $a, b \in \mathbf{R}$ with $a < b$, and let $f : [a, b] \to \mathbf{R}$ 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 $k \in \mathbf{Z}^+$, choose $x_k \in [a, b]$ such that $f(x_k) \geq k$. By the Bolzano Weierstrass Theorem, we can choose a convergent subsequence $\langle x_{k_j} \rangle$. 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 $f(x_{k_j}) \to f(p) \in \mathbf{R}$, 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 $k \in \mathbf{Z}^+$ 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 $\langle y_{k_j} \rangle$. 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.

4.26 Example: For the function $f : [-1, 1] \subseteq \mathbf{R} \to \mathbf{R}$ given by $f(x) = x^3 - x$, you can check using high school calculus that f attains its maximum and minimum values at $a = -\frac{1}{\sqrt{3}}$ and $b = \frac{1}{\sqrt{3}}$. The function $f : [-1, 1] \subseteq \mathbf{Q} \to \mathbf{Q}$ is continuous in the closed rational interval $[-1, 1] = \{t \in \mathbf{Q} | -1 \leq t \leq 1\}$, but it does not attain its maximum and minimum values in this interval, so the conclusion of the Extreme Value Theorem does not hold for this function.

4.27 Definition: Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, and let $f : A \to F$. We say that f is **uniformly continuous** in A when

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

4.28 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 $|x - a| = \frac{2}{3} \leq \delta$ but $|f(x) - f(a)| = 2 > \epsilon$. If $0 < \delta < 1$ then for $x = \frac{\delta}{3}$ and $a = \delta$ we have $|x - a| = \frac{2}{3}\delta \leq \delta$ but $|f(x) - f(a)| = \frac{2}{\delta} \geq 2 > \epsilon$. This proves that f is not uniformly continuous (but f is continuous because it is elementary).

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

Proof: Suppose, for a contradiction, that $f : [a, b] \to \mathbf{R}$ 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 $|x - y| \leq \delta$ but $|f(x) - f(y)| > \epsilon$. For each $k \in \mathbf{Z}^+$ choose x_k and y_k in $[a, b]$ with $|x_k - y_k| \leq \frac{1}{k}$ and $|f(x_k) - f(y_k)| > \epsilon$. By the Bolzano Weierstrass Theorem, we can choose a convergent subsequence $\langle y_{k_j} \rangle$ of $\langle y_k \rangle$. 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 $|f(x_{k_j}) - f(y_{k_j})| \leq \epsilon$, giving the desired contradiction.

Chapter 5: Differentiation

5.1 Definition: Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f : A \to F$ and let $a \in A$ be a limit point of 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 F. In this case we call the limit the **derivative** of f at a, and we denote to by $f'(a)$, so we have

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

When $a \in A$ is a limit point of A from the right, we say that f is **differentiable from the right** at a and that $f'_+(a)$ is the **derivative from the right** of f at a, when

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

Similarly, when $a \in A$ is a limit point of A from the left, we say that f is **differentiable from the left** at a and that $f'_-(a)$ is the **derivative from the left** of f at a when

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

5.2 Definition: We say that f is **differentiable** (in A) when f is differentiable at every point $a \in A$. In this case, the **derivative** of f is the function $f' : A \to F$ 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** (in A), and the function $f'' : A \to F$ is called the **second derivative** of f. Similarly, $f'''(a)$ is the derivative of f'' at a and so on. More generally, for any function $f : A \to F$, we define its **derivative** to be the function $f' : B \to F$ where $B = \big\{a \in A \big| f \text{ is differentiable at } a\big\}$, and we define its **second derivative** to be the function $f'' : C \to F$ where $C = \big\{a \in B \big| f' \text{ is differentiable at } a\big\}$ and so on.

5.3 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 F if and only if the limit on the right exists in F, and in this case the two limits are equal.

5.4 Theorem: *Let F be a subfield of* $\mathbf{R}$, *let $A \subseteq F$, let $f : A \to F$, and let $a \in A$ be a limit point of A. Then f is differentiable at a with derivative $f'(a)$ if and only if*

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

Proof: We have

$$\begin{aligned} & f \text{ is differentiable at } a \text{ with derivative } f'(a) \iff \lim_{x \to a} \frac{f(x)-f(a)}{x-a} = f'(a) \\ & \iff \forall \epsilon > 0 \ \exists \delta > 0 \ \forall x \in A \left(0 < |x-a| \leq \delta \to \left| \frac{f(x)-f(a)}{x-a} - f'(a) \right| \leq \epsilon \right) \\ & \iff \forall \epsilon > 0 \ \exists \delta > 0 \ \forall x \in A \left(0 < |x-a| \leq \delta \to \left| \frac{f(x)-f(a)-f'(a)(x-a)}{x-a} \right| \leq \epsilon \right) \\ & \iff \forall \epsilon > 0 \ \exists \delta > 0 \ \forall x \in A \Big(0 < |x-a| \leq \delta \to \big| f(x)-f(a)-f'(a)(x-a) \big| \leq \epsilon |x-a| \Big) \\ & \iff \forall \epsilon > 0 \ \exists \delta > 0 \ \forall x \in A \Big(|x-a| \leq \delta \to \big| f(x)-f(a)-f'(a)(x-a) \big| \leq \epsilon |x-a| \Big) \end{aligned}$$

where on the last line, we can remove the condition that $0 < |x-a|$ because when $x = a$ we have $\big| f(x) - f(a) - f'(a)(x-a) \big| = 0$.

5.5 Definition: When $f : A \to F$ is differentiable at a with dervative $f'(a)$, the function

$$l(x) = f(a) + f'(a)(x-a)$$

is called the **linearization** of f at a. Note that the graph $y = l(x)$ of the linearization is the line through the point $(a, f(a))$ with slope $f'(a)$. This line is called the **tangent line** to the graph $y = f(x)$ at the point $(a, f(a))$.

5.6 Theorem: *(Differentiability Implies Continuity) Let F be a subfield of* $\mathbf{R}$, *let $A \subseteq F$, let $f : A \to F$ and let $a \in A$ be a limit point of A. Suppose that f is differentiable at a. Then f is continuous at a.*

Proof: We have

$$f(x) - f(a) = \frac{f(x)-f(a)}{x-a} \cdot (x-a) \longrightarrow f'(a) \cdot 0 = 0 \ \text{ as } x \to a$$

and so

$$f(x) = \big(f(x) - f(a) \big) + f(a) \longrightarrow 0 + f(a) = f(a) \ \text{ as } x \to a.$$

This proves that f is continuous at a.

5.7 Theorem: *(Local Determination of the Derivative) Let F be a subfield of* $\mathbf{R}$, *let $A, B \subseteq F$, let $f : A \to F$ and $g : B \to F$, and let $a \in A \cap B$ be a limit point of both A and B. Suppose that for some $\delta > 0$ we have $\big\{ x \in A \big| |x-a| \leq \delta \big\} \subset \big\{ x \in B \big| |x-a| \leq \delta \big\}$. If g is differentiable at a then so is f and we have $f'(a) = g'(a)$.*

Proof: The proof is left as an exercise.

5.8 Theorem: *(Operations on Derivatives) Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f, g : A \to F$, let $a \in A$ be a limit point of A, and let $c \in F$. Suppose that f and g are differentiable at a. Then*

(1) (Linearity) the functions cf, $f+g$ and $f-g$ are differentiable at a with

$$(cf)'(a) = c\,f'(a) \ , \ (f+g)'(a) = f'(a) + g'(a) \ , \ (f-g)'(a) - f'(a) - g'(a),$$

(2) (Product Rule) the function fg is differentiable at a with

$$(fg)'(a) = f'(a)g(a) + f(a)g'(a),$$

(3) (Reciprocal Rule) if $g(a) \neq 0$ then the function $1/g$ is differentiable at a with

$$\left(\frac{1}{g}\right)'(a) = -\frac{g'(a)}{g(a)^2},$$

(4) (Quotient Rule) if $g(a) \neq 0$ then the function f/g is differentiable at a with

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.$$

Proof: We prove Parts (2), (3) and (4). For $x \in A$ with $x \neq a$, we have

$$\begin{aligned}\frac{(fg)(x) - (fg)(a)}{x-a} &= \frac{f(x)g(x) - f(a)g(a)}{x-a} \\ &= \frac{f(x)g(x) - f(x)g(a) + f(x)g(a) - f(a)g(a)}{x-a} \\ &= f(x) \cdot \frac{g(x) - g(a)}{x-a} + g(a) \cdot \frac{f(x) - f(a)}{x-a} \\ &\longrightarrow f(a) \cdot g'(a) + g(a) \cdot f'(a) \ \text{ as } x \to a.\end{aligned}$$

Note that $f(x) \to f(a)$ as $x \to a$ because f is continuous at a since differentiability implies continuity. This proves the Product Rule.

Suppose that $g(a) \neq 0$. Since g is continuous at a (because differentiability implies continuity) we can choose $\delta > 0$ so that $|x-a| \leq \delta \to |g(x) - g(a)| \leq \frac{|g(a)|}{2}$ and then when $|x-a| \leq \delta$ we have $|g(x)| \geq \frac{|g(a)|}{2}$ so that $g(x) \neq 0$. For $x \in A$ with $|x-a| \leq \delta$ we have

$$\frac{\left(\frac{1}{g}\right)(x) - \left(\frac{1}{g}\right)(a)}{x - -a} = \frac{\frac{1}{g(x)} - \frac{1}{g(a)}}{x-a} = \frac{-1}{g(x)g(a)} \cdot \frac{g(x) - g(a)}{x-a} \longrightarrow \frac{-1}{g(a)^2} \cdot g'(a)$$

as $x \to a$. This Proves the Reciprocal Rule.

Finally, note that Part (4) follows from Parts (2) and (3). Indeed when $g(a) \neq 0$, we have

$$\begin{aligned}\left(\frac{f}{g}\right)'(a) &= \left(f \cdot \frac{1}{g}\right)'(a) = f'(a) \cdot \left(\frac{1}{g}\right)(a) + f(a) \cdot \left(\frac{1}{g}\right)'(a) \\ &= f'(a) \cdot \frac{1}{g(a)} + f(a) \cdot \frac{-g'(a)}{g(a)^2} = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}.\end{aligned}$$

5.9 Theorem: *(Chain Rule) Let F be a subfield of $\mathbf{R}$, let $A, B \subseteq F$, let $f : A \to F$, let $g : B \to F$ and let $h = g \circ f : C \to F$ where $C = A \cap f^{-1}(B)$. Let $a \in C$ be a limit point of C (hence also of A) and let $b = f(a) \in B$ be a limit point of 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 from Theorem 5.3. Note first that for $x \in C$ and $y = f(x) \in B$ we have

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

and also

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

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

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

Since f is continuous at a, we can choose δ_1 so that

$$|x-a| \le \delta_1 \to |f(x) - f(a)| \le \delta_0 \to |y-b| \le \delta_0.$$

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

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

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

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

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

5.10 Theorem: *Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$ and let $f : A \to F$. Then f is monotonic if and only if f has the property that for all $a, b, c \in A$, if b lies between a and c then $f(b)$ lies between $f(a)$ and $f(c)$.*

Proof: The proof is left as an exercise.

5.11 Theorem: *(The Inverse Function Theorem) Let I be an interval in $\mathbf{R}$, let $f : I \to \mathbf{R}$ and let $J = f(I)$.*

(1) If f is continuous then its range $J = f(I)$ is an interval in $\mathbf{R}$.
(2) If f is injective and continuous then f is strictly monotonic.
(3) If $f : I \to J$ is strictly monotonic, then so is its inverse $g : J \to I$.
(4) If f is bijective and continuous then its inverse g is continuous.

(5) If f is bijective and continuous, and f is differentiable at a with $f'(a) \neq 0$, then its inverse g is differentiable at $b = f(a)$ with $g'(b) = \frac{1}{f'(a)}$.

Proof: Suppose that $f : I \to \mathbf{R}$ is continuous. If f is the empty function or if f is constant, then J is a degenerate interval. Suppose that J contains at least two points. Let $u, v \in J$ and let $y \in \mathbf{R}$ with $u < y < v$. Since $J = f(I)$ we can choose $a, b \in I$ with $f(a) = u$ and $f(b) = v$. Since $f(a) = u \neq v = f(b)$ we have $a \neq b$. Since y lies between $f(a) = u$ and $f(b) = v$, and since f is continuous, it follows from the Intermediate Value Theorem that we can choose x between a and b with $f(x) = y$. Since I is an interval in $\mathbf{R}$, it has the intermediate value property, and so we have $x \in I$. Since $x \in I$ and $y = f(x)$ we have $y \in f(I) = J$. This proves that J has the intermediate value property, and so J is an interval, as required. This proves Part (1).

Suppose that f is injective and continuous. Let $a, b, c \in I$ with $a < b < c$. Since f is injective and $a \neq c$, we have $f(a) \neq f(c)$. We claim that $f(b)$ lies between $f(a)$ and $f(c)$. Consider the case that $f(a) < f(c)$ (the case that $f(a) > f(c)$ is similar). 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 $x_2 \in b$ 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 $f(b) \geq f(c)$ and so we have $f(b) < f(c)$. A similar argument by contradiction shows that we cannot have $f(b) \leq f(a)$ and so we have $f(a) < f(b) < f(c)$, and so $f(b)$ lies between $f(a)$ and $f(c)$ as claimed. We have proven that for all $a, b, c \in I$ with $a < b < c$, $f(b)$ lies between $f(a)$ and $f(c)$. It follows from the above theorem that f is monotonic (hence strictly monotonic since it is injective). This proves Part (2).

To prove Part (3), suppose that $f : I \to J$ is strictly monotonic and let $g : J \to I$ be the inverse of f. Suppose that f is strictly increasing. Let $u, v \in J = f(I)$ with $u < v$. Let $a = g(u)$ and $b = g(v)$ so we have $u = f(a)$ and $v = f(b)$. Since f is strictly increasing, we must have $a < b$ (since $a = b \to f(a) = f(b) \to u = v$ and $a > b \to f(a) > f(b) \to u > v$). Thus $g(u) = a < b = g(v)$ and so g is strictly increasing. A similar argument shows that if f is strictly decreasing then so is g.

Part (4) follows from Part (3) by the Monotone Surjective Functions Theorem.

To prove Part (5), suppose that f is bijective and continuous and that f is differentiable at a with $f'(a) \neq 0$. By Part (4), we know that g is continuous at $b = f(a)$, and so as $y \to b$ in J we have $g(y) \to g(b)$ in I, and so for $x = g(y)$ we have

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

5.12 Theorem: *(Derivatives of the Basic Elementary Functions) The basic elementary functions have the following derivatives.*

(1) $(x^a)' = a\,x^{a-1}$ *where* $a \in \mathbf{R}$ *and* $x \in \mathbf{R}$ *is such that* x^{a-1} *is defined,*

(2) $(a^x)' = \ln a \cdot a^x$ *where* $a > 0$ *and* $x \in \mathbf{R}$ *and*
$(\log_a x)' = \frac{1}{\ln a} \cdot \frac{1}{x}$ *where* $0 < a \neq 1$ *and* $x > 0$*, and in particular*
$(e^x)' = e^x$ *for all* $x \in \mathbf{R}$ *and* $(\ln x)' = \frac{1}{x}$ *for all* $x > 0$*,*

(3) $(\sin x)' = \cos x$ *and* $(\cos x)' = -\sin x$ *for all* $x \in \mathbf{R}$*, and*
$(\tan x)' = \sec^2 x$ *and* $(\sec x)' = \sec x\,\tan x$ *for all* $x \in \mathbf{R}$ *with* $x \neq \frac{\pi}{2} + k\pi, k \in \mathbf{Z}$*,*
$(\cot x)' = -\csc^2 x$ *and* $(\csc x)' = -\cot x\,\csc x$ *for all* $x \in \mathbf{R}$ *with* $x \neq \pi + k\pi, k \in \mathbf{Z}$*,*

(4) $(\sin^{-1} x)' = \frac{1}{\sqrt{1-x^2}}$ *and* $(\cos^{-1} x)' = \frac{-1}{\sqrt{1-x^2}}$ *for* $|x| < 1$*,*
$(\sec^{-1} x)' = \frac{1}{x\sqrt{x^2-1}}$ *and* $(\csc^{-1} x)' = \frac{-1}{x\sqrt{x^2-1}}$ *for* $|x| > 1$*, and*
$(\tan^{-1} x)' = \frac{1}{1+x^2}$ *and* $(\cot^{-1} x)' = \frac{-1}{1+x^2}$ *for all* $x \in \mathbf{R}$*.*

Proof: First we prove Part (1) in the case that $a \in \mathbf{Q}$. When $n \in \mathbf{Z}^+$ and $f(x) = x^n$ we have

$$\frac{f(u)-f(x)}{u-x} = \frac{u^n - x^n}{u-x} = \frac{(u-x)(u^{n-1} + u^{n-2}x + u^{n-3}x^2 + \cdots + x^{n-1})}{u-x}$$
$$= u^{n-1} + u^{n-2}x + u^{n-3}x^2 + \cdots + x^{n-1} \longrightarrow n\,x^{n-1} \quad \text{as } u \to x.$$

This shows that $(x^n)' = n\,x^{n-1}$ for all $x \in \mathbf{R}$ when $n \in \mathbf{Z}^+$. By the Reciprocal Rule, for $x \neq 0$ we have

$$(x^{-n})' = \left(\frac{1}{x^n}\right)' = -\frac{(x^n)'}{(x^n)^2} = -\frac{n\,x^{n-1}}{x^{2n}} = -n\,x^{-n-1}.$$

The function $g(x) = x^{1/n}$ is the inverse of the function $f(x) = x^n$ (when n is odd, $x^{1/n}$ is defined for all $x \in \mathbf{R}$, and when n is even, $x^{1/n}$ is defined only for $x \geq 0$). Since $f'(x) = (x^n)' = n\,x^{n-1}$ we have $f'(x) = 0$ when $x = 0$. By the Inverse Function Theorem, when $x \neq 0$ we have

$$(x^{1.n})' = g'(x) = \frac{1}{f'(g(x))} = \frac{1}{n\,g(x)^{n-1}} = \frac{1}{n(x^{1/n})^{n-1}} = \frac{1}{n\,x^{1-\frac{1}{n}}} = \tfrac{1}{n}\,x^{\frac{1}{n}-1}.$$

Finally, when $n \in \mathbf{Z}^+$ and $k \in \mathbf{Z}$ with $\gcd(k,n) = 1$, by the Chain Rule we have

$$(x^{k/n})' = \big((x^{1/n})^k\big)' = k(x^{1/n})^{k-1}(x^{1/n})' = k\,x^{\frac{k-1}{n}} \cdot \tfrac{1}{n}\,x^{\frac{1-n}{n}} = \tfrac{k}{n}\,x^{\frac{k}{n}-1}.$$

We have proven Part (1) in the case that $a \in \mathbf{Q}$.

Next we shall prove Part (2). For $f(x) = a^x$ where $a > 0$, we have

$$\frac{f(x+h)-f(x)}{h} = \frac{a^{x+h} - a^x}{h} = \frac{a^x a^h - a^x}{h} = a^x \cdot \frac{a^h - 1}{h}$$

and so we have $f'(x) = a^x \left(\lim\limits_{h \to 0} \frac{a^h-1}{h}\right)$ provided that the limit exists and is finite. For $g(x) = \log_a x$, where $0 < a \neq 1$ and $x > 0$, we have

$$\frac{g(x+h)-g(x)}{h} = \frac{\log_a(x+h) - \log_a x}{h} = \frac{\log_a\left(\frac{x+h}{x}\right)}{h} = \frac{\log_a\left(1+\frac{h}{x}\right)}{x \cdot \frac{h}{x}} = \tfrac{1}{x} \cdot \log_a\left(1+\tfrac{h}{x}\right)^{x/h}$$

and so we have $g'(x) = \frac{1}{x} \cdot \log_a \left(\lim_{h \to 0} \left(1 + \frac{h}{x}\right)^{x/h} \right)$ provided the limit exists and is finite. By letting $u = \frac{h}{x}$ we see that

$$\lim_{h \to 0^+} \left(1 + \tfrac{h}{x}\right)^{x/h} = \lim_{u \to \infty} \left(1 + \tfrac{1}{u}\right)^u = e$$

as you showed in Assignment 5. By letting $u = -\frac{h}{x}$, a similar argument shows that

$$\lim_{h \to 0^-} \left(1 + \tfrac{h}{x}\right)^{x/h} = \lim_{u \to \infty} \left(1 - \tfrac{1}{u}\right)^{-u} = e.$$

Thus the derivative $g'(x)$ does exist and we have

$$(\log_a x)' = g'(x) = \tfrac{1}{x} \log_a \left(\lim_{h \to 0} \left(1 + \tfrac{h}{x}\right)^{x/h} \right) = \tfrac{1}{x} \log_a e = \tfrac{1}{x} \cdot \tfrac{\ln e}{\ln a} = \tfrac{1}{x \ln a}.$$

Since $g(x) = \log_a x$ is differentiable with $g'(x) \neq 0$ it follows from the Inverse Function Theorem that $f(x) = a^x$ is differentiable with derivative

$$(a^x)' = f'(x) = \frac{1}{g'(f(x))} = \frac{1}{\frac{1}{f(x)\ln a}} = \ln a \cdot f(x) = \ln a \cdot a^x.$$

Now we return to the proof of Part (1), in the case that $a \notin \mathbf{Q}$. When $a > 0$ we have $a^x = e^{x \ln a}$ for all $x > 0$ and so by the Chain Rule

$$(x^a)' = \left(e^{a \ln x}\right)' = e^{a \ln x}(a \ln x)' = x^a \cdot \tfrac{a}{x} = a\, x^{a-1},$$

I may finish the proof later.

5.13 Definition: Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f : A \to F$ and let $a \in A$. We say that f has a **local maximum** value at a when

$$\exists \delta > 0\ \forall x \in A\ \Big(|x - a| \leq \delta \to f(x) \leq f(a) \Big).$$

Similarly, we say that f has a **local minimum** value at a when

$$\exists \delta > 0\ \forall x \in A\ \Big(|x - a| \leq \delta \to f(x) \geq f(a) \Big).$$

5.14 Theorem: *(Fermat's Theorem) Let F be a subfield of $\mathbf{R}$, let $A \subseteq F$, let $f : A \to F$. Suppose that $a \in A$ is a limit point of A, both from above and from below. Suppose that f is differentiable at a and that f has a local maximum or minimum value at a. Then $f'(a) = 0$.*

Proof: We suppose that f has a local maximum value at a (the case that f has a local minimum value at a is similar). Choose $\delta > 0$ so that $|x - a| \leq \delta \to f(x) \leq f(a)$. For $x \in A$ with $a < x < a + \delta$, since $x > a$ and $f(x) \geq f(a)$ we have $\frac{f(x) - f(a)}{x - a} \geq 0$, and so

$$f'(a) = \lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} \geq 0$$

by the Comparison Theorem. Similarly, for $x \in A$ with $a - \delta \leq x < a$, since $x < a$ and $f(x) \geq f(a)$ we have $\frac{f(x) - f(a)}{x - a} \leq 0$, and so

$$f'(a) = \lim_{x \to a^-} \frac{f(x) - f(a)}{x - a} \leq 0.$$

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

(1) (Rolle's Theorem) If $f : [a, b] \to \mathbf{R}$ 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 $f : [a, b] \to \mathbf{R}$ 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 $f, g : [a, b] \to \mathbf{R}$ 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 $f : [a, b] \to \mathbf{R}$ 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.

Now we use Rolle's Theorem to prove the Mean Value Theorem. Suppose that $f : [a, b] \to \mathbf{R}$ 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 $f'(c) = 0$, 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 that $f, g : [a, b] \to \mathbf{R}$ 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.

5.16 Corollary: *Let $a, b \in \mathbf{R}$ with $a < b$. Let $f : [a, b] \to \mathbf{R}$. Suppose that f is differentiable in (a, b) and continuous at a and b.*

(1) If $f'(x) \geq 0$ for all $x \in (a, b)$ then f is increasing on $[a, b]$.
(2) If $f'(x) > 0$ for all $x \in (a, b)$ then f is strictly increasing on $[a, b]$.
(3) If $f'(x) \leq 0$ for all $x \in (a, b)$ then f is decreasing on $[a, b]$.
(4) If $f'(x) < 0$ for all $x \in (a, b)$ then f is strictly decreasing on $[a, b]$.
(5) if $f'(x) = 0$ for all $x \in (a, b)$ then f is constant on $[a, b]$.
(6) If $g : [a, b] \to \mathbf{R}$ is continuous at a and b and differentiable in (a, b) with $g'(x) = f'(x)$ for all $x \in (a, b)$, then for some $c \in \mathbf{R}$ we have $g(x) = f(x) + c$ for all $x \in (a, b)$.

Proof: We prove Part (1) (the proofs of the other parts are similar. Suppose that $f'(x) \geq 0$ for all $x \in (a, b)$. Let $a \leq x < y \leq b$. Choose $c \in (x, y)$ so that $f'(c) = \frac{f(y)-f(x)}{y-x}$. Then $f(y) - f(x) = f'(c)(y - x) \geq 0$ and so $f(y) \geq f(x)$. Thus f is increasing on $[a, b]$.

5.17 Corollary: *(The Second Derivative Test) Let I be an interval in $\mathbf{R}$, let $f : I \to \mathbf{R}$ and let $a \in I$. Suppose that f is differentiable in I with $f'(a) = 0$.*
(1) If $f''(a) > 0$ then f has a local minimum at a.
(2) If $f''(a) < 0$ then f has a local maximum at a.

Proof: The proof is left as an exercise.

5.18 Theorem: *(l'Hôpital's Rule) Let I be a non degenerate interval in $\mathbf{R}$. Let $a \in I$, or let a be an endpoint of I. Let $f, g : I \setminus \{a\} \to \mathbf{R}$. Suppose that f and g are differentiable in $I \setminus \{a\}$ with $g'(x) \neq 0$ for all $x \in I \setminus \{a\}$. 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 $\lim\limits_{x \to a} \dfrac{f'(x)}{g'(x)} = u \in \hat{\mathbf{R}}$. 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 $x \to -\infty$.

Proof: We give the proof for $x \to a^+$ (assuming that a is a limit point of I from the right) and that $u \in \mathbf{R}$. Suppose first that $\lim\limits_{x \to a^+} f(x) = 0 = \lim\limits_{x \to a^+} g(x)$. Choose $b \in I$ with $a < b$. Extend the maps f and g to obtain maps $f, g : [a, b] \to \mathbf{R}$ by defining $f(a) = 0 = g(b)$. 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 $\langle x_k \rangle$ 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 $\langle x_k \rangle$ in $(a, b]$ with $x_k \to a$ we have $\frac{f(x_k)}{g(x_k)} \to u$, and it follows that $\lim\limits_{x \to a^+} \dfrac{f(x)}{g(x)} = u$ by the Sequential Characterization of Limits.

Now suppose that $\lim\limits_{x \to a^+} g(x) = \infty$. Since $\lim\limits_{x \to a^+} g(x) = \infty$ we can choose $b \in I$ with $b > a$ so that $g(x) > 0$ for all $x \in (a, b]$. Let $\langle x_k \rangle$ 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 $|x - a| \leq \delta \to \left|\frac{f'(x)}{g'(x)} - u\right| \leq \frac{\epsilon}{3}$. Since $x_k \to a$ we can choose $m \in \mathbf{Z}^+$ so $k \geq m \to |x_k - a| \leq \delta$. 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. Introduction to the Foundations of Mathematics

1.1 Remark: A little over 100 years ago, it was found that some mathematical proofs contained paradoxes, and these paradoxes could be used to prove statements that were known to be false. One well known paradox, outside of the realm of mathematics, is the statement

"This statement is false".

The above statement is true if and only if it is false. It is one form of a paradox known as the **liar's paradox**. After examining some lengthy and convoluted mathematical proofs which contained paradoxes, Bertrand Russel came up with the following mathematical paradox, which is somewhat similar to the liar's paradox:

Let X be the set of all sets, and let $S = \big\{A \in X \big| A \notin A\big\}$.
Note for example that $\mathbf{Z} \notin \mathbf{Z}$ so $\mathbf{Z} \in S$, and $X \in X$ so $X \notin S$.
Then we have $S \in S$ if and only if $S \notin S$.

This paradox is known as **Russel's paradox**. With Russel's paradox, it was possible to construct a proof by contradiction, which followed all the accepted rules of mathematical proof, of any statement whatsoever. Mathematicians realized that they would need to modify the accepted framework of mathematics in order to ensure that mathematical paradoxes could no longer arise. They were led to consider the following three questions.

1. Exactly what is an allowable mathematical object?
2. Exactly what is an allowable mathematical statement?
3. Exactly what is an allowable mathematical proof?

Eventually, after a great deal of work by many mathematicians, a consensus was reached as to the answers to these three questions. Roughly, the answers are as follows. Every mathematical object is a mathematical **set** (this includes objects that we would not normally consider to be sets, such as the integer 1), and a mathematical set can be constructed using certain specific rules, known as the **ZFC axioms** of set theory. Every mathematical statement can be expressed as a so-called **formula** in a certain specific formal symbolic language, which uses symbols rather than words from a spoken language, such as English. Every mathematical proof is a finite list of ordered pairs $(\mathcal{S}_n, F_n)$ (which we think of as *proven theorems*), where each S_n is a finite set of formulas (called the *premises*) and each F_n is a single formula (called the *conclusion*), such that each pair $(\mathcal{S}_n, F_n)$ can be obtained from previous pairs $(\mathcal{S}_i, F_i)$ with $i < n$, using certain specific proof rules.

In the remainder of this appendix, we provide a fairly detailed answer to the first two of the above three questions, beginning with the second question.

1.2 Definition: We allow ourselves to use only symbols from the following **symbol set**

$\neg$	not
$\wedge$	and
$\vee$	or
$\rightarrow$	implies
$\leftrightarrow$	if and only if
$=$	equals
$\in$	is an element of
$\forall$	for all
$\exists$	there exists
$($, $)$	parenthises

along with some variable symbols such as $x, y, z, u, v, w, \cdots$ or $x_1, x_2, x_3, \cdots$.

1.3 Definition: A **formula** (in the formal symbolic language of first order set theory) is a non-empty finite string of symbols, from the above list, which can be obtained using finitely many applications of the following three rules.

1. If x and y are variable symbols, then each of the following strings are formulas.

$$x = y \ , \quad x \in y$$

2. If F and G are formulas then each of the following strings are formulas.

$$\neg F \ , \quad (F \wedge G) \ , \quad (F \vee G) \ , \quad (F \rightarrow G) \ , \quad (F \leftrightarrow G)$$

3. If x is a variable symbol and F is a formula then each of the following is a formula.

$$\forall x \, F \ , \quad \exists x \, F$$

1.4 Definition: Let x be a variable symbol and let F be a formula. For each occurrence of the symbol x, which does not immediately follow a quantifier, in the formula F, we define whether the occurrence of x is **free** or **bound** inductively as follows.

1. If F is a formula of one of the forms $y = z$ or $y \in z$, where y and z are variable symbols (possibly equal to x), then every occurrence of x in F is free, and no occurrence is bound.

2. If F is a formula of one of the forms $\neg G$, $(G \wedge H)$, $(G \vee H)$, $(G \rightarrow H)$ or $(G \leftrightarrow H)$, where G and H are formulas, then each occurrence of the symbol x is either an occurrence in the formula G or an occurrence in the formula H, and each free (respectively, bound) occurrence of x in G remains free (respectivly, bound) in F, and similarly for each free (or bound) occurrence of x in H.

3. If F is a formula of one of the forms $\forall y \, G$ or $\exists y \, G$, where G is a formula and y is a variable symbol (possibly equal to x), then if y is different than x then each free (or bound) occurrence of x in G remains free (or bound) in the formula F, and if y is equal to x then every free occurrence of x in G becomes bound in the formula F, and every bound occurrence of x in G remains bound in the formula F.

1.5 Definition: When a quantifier symbol occurs in a given formula F, and is followed by the variable symbol x and then by the formula G, any free occurrence of x in G will become bound in the given formula F (by an application of part 3 of the above definition), and we shall say that that occurrence of x **is bound by** (that occurrence of) the quantifier symbol, or that (that occurrence of) the quantifier symbol **binds** that occurrence of x.

1.6 Definition: A **free variable** in a formula F is any variable symbol that has at least one free occurence in F. A formula F with no free variables is called a **statement**. When the free variables in F all lie in the set $\{x_1, x_2, \cdots, x_n\}$, we shall write F as $F(x_1, \cdots, x_n)$ and we shall say that F is a **statement about** the variables $x_1, x_2, \cdots, x_n$.

1.7 Example: In the following formula, determine which occurrences of the variable symbols are free and which are bound, and for each bound occurrence, indicate which quantifier binds it.

$$\forall x\, \exists y \big(\forall z (x \in y \rightarrow \exists y\ y = z) \wedge \forall x (\exists z\ z = u \vee z \in x)\big)$$

Solution: We indicate the free and bound occurrences and their binding quantifiers by placing integral labels under the relevant symbols: the free variables are given the label 0, each quantifier is given its own non-zero label, and each bound variable is given the same label as its binding quantifier:

$$\begin{array}{l} \forall x\, \exists y \big(\forall z (x \in y \rightarrow \exists y\ y = z) \wedge \forall x (\exists z\ z = u \vee z \in x)\big) \\ 1\ \ 2\ \ \ 3\ \ 1\ \ \ 2\ \ \ \ 4\ \ 4\ \ \ 3\ \ \ \ 5\ \ 6\ \ 6\ \ \ 0\ \ \ 0\ \ \ 5 \end{array}$$

We remark that the free variables in this formula are z and u, so we say that it is a statement about z and u.

1.8 Example: Express that statement $x = \big\{y, \{z\}\big\}$ as a formal symbolic formula.

Solution: We can express the given statement in each of the following ways.

$$\begin{gathered} x = \big\{y, \{z\}\big\} \\ \forall u \left(u \in x \leftrightarrow u \in \big\{y, \{z\}\big\}\right) \\ \forall u \big(u \in x \leftrightarrow \big(u = y \vee u = \{z\}\big)\big) \\ \forall u \big(u \in x \leftrightarrow \big(u = y \vee \forall v (v \in u \leftrightarrow v = z)\big)\big) \end{gathered}$$

The last expression is a formula.

1.9 Definition: When $F(x)$ is a statement about x we sometimes write $F(y)$ as a short form for the formula $\forall x \big(x = y \rightarrow F(x)\big)$, and we sometimes write

$$\exists! y\ F(y)$$

which we read as "there exists a unique y such that $F(y)$", as a short form for the formula

$$\exists y \big(F(y) \wedge \forall z (F(z) \rightarrow z = y\big)$$

which is short, in turn, for the formula

$$\exists y \Big(\forall x (x = y \rightarrow F(x)) \wedge \forall z \big(\forall x (x = z \rightarrow F(x)) \rightarrow z = y\big)\Big).$$

The ZFC Axioms of Set Theory

1.10 Remark: Every mathematical **set** can be constructed using specific rules, which are known as the **ZFC axioms** of set theory, or the Zermelo-Fraenkel axioms of set theory, with the axiom of choice. We begin by listing the ZFC axioms, stating them informally.

Empty Set Axiom: There exists a set $\emptyset$ with no elements.

Extension Axiom: Two sets are equal if and only if they have the same elements.

Separation Axiom: If u is a set and $F(x)$ is a statement about x, $\{x \in u \,|\, F(x)\}$ is a set.

Pair Axiom: If u and v are sets then $\{u, v\}$ is a set.

Union Axiom: If u is a set then $\bigcup u = \bigcup\limits_{v \in u} v$ is a set.

Power Set Axiom: If u is a set then $\mathcal{P}(u) = \{v \,|\, v \subseteq u\}$ is a set.

Axiom of Infinity: If we define the natural numbers to be the sets $0 = \emptyset$, $1 = \{0\}$, $2 = \{0, 1\}$, $3 = \{0, 1, 2\}$ and so on, then $\mathbf{N} = \{0, 1, 2, 3, \cdots\}$ is a set.

Replacement Axiom: If u is a set and $F(x, y)$ is a statement about x and y with the property that $\forall x \, \exists! y \; F(x, y)$ then $\{y \,|\, \exists x \in u \; F(x, y)\}$ is a set.

Axiom of Choice: Given a set u of non-empty pairwise disjoint sets, there exists a set which contains exactly one element from each of the sets in u.

We now proceed to state each of the ZFC axioms formally (as a symbolic formula) and give some indication as to how these axioms can be used as a rigorous framework for essentially all of mathematics.

1.11 Definition: The **Empty Set Axiom** is the formula

$$\exists u \, \forall x \; \neg x \in u \, .$$

1.12 Definition: The **Extension Axiom** is the formula

$$\forall u \, \forall v \, \big(u = v \leftrightarrow \forall x (x \in u \leftrightarrow x \in v) \big) \, .$$

1.13 Theorem: *The empty set is unique.*

Proof: Suppose that u and v are both empty. Let x be arbitrary. Since u is empty, we have $\neg x \in u$ and hence $x \in u \to x \in v$. Similarly, since v is empty, we have $\neg x \in v$ and hence $x \in v \to x \in u$. Since $x \in u \to x \in v$ and $x \in v \to x \in u$, we have $x \in u \leftrightarrow x \in v$. Since x was arbitrary, we have $\forall x \, (x \in u \leftrightarrow x \in v)$. By the Axiom of Extension, $u = v$.

1.14 Definition: We denote the unique empty set by $\emptyset$.

1.15 Remark: In a formal and rigorous treatment of the foundations of mathematics, we would need to decide at this point how to interpret the use of the symbol $\emptyset$. One approach is to add the symbol $\emptyset$ to our list of symbols, modify our definition of a formula to allow the use of the new symbol $\emptyset$, and add the axiom $\forall x \; \neg x \in \emptyset$ to our list of axioms. Another option is to interpret the use of the symbol as a shorthand notation for an expression which can be expressed formally using the existing symbols, so that for example the expression $u = \emptyset$ would be shorthand for the formula $\forall x \; \neg x \in u$.

1.16 Definition: Given sets u and v, we say that u is a **subset** of v, and we write $u \subseteq v$, when every element of u also lies in v, that is when $\forall x\,(x \in u \to x \in v)$.

1.17 Definition: For any statement $F(x)$ about x, the following formula is an axiom.

$$\forall u\,\exists v\,\forall x\big(x \in v \leftrightarrow (x \in u \land F(x))\big)$$

More generally, for any statement $F(x, u_1, u_2, \cdots, u_n)$ about $x, u_1, u_2, \cdots, u_n$, where $n \geq 0$, the following formula is an axiom.

$$\forall u \forall u_1 \cdots \forall u_n \exists v\,\forall x\big(x \in v \leftrightarrow \big(x \in u \land F(x, u_1, \cdots, u_n)\big)\big)$$

Any axiom of this form is called an **Axiom of Separation**.

1.18 Notation: Given sets $u, u_1, \cdots, u_n$ and given a formula $F(x, u_1, \cdots, u_n)$ about $x, u_1, \cdots, u_n$, by the appropriate Axiom of Separation, there exists a set v with the property that $\forall x\big(x \in v \leftrightarrow \big(x \in u \land F(x, u_1, \cdots, u_n)\big)\big)$, and by the Extension Axiom, this set v is unique, and we denote it by

$$\big\{x \in u \,\big|\, F(x, u_1, \cdots, u_n)\big\}\,.$$

1.19 Note: It is important to realize that a Separation Axiom only allows us to construct a subset of a given set u, so for example we cannot use a Separation Axiom to show that the collection $S = \{x \big| \neg x \in x\}$, which is used to formulate Russel's paradox, is a set.

1.20 Definition: The **Pair Axiom** is the formula

$$\forall u\,\forall v\,\exists w\,\forall x\big(x \in w \leftrightarrow (x = u \lor x = v)\big)\,.$$

1.21 Notation: Given sets u and v, by the Pair Axiom there exists a set w with the property that $\forall x\big(x \in w \leftrightarrow (x = u \lor x = v)\big)$, and by the Extension Axiom, this set w is unique, and we denote it by

$$\{u, v\}$$

1.22 Example: With this axiom, we can construct some non-empty sets. For example, taking $u = v = \emptyset$ gives the set $\{\emptyset, \emptyset\} = \{\emptyset\}$ (note that $\{\emptyset\} \neq \emptyset$ by the Extension Axiom, since $\emptyset \in \{\emptyset\}$ but $\emptyset \notin \emptyset$). Then taking $u = \emptyset$ and $v = \{\emptyset\}$ gives the set $\big\{\emptyset, \{\emptyset\}\big\}$.

1.23 Definition: The **Union Axiom** is the formula

$$\forall u\,\exists w\,\forall x\big(x \in w \leftrightarrow \exists v\,(v \in u \land x \in v)\big)\,.$$

1.24 Definition: Given a set u, by the Union Axiom there exists a set w with the property that $\forall x\big(x \in w \leftrightarrow \exists v\,(v \in u \land x \in v)\big)$, and by the Extension Axiom this set w is unique. We call the set w the **union** of the elements in u, and we denote it by

$$\bigcup u = \bigcup_{v \in u} v\,.$$

Given two sets u and v, we define the **union** of u and v to be the set

$$u \cup v = \bigcup \{u, v\}\,.$$

Given three sets u, v and w, note that $\{z\} = \{z, z\}$ is a set and so $\{x, y, z\} = \{x, y\} \cup \{z\}$ is also a set. More generally, if $u_1, u_2, \cdots, u_n$ are sets then $\{u_1, u_2, \cdots, u_n\}$ is a set and we define the **union** of the sets $u_1, \cdots, u_n$ to be

$$u_1 \cup u_2 \cup \cdots \cup u_n = \bigcup_{k=1}^{n} u_k = \bigcup \{u_1, u_2, \cdots, u_n\}\,.$$

1.25 Definition: Given a set u, we define the **intersection** of the elements in u to be the set

$$\bigcap u = \left\{ x \in \bigcup u \,\middle|\, \forall v (v \in u \to x \in v) \right\}$$

Given two sets u and v, we define the **intersection** of u and v to be the set

$$u \cap v = \bigcap \{u, v\},$$

and more generally, given sets $u_1, u_2, \cdots, u_n$, we define the **intersection** of $u_1, u_2, \cdots, u_n$ to be the set

$$u_1 \cap u_2 \cap \cdots \cap u_n = \bigcap_{k=1}^{n} u_k = \bigcap \{u_1, u_2, \cdots, u_n\}.$$

1.26 Definition: The **Power Set Axiom** is the formula

$$\forall u\, \exists w\, \forall v\, (v \in w \leftrightarrow v \subseteq u).$$

1.27 Definition: Given a set u, the set w with the property that $\forall v (v \in w \leftrightarrow v \subseteq u)$ (which exists by the Power Set Axiom and is unique by the Extension Axiom) is called the **power set** of u and is denoted by $\mathcal{P}(u)$, so we have

$$\mathcal{P}(u) = \{ v | v \subseteq u \}.$$

1.28 Example: Find the power set of the set $\{\emptyset, \{\emptyset\}\}$.

Solution: We have

$$\mathcal{P}\Big(\{\emptyset, \{\emptyset\}\}\Big) = \Big\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\Big\}.$$

1.29 Definition: Given two sets x and y, we define the **ordered pair** (x, y) to be the set

$$(x, y) = \big\{ \{x\}, \{x, y\} \big\}.$$

Given two sets u and v, note that if $x \in u$ and $y \in v$ then we have $\{x\} \in \mathcal{P}(u \cup v)$ and $\{x, y\} \in \mathcal{P}(u \cup v)$ and so $(x, y) = \big\{\{x\}, \{x, y\}\big\} \in \mathcal{P}\big(\mathcal{P}(u \cup v)\big)$. We define the **product** $u \times v$ to be the set

$$u \times v = \big\{ (x, y) \big| x \in u \wedge y \in v \big\},$$

that is

$$u \times v = \Big\{ z \in \mathcal{P}\big(\mathcal{P}(u \cup v)\big) \Big| \, \exists x\, \exists y \big((x \in u \wedge y \in v) \wedge z = (x, y)\big) \Big\}.$$

1.30 Exercise: Find $\bigcup \Big(\{\emptyset\} \times \big\{\{\emptyset\}, \{\emptyset, \{\emptyset\}\}\big\} \Big)$.

1.31 Definition: We define

$$0 = \emptyset\ ,\ 1 = \{0\} = 0 \cup \{0\}\ ,\ 2 = \{0, 1\} = 1 \cup \{1\}\ ,\ 3 = \{0, 1, 2\} = 2 \cup \{2\},$$

and so on. For a set x, we define the **successor** of x to be the set

$$x + 1 = x \cup \{x\}.$$

A set u is called **inductive** when it has the property that

$$\big(0 \in u \wedge \forall x\, (x \in u \to x + 1 \in u)\big).$$

1.32 Definition: The **Axiom of Infinity** is the formula

$$\exists u \big(0 \in u \wedge \forall x\, (x \in u \to x + 1 \in u)\big),$$

so the Axiom of Infinity states that there exists an inductive set.

1.33 Theorem: *There exists a unique set w of the form*

$$w = \big\{x \,\big|\, x \in v \text{ for every inductive set } v\big\}\,.$$

Moreover, this set w is an inductive set.

Proof: By the axiom of infinity, there exists an inductive set, say u. Let w be the set

$$\begin{aligned} w &= \big\{x \in u \,\big|\, x \in v \text{ for every inductive set } v\big\} \\ &= \big\{x \in u \,\big|\, \forall v \,\big(\big(0 \in v \wedge \forall y\,(y \in v \to y+1 \in v)\big) \to x \in v\big)\big\}\,. \end{aligned}$$

We claim that this set w does not depend on the choice of u. To prove this, let u_1 and u_2 be two inductive sets and let

$$\begin{aligned} w_1 &= \big\{x \in u_1 \,\big|\, x \in v \text{ for every inductive set } v\big\} \\ w_2 &= \big\{x \in u_2 \,\big|\, x \in v \text{ for every inductive set } v\big\}\,. \end{aligned}$$

Then for any set x we have

$$\begin{aligned} x \in w_1 &\iff x \in u_1 \text{ and } x \in v \text{ for every inductive set } v \\ &\iff x \in v \text{ for every inductive set } v \text{ (since } u_1 \text{ is inductive)} \\ &\iff x \in u_2 \text{ and } x \in v \text{ for every inductive set } v \text{ (since } u_2 \text{ is inductive)} \\ &\iff x \in w_2\,. \end{aligned}$$

Thus $w_1 = w_2$, showing that w is unique. We leave it as an exercise to show that w is inductive.

1.34 Definition: The unique set w in the above theorem is called the set of **natural numbers**, and we denote it by $\mathbf{N}$. We write

$$\begin{aligned} \mathbf{N} &= \big\{x \,\big|\, x \in v \text{ for every inductive set } v\big\} \\ &= \{0, 1, 2, 3 \cdots\}\,. \end{aligned}$$

For $x, y \in \mathbf{N}$, we write $x < y$ when $x \in y$ and we write $x \leq y$ when $x < y$ or $x = y$.

1.35 Notation: For a formula F, we write $\forall x{\in}u\; F$ as a shorthand notation for the formula $\forall x\,(x{\in}u \to F)$. Similarly, we write $\exists x{\in}u\; F$ as a shorthand notation for $\exists x\,(x{\in}u \wedge F)$.

1.36 Theorem: *(Principle of Induction) Let $F(x)$ be a statement about x. Suppose that*
(1) $F(0)$, and
(2) $\forall x{\in}\mathbf{N}\,\big(F(x) \to F(x+1)\big)$.
Then $\forall x{\in}\mathbf{N}\; F(x)$.

Proof: Let $u = \big\{x \in \mathbf{N} \big| F(x)\big\}$. By (1) we have $0 \in u$. Let $x \in u$. Then $x \in \mathbf{N}$ and $F(x)$. Since $x \in \mathbf{N}$ we have $x+1 \in \mathbf{N}$ (since $\mathbf{N}$ is inductive). Since $x \in \mathbf{N}$ and $F(x)$ we have $F(x+1)$ by (2). Since $x+1 \in \mathbf{N}$ and $F(x+1)$, we have $x+1 \in u$ (by the definition of u). We have shown that $0 \in u$ and that $\forall x\,(x \in u \to x+1 \in u)$, so u is inductive. Since u is inductive, we have $\mathbf{N} \subseteq u$ (by the definition of $\mathbf{N}$). Thus $x \in \mathbf{N} \Longrightarrow x \in u \Longrightarrow F(x)$.

1.37 Remark: In the above theorem, the expression $F(0)$ is short for $\forall x\,\big(x = 0 \to F(x)\big)$ which in turn is short for $\forall x\,\big(\forall y\, \neg y \in x \to F(x)\big)$. Similarly, $F(x+1)$ is short for the formula $\forall y\,\big(y = x+1 \to F(y)\big)$, where $F(y)$ is short for $\forall x\big(x = y \to F(x)\big)$.

1.38 Definition: Given a statement $F(x,y)$ about x and y, the following formula is an axiom:

$$\forall u\Big(\forall x\,\exists! y\ F(x,y) \to \exists w\,\forall y\,\big(y\in w \leftrightarrow \exists x\!\in\! u\ F(x,y)\big)\Big)\,,$$

where $\exists! y F(x,y)$ is short for $\exists y\big(F(x,y)\wedge\forall z\,\big(F(x,z)\to z=y\big)\big)$ with $F(x,z)$ short for the formula $\forall y\,\big(y=z\to F(x,y)\big)$. More generally, given a statement $F(x,y,u_1,\cdots,u_n)$ about $x,y,u_1,\cdots,u_n$ with $n\geq 0$, the following formula is an axiom:

$$\forall u\,\forall u_1\cdots\forall u_n\Big(\forall x\,\exists! y\ F(x,y,u_1,\cdots,u_n)\to\exists w\,\forall y\,\big(y\in w\leftrightarrow\exists x\!\in\! u\ F(x,y,u_1,\cdots,u_n)\big)\Big)\,.$$

An axiom of this form is called a **Replacement Axiom**.

1.39 Notation: Given sets $u,u_1,\cdots,u_n$ and given a statement $F(x,y,u_1,\cdots,u_n)$ about $x,y,u_1,\cdots,u_n$ with the property that $\forall x\,\exists! y\ F(x,y,u_1,\cdots,u_n)$, for each set x we let $y=f(x)$ denote the unique set for which $F(x,y,u_1,\cdots,u_n)$ holds, and then we denote the unique set w, whose existence is stipulated by the above Replacement Axiom, by

$$\big\{f(x)\big|x\in u\big\}\,.$$

1.40 Example: If u is a set then the collection

$$\big\{\mathcal{P}(x)\big|x\in u\big\}$$

is also a set, by the Replacement Axiom taking $F(x,y)$ to be the formula $y=\mathcal{P}(x)$.

1.41 Definition: The **Axiom of Choice** is the formula given by

$$\forall u\Big(\big(\neg\phi\in u\wedge\forall x\!\in\! u\ \forall y\!\in\! u\,(\neg x=y\to x\cap y=\emptyset)\big)\to\exists w\,\forall v\!\in\! u\ \exists!\,x\!\in\! v\ x\in w\Big)$$

1.42 Remark: We have now stated each of the ZFC axioms formally. Up until now, we have used lower-case letters to denote all sets (and all elements of sets, which are also sets). From now on, we shall often use upper-case letters to denote sets, as is more customary.

1.43 Definition: A **binary relation** R on a set X is a subset $R \subseteq X \times X$. More generally, a **binary relation** is any set R whose elements are ordered pais. For a binary relation R, we usually write xRy instead of $(x, y) \in R$.

1.44 Definition: Let R and S be binary relations. The **domain** of R is

$$\text{Domain}(R) = \{x \mid \exists y \; xRy\}$$

and the **range** of R is

$$\text{Range}(R) = \{x \mid \exists y \; xRy\} .$$

For any set A, the **image** of A under R is

$$R(A) = \{y \mid \exists x \in A \; xRy\}$$

and the **inverse image** of A under R is

$$R^{-1}(A) = \{x \mid \exists y \in A \; xRy\} .$$

The **inverse** of R is

$$R^{-1} = \{(y, x) \big| (x, y) \in R\}$$

and the composite S **composed with** R is

$$S \circ R = \{(x, z) \big| \exists y \; xRy \wedge ySz\} .$$

1.45 Theorem: *Let A be a set and let R and S be binary relations. Then*
(1) Domain(R), Range(R), $R(A)$ *and* $R^{-1}(A)$ *are sets, and*
(2) R^{-1} *and* $S \circ R$ *are binary relations.*

Proof: The proof is left as an exercise.

1.46 Definition: An **equivalence relation** on a set X is a binary relation R on X such that

(1) R is **reflexive**, that is $\forall x \in X \; xRx$,
(2) R is **symmetric**, that is $\forall x, y \in X \; (xRy \to yRx)$, and
(3) R is **transitive**, that is $\forall x, y, z \in X \; \big((xRy \wedge yRz) \to xRz\big)$.

1.47 Definition: Let R be an equivalence relation on the set X. For $a \in X$, the **equivalence class** of a modulo R is the set

$$[a]_R = \{x \in X \big| xRa\} .$$

1.48 Definition: A **partition** of a set X is a set S of non-empty pairwise disjoint sets whose union is X, that is a set S such that

(1) $\forall X, Y \in S \; \big(X \neq Y \to X \cap Y = \emptyset\big)$, and
(2) $\bigcup S = X$.

1.49 Theorem: *Given a set X, we have the following correspondence between equivalence relations on X and partitions of X.*

(1) Given an equivalence relation R on X, the set of all equivalence classes

$$S_R = \big\{[a]_R \,\big|\, a \in X\big\}$$

is a partition of X.
(2) Given a partition S of X, the relation R_S on X defined by

$$R_S = \Big\{(x,y) \in X \times X \Big| \exists A \in S\, (x \in A \wedge y \in A)\Big\}$$

is an equivalence relation on X.
(3) Given an equivalence relation R on X we have $R_{S_R} = R$, and given a partition S of X we have $S_{R_S} = S$.

Proof: The proof is left as an exercise.

1.50 Notation: Given an equivalence relation R on X, the set of all equivalence classes, which we denoted by S_R in the above theorem, is usually denoted by X/R, so

$$X/R = \big\{[a]_R \,\big|\, a \in X\big\}\,.$$

1.51 Definition: Let R be an equivalence relation. A **set of representatives** for R is a subset of X which contains exactly one element from each equivalence class in X/R.

1.52 Remark: Notice that the Axiom of Choice is equivalent to the statement that every equivalence relation has a set of representatives.

1.53 Definition: Given sets X and Y, a **function** from X to Y is a binary relation $f \subseteq X \times Y$ with the property that

$$\forall x \in X\ \exists!\, y \in Y\ (x,y) \in f\,.$$

More generally, a **function** is a binary relation with the property that

$$\forall x \in \text{Domain}(f)\ \exists!\, y\ (x,y) \in f\,.$$

For a function f, we usually write $y = f(x)$ instead of xfy. It is customary to use the notation $f : X \to Y$ when $X = \text{Domain}(f)$ and Y is any set with Range(f) $\subseteq Y$.

1.54 Definition: Let $f : X \to Y$. The function f is called **one-to-one** (or **injective**) when

$$\forall y \in Y\ \exists \text{ at most one } x \in X\ y = f(x)$$

and f is called **onto** (or **surjective**) when

$$\forall y \in Y\ \exists \text{ at least one } x \in X\ y = f(x)\,.$$

1.55 Definition: Let $f : X \to Y$. Let I_X and I_Y denote the identity functions on X and Y respectively (that is $I_X(x) = x$ for all $x \in X$ and $I_Y(y) = y$ for all $y \in Y$). A **left inverse** of f is a function $g : Y \to X$ such that $g \circ f = I_X$. A **right inverse** of f is a function $H : Y \to X$ such that $f \circ H = I_Y$. Note that if f has a left inverse g and a right inverse H, then we have $g = g \circ I_Y = g \circ f \circ H = I_X \circ H = H$. In this case we say that g is the (unique two-sided) **inverse** of f.

1.56 Theorem: *Let $f : X \to Y$. Then*

(1) f is one-to-one if and only if f has a left inverse.
(2) f is onto if and only if f has a right inverse.
(3) f is one-to-one and onto if and only if f has a (two-sided) inverse.

Proof: The proof is left as an exercise. We remark that the Axiom of Choice is needed.

1.57 Definition: A function $f : X \to Y$ is called **invertible** (or **bijective**) when it is one-to-one and onto, or equivalently, when it has a (unique two-sided) inverse.

1.58 Remark: The Axiom of Choice is equivalent to the statement that for every set S, there exists a function $f : S \to \bigcup S$ with the property that $\forall X \in S \big(X \neq \emptyset \to f(X) \in X \big)$. Such a function f is called a **choice function** for the set S.

1.59 Theorem: *(The Recursion Theorem)*

(1) Let A be a set, let $a \in A$, and let $g : A \times \mathbf{N} \to A$. Then there exists a unique function $f : \mathbf{N} \to A$ such that

$$f(0) = a \text{ and } f(n+1) = g\big(f(n), n\big) \text{ for all } n \in \mathbf{N}\,.$$

(2) Let A and B be sets, let $g : A \to B$, and let $h : A \times B \times \mathbf{N} \to B$. Then there exists a unique function $f : A \times \mathbf{N} \to B$ such that for all $a \in A$ we have

$$f(a, 0) = g(a) \text{ and } f(a, n+1) = h\big(a, f(a, n), n\big) \text{ for all } n \in \mathbf{N}\,.$$

Proof: To prove part (1), note first that for each $n \in \mathbf{N}$ we can construct a (unique) function $f_n : \{0, 1, \cdots, n\} \to A$ such that $f(0) = a$ and $f_n(k+1) = g\big(f_n(k), k\big)$ for all k with $0 \leq k < n$ (that the functions f_n exist and are unique can be proven by induction). Notice that since $\{0, 1, \cdots, n\} = n+1$, we have $f_n : (n+1) \to A$, so $f_n \subseteq (n+1) \times A \subseteq \mathbf{N} \times A$, and so all of the functions f_n are subsets of $\mathbf{N} \times A$. We can combine all these functions into a single function $f : \mathbf{N} \to A$ as follows. First we let

$$F = \Big\{ f \subseteq \mathbf{N} \times A \,\Big|\, \exists n \in \mathbf{N} \Big(f : (n+1) \to A\,,\ f(0) = a\,,\ \forall k \in (n+1)\ f(k+1) = g\big(f(k), k\big) \Big) \Big\}\,,$$

and then we let

$$f = \bigcup F\,.$$

We leave it as an exercise to prove that indeed f is a function which satisfies the conditions of the theorem.

We can prove part (2) in a similar manner. First we let

$$F = \Big\{ f \subseteq A \times \mathbf{N} \times B \,\Big|\, \exists n \in \mathbf{N} \Big(f : A \times (n+1) \to B \text{ and } \forall a \in A \Big(f(a, 0) = g(a) \wedge \forall k \in (n+1)\ f(a, k+1) = h\big(a, f(a, k), k\big) \Big) \Big\}\,,$$

then we let $f = \bigcup F$.

1.60 Definition: By part (2) of the Recursion Theorem, there is a unique function $s : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$ such that for all $a, b \in \mathbf{N}$ we have

$$s(a,0) = a\,, \quad s(a,b+1) = s(a,b) + 1\,.$$

We call $s(a,b)$ the **sum** of a and b in $\mathbf{N}$, and we write it as

$$a + b = s(a,b)\,.$$

Also, there is a unique function $p : \mathbf{N} \times \mathbf{N} \to \mathbf{N}$ such that for all $a, b \in \mathbf{N}$ we have

$$p(a,0) = 0\,, \quad p(a,b+1) = p(a,b) + a\,.$$

We call $p(a,b)$ the **product** of a and b in $\mathbf{N}$, and we write it as

$$a \cdot b = p(a,b)\,.$$

1.61 Remark: It can be shown (using induction) that the sum and product satisfy all the usual properties in $\mathbf{N}$.

1.62 Definition: We define the set of **integers** to be the set

$$\mathbf{Z} = \big(\mathbf{N} \times \mathbf{N}\big)/R$$

where R is the equivalence relation given by

$$(a,b)R(c,d) \iff a + d = b + c\,.$$

For $[(a,b)]$ and $[(c,d)]$ in $\mathbf{Z}$, we define

$$[(a,b)] \leq [(c,d)] \iff b + c \leq a + d$$
$$[(a,b)] + [(c,d)] = [(a+c, b+d)]$$
$$[(a,b)] \cdot [(c,d)] = [(ac+bd, ad+bc)]\,.$$

For $n \in \mathbf{N}$, we write $n = [(n,0)]$ and $-n = [(0,n)]$, so that every element of $\mathbf{Z}$ can be written as $\pm n$ for some $n \in \mathbf{N}$, and we can identify $\mathbf{N}$ with a subset of $\mathbf{Z}$.

1.63 Remark: It can be shown that the ordering and the sum and product defined above are well-defined and satisfy the usual properties in $\mathbf{Z}$.

1.64 Definition: We define the set of **rational numbers** to be the set

$$\mathbf{Q} = \big(\mathbf{N} \times \mathbf{P}\big)/R$$

where $\mathbf{P} = \{x \in \mathbf{N} \big| x \neq 0\}$ and R is the equivalence relation given by

$$(a,b)R(c,d) \iff ad = bc\,.$$

For $[(a,b)]$ and $[(c,d)]$ in $\mathbf{Q}$, we define

$$[(a,b)] \leq [(c,d)] \iff a \cdot d \leq b \cdot c$$
$$[(a,b)] + [(c,d)] = [(a \cdot d + b \cdot c, b \cdot d)]$$
$$[(a,b)] \cdot [(c,d)] = [(a \cdot c, b \cdot d)]\,.$$

For $a \in \mathbf{N}$ and $b \in \mathbf{P}$, it is customary to write $\frac{a}{b} = [(a,b)]$. Also for $a \in \mathbf{Z}$ we write $a = [(a,1)]$, and we identify $\mathbf{Z}$ with a subset of $\mathbf{Q}$.

1.65 Remark: It can be shown that the above ordering, sum and product are well-defined and satisfy the usual rules in $\mathbf{Q}$.

1.66 Definition: We define the set of **real numbers** to be the set

$$\mathbf{R} = \left\{ x \subseteq \mathbf{Q} \,\middle|\, x \neq \emptyset \,,\; x \neq \mathbf{Q} \,,\; \forall a \in x \; \forall b \in \mathbf{Q} \; (b \leq a \to b \in x) \,,\; \forall a \in x \; \exists b \in x \; a < b \right\}.$$

For $x, y \in \mathbf{R}$, we define

$$x \leq y \Longleftrightarrow x \subseteq y$$
$$x + y = \left\{ a + b \,\middle|\, a, b \in \mathbf{Q} \,,\; a \in x \,,\; b \in y \right\}.$$

For $0 \leq x, y \in \mathbf{R}$ we define

$$x \cdot y = \left\{ a \cdot b \,\middle|\, 0 \leq a, b \in \mathbf{Q} \,,\; a \in x \,,\; b \in y \right\} \cup \left\{ c \in \mathbf{Q} \middle| c < 0 \right\},$$

and we leave, as an exercise, the definition of $x \cdot y$ in the case that $x < 0$ or $y < 0$.

1.67 Remark: It can be shown that the above ordering,sum and product are well-defined and satisfy the usual rules in $\mathbf{R}$.

1.68 Definition: We define the set of **complex** numbers to be the set

$$\mathbf{C} = \mathbf{R} \times \mathbf{R}.$$

We define addition and multiplication in $\mathbf{C}$ by

$$(a, b) + (c, d) = (a + c, b + d)$$
$$(a, b) \cdot (c, d) = (ac - bd, ad + bc).$$

We write $i = (0, 1)$. For $x \in \mathbf{R}$ we write $x = (x, 0)$, and we identify $\mathbf{R}$ with a subset of $\mathbf{C}$.

1.69 Remark: It can be shown that the above sum and product are well-defined and satisfy the usual rules in $\mathbf{C}$.

1.70 Definition: The **system of sets**, **indexed by** the set A, which we write as

$$\langle X_\alpha \big| \alpha \in A \rangle\,,$$

is defined to be the function F with domain A given by $F(\alpha) = X_\alpha$ for all $\alpha \in A$. We denote the range of this function F by

$$\big\{ X_\alpha \big| \alpha \in A \big\}\,.$$

1.71 Definition: Given an indexed system of sets $\langle X_\alpha \big| \alpha \in A\rangle$, we define the **union**, the **intersection** and the **product** of the family to be the sets

$$\bigcup_{\alpha\in A} X_\alpha = \bigcup \big\{ X_\alpha \big| \alpha \in A \big\} = \big\{ x \,\big|\, x \in X_\alpha \text{ for some } \alpha \in A \big\}$$
$$\bigcap_{\alpha\in A} X_\alpha = \bigcap \big\{ X_\alpha \big| \alpha \in A \big\} = \big\{ x \,\big|\, x \in X_\alpha \text{ for all } \alpha \in A \big\}$$
$$\prod_{\alpha\in A} X_\alpha = \Big\{ g : A \to \bigcup_{\alpha\in A} X_\alpha \Big| g(\alpha) \in X_\alpha \text{ for all } \alpha \in A \Big\}\,.$$

1.72 Remark: The Axiom of Choice is equivalent to the statement that for every indexed system of sets $\langle X_\alpha \big| \alpha \in A\rangle$ there exists a function $f : A \to \bigcup\limits_{\alpha\in A} X_\alpha$ such that

$$\forall \alpha \in A \left(X_\alpha \neq \emptyset \to f(\alpha) \in X_\alpha \right).$$

1.73 Definition: In the special case that A is a subset of $\mathbf{Z}$ of the form

$$A = \{ n \in \mathbf{Z} | n \geq a \} = \{ a, a+1, a+2, \cdots \}$$

for some $a \in \mathbf{Z}$, the system $\langle X_n \big| n \in A\rangle$ is called a **sequence** and it is denoted by

$$\langle X_n \rangle_{n\geq a} = \langle X_a, X_{a+1}, X_{a+2}, \cdots \rangle\,.$$

We use the notation

$$\bigcup_{n=a}^{\infty} X_n = \bigcup_{n\geq a} X_n = \bigcup_{n\in A} X_k$$
$$\bigcap_{n=a}^{\infty} X_k = \bigcap_{n\geq a} X_k = \bigcap_{n\in A} X_k$$
$$\prod_{n=a}^{\infty} X_k = \prod_{n\geq a} X_k = \prod_{n\in A} X_k\,.$$

1.74 Definition: In the special case that $A = \{n \in \mathbf{Z} | a \le n \le b\} = \{a, a+1, a+2, \cdots, b\}$ for some $a \le b \in \mathbf{Z}$, the system $\langle X_n \big| n \in A\rangle$ is called an **ordered** $(b-a)$**-tuple**, and it is written as

$$\langle X_n \big| a \le n \le b\rangle = \langle X_a, X_{a+1}, \cdots, X_b\rangle = (X_a, X_{a+1}, \cdots, X_b) \; .$$

We use the notation

$$X_a \cup X_{a+1} \cup \cdots \cup X_b = \bigcup_{n=a}^{b} X_n = \bigcup_{n \in A} X_n$$
$$X_a \cap X_{a+1} \cap \cdots \cap X_b = \bigcap_{n=a}^{b} X_n = \bigcap_{n \in A} X_n$$
$$X_a \times X_{a+1} \times \cdots \times X_b = \prod_{n=a}^{b} X_n = \prod_{n \in A} X_n \; .$$

1.75 Remark: This notation is not equivalent to our earlier notation for (u, v) and $u \times v$, but this inconsistency of notation usually causes no difficulty.

1.76 Definition: For a set X and for $n \in \mathbf{N}$ we define

$$\begin{aligned} X^n &= \big\{(x_0, \cdots, x_{n-1}) \big| \text{ each } x_i \in X\big\} \\ &= \big\{f \,\big|\, f : n \to X\big\} \, . \end{aligned}$$

More generally, for any sets A and B we sometimes write

$$B^A = \big\{f \,\big|\, f : A \to B\big\} \, .$$

Appendix 2: Exponential and Trigonometric Functions

2.1 Definition: Let X and Y be sets and let $f : X \to Y$. We say that f is **injective** (or **one-to-one**, written as $1\!:\!1$) when for every $y \in Y$ there exists at most one $x \in X$ such that $f(x) = y$. Equivalently, f is injective when for all $x_1, x_2 \in X$, if $f(x_1) = f(x_2)$ then $x_1 = x_2$. We say that f is **surjective** (or **onto**) when for every $y \in Y$ there exists at least one $x \in X$ such that $f(x) = y$. Equivalently, f is surjective when $\text{Range}(f) = Y$. We say that f is **bijective** (or **invertible**) when f is both injective and surjective, that is when for every $y \in Y$ there exists exactly one $x \in X$ such that $f(x) = y$. When f is bijective, we define the **inverse** of f to be the function $f^{-1} : Y \to X$ such that for all $y \in Y$, $f^{-1}(y)$ is equal to the unique element $x \in X$ such that $f(x) = y$. Note that when f is bijective so is f^{-1}, and in this case we have $(f^{-1})^{-1} = f$.

2.2 Example: Let $f(x) = \frac{1}{3}\sqrt{12x - x^2}$ for $0 \le x \le 6$. Show that f is injective and find a formula for its inverse function.

Solution: Note that when $0 \le x \le 6$ (indeed when $0 \le x \le 12$) we have $12x - x^2 = x(12 - x) \ge 0$, so that $\frac{1}{3}\sqrt{12x - x^2}$ exists, and we have $12x - x^2 = 36 - (x - 6)^2 \le 36$ so that $\frac{1}{3}\sqrt{12x - x^2} \le \frac{1}{3}\sqrt{36} = 2$. Thus if $0 \le x \le 6$ then $f(x) = \frac{1}{3}\sqrt{12x - x^2}$ exists and we have $0 \le f(x) \le 2$. Let $x, y \in \mathbf{R}$ with $0 \le x \le 6$ and $0 \le y \le 2$. Then we have

$$\begin{aligned} y = f(x) &\leftrightarrow y = \tfrac{1}{3}\sqrt{12x - x^2} \\ &\leftrightarrow 3y = \sqrt{12x - x^2} \\ &\leftrightarrow 9y^2 = 12x - x^2 \quad \text{, since } y \ge 0 \\ &\leftrightarrow x^2 - 12x + 9y^2 = 0 \\ &\leftrightarrow x = \frac{12 \pm \sqrt{144 - 36y^2}}{2} = 6 \pm 3\sqrt{4 - y^2} \quad \text{, by the Quadratic Formula} \\ &\leftrightarrow x = 6 - 3\sqrt{4 - y^2} \ \text{ since } x \le 6. \end{aligned}$$

Verify that when $0 \le y \le 2$ we have $0 \le 4 - y^2 \le 4$ so that $\sqrt{4 - y^2}$ exists and we have $0 \le 6 - 3\sqrt{4 - y^2} \le 6$. Thus when we consider f as a function $f : [0, 6] \to [0, 2]$, it is bisectve and its inverse $f^{-1} : [0, 2] \to [0, 6]$ is given by $f^{-1}(y) = 6 - 3\sqrt{4 - y^2}$.

2.3 Definition: Let F be a field and let $f : A \subseteq F \to F$. We say that f is **even** when $f(-x) = f(x)$ for all $x \in F$ and we say that f is **odd** when $f(-x) = -f(x)$ for all $x \in F$.

2.4 Definition: Let F be an ordered field and let $f : A \subseteq\to F$. We say that f is **increasing** when it has the property that for all $x, y \in A$, if $x < y$ then $f(x) < f(y)$, and we say f is decreasing when for all $x, y \in A$ with $x < y$ we have $f(x) > f(y)$. We say that f is **monotonic** when f is either increasing or decreasing. Note that every monotonic function is injective.

2.5 Remark: We assume familiarity with exponential, logarithmic, trigonometric and inverse trigonometric functions. These functions can be defined rigorously. We shall give a brief description of how one can define the exponential and logarithmic function rigorously, and we shall provide an informal (non-rigorous) description of the trigonometric and inverse trigonometric functions.

2.6 Definition: Let us outline one possible way to define the value of x^y for suitable real numbers $x, y \in \mathbf{R}$. First we define $x^0 = 1$ for all $x \in \mathbf{R}$. Then for $n \in \mathbf{Z}$ with $n \geq 1$ we define x^n recursively by $x^n = x \cdot x^{n-1}$ for all $x \in \mathbf{R}$. Also, for $n \in \mathbf{Z}$ with $n \geq 1$ we define $x^{-n} = \frac{1}{x^n}$ for all $x \neq 0$. At this stage we have defined x^y for $y \in \mathbf{Z}$.

When $0 < n \in \mathbf{Z}$ is odd, for all $x \in \mathbf{R}$ we define $x^{1/n} = y$ where y is the unique real number such that $y^n = x$ (to be rigorous, one must prove that this number y exists and is unique). When $0 < n \in \mathbf{Z}$ is even, for $x \geq 0$ we define $x^{1/n} = y$ where y is the unique nonnegative real number such that $y^n = x$ (again, to be rigorous a proof is required). Also, for $0 < n \in \mathbf{Z}$ we define $x^{-1/n} = \frac{1}{x^{1/n}}$, which is defined for $x \neq 0$ if n is odd, and is defined for $x > 0$ when n is even. When $n, m \in \mathbf{Z}$ with $n > 0$ and $m > 0$ and $\gcd(n, m) = 1$, we define $x^{n/m} = (x^n)^{1/m}$, which is defined for all $x \in \mathbf{R}$ when m is odd and for $x \geq 0$ when m is even, and we define $x^{-n/m} = \frac{1}{x^{n/m}}$, defined for $x \neq 0$ when m is odd and for $x > 0$ when m is even. At this stage, we have defined x^y for $y \in \mathbf{Q}$.

When $x > 1$ and $y \in \mathbf{R}$, we define $x^y = \sup\left\{x^t \middle| t \in \mathbf{Q}, t \leq y\right\}$ (to be rigorous, one needs to prove that the supremum exists and that when $y \in \mathbf{Q}$ this agrees with our previous definition). When $0 < x < 1$ and $y \in \mathbf{R}$ we define $x^y = \inf\left\{x^t \middle| t \in \mathbf{Q}, t \leq y\right\}$. Finally, we define $1^y = 1$ for all $y \in \mathbf{R}$ and we define $0^y = 0$ for all $y > 0$.

2.7 Theorem: *(Properties of Exponentials) Let $a, b, x, y \in \mathbf{R}$ with $a, b > 0$. Then*
(1) $a^0 = 1$,
(2) $a^{x+y} = a^b\, a^c$,
(3) $a^{x-y} = a^x / a^y$,
(4) $(a^x)^y = a^{xy}$,
(5) $(ab)^x = a^x b^x$.

Proof: We omit the proof.

2.8 Theorem: *(Properties of Power Functions)*
(1) When $a > 0$, the function $f : [0, \infty) \to [0, \infty)$ given by $f(x) = x^a$ is increasing and bijective and its inverse function is given by $f^{-1}(x) = x^{1/a}$.
(2) When $a < 0$, the function $f : (0, \infty) \to (0, \infty)$ given by $f(x) = x^a$ is decreasing and bijective and its inverse is given by $f^{-1}(x) = a^{1/x}$.

Proof: We omit the proof.

2.9 Definition: A function of the form $f(x) = x^a$ is called a **power function**.

2.10 Theorem: *(Properties of Exponential Functions)*

(1) When $a > 1$ *the function* $f : \mathbf{R} \to (0, \infty)$ *given by* $f(x) = a^x$ *is increasing and bijective.*
(2) When $0 < a < 1$ *the function* $f : \mathbf{R} \to (0, \infty)$ *given by* $f(x) = a^x$ *is decreasing and bijective.*

Proof: We omit the proof.

2.11 Definition: For $a > 0$ with $a \neq 1$, the function $f : \mathbf{R} \to (0, \infty)$ given by $f(x) = a^x$ is called the base a **exponential function**, its inverse function $f^{-1} : (0, \infty) \to \mathbf{R}$ is called the base a **logarithmic function**, and we write $f^{-1}(x) = \log_a x$. By the definition of the inverse function, we have $\log_a(a^x) = x$ for all $x \in \mathbf{R}$ and $e^{\log_a y} = y$ for all $y > 0$, and for all $x, y \in \mathbf{R}$ with $y > 0$ we have $y = a^x \leftrightarrow x = \log_a y$.

2.12 Theorem: *(Properties of Logarithms) Let* $a, b, x, y \in (0, \infty)$*. Then*

(1) $\log_a 1 = 0$,
(2) $\log_a(xy) = \log_a x + \log_a y$,
(3) $\log_a(x/y) = \log_a x - \log_a y$,
(4) $\log_a(x^y) = y \log_a x$, *and*
(5) $\log_b x = \log_a x / \log_a b$,
(6) if $a > 1$, *the function* $g : (0, \infty) \to \mathbf{R}$ *given by* $g(x) = \log_a x$ *is increasing and bijective.*

Proof: The proof is left as an exercise.

2.13 Definition: There is a number $e \in \mathbf{R}$ called **natural base**, with $e \cong 2.71828$, which can be defined in many ways, for example we can define

$$e = \sup \left\{ \left(1 + \tfrac{1}{n}\right)^n \middle| 1 \leq n \in \mathbf{Z} \right\}$$

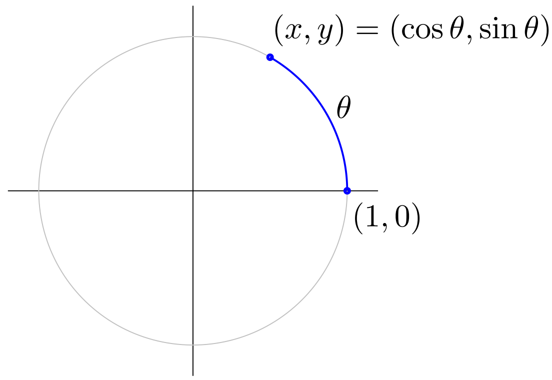
(to be rigorous, one must prove that the set $A = \left\{(1 + \frac{1}{n})^n \middle| 1 \leq n \in \mathbf{Z}\right\}$ is bounded above). The logarithm to the base e is called the **natural logarithm**, and we write

$$\ln x = \log_e x \text{ for } x > 0.$$

The properties of exponentials and logarithms in Theorems 2.13 and 2.18 give

$$e^0 = 1 \ , \ a^{x+y} = e^x e^y \ , \ e^{x-y} = e^x / e^y \ , \ (e^x)^y = e^{xy},$$
$$\ln 1 = 0 \ , \ \ln(xy) = \ln x + \ln y \ , \ \ln(x/y) = \ln x - \ln y \ , \ \ln x^y = y \ln x$$
$$\log_a x = \frac{\ln x}{\ln a} \ \text{ and } \ a^x = e^{x \ln a}.$$

2.14 Definition: We define the trigonometric functions informally as follows. For $\theta \geq 0$, we define $\cos\theta$ and $\sin\theta$ to be the x- and y-coordinates of the point at which we arrive when we begin at the point $(1,0)$ and travel for a distance of θ units counterclockwise around the unit circle $x^2 + y^2 = 1$. For $\theta \leq 0$, $\cos\theta$ and $\sin\theta$ are the x and y-coordinates of the point at which we arrive when we begin at $(1,0)$ and travel clockwise around the unit circle for a distance of $|\theta$ units. When $\cos\theta \neq 0$ we define $\sec\theta = 1/\cos\theta$ and $\tan\theta = \sin\theta/\cos\theta$, and when $\sin\theta \neq 0$ we define $\csc\theta = 1/\sin\theta$ and $\cot\theta = \cos\theta/\sin\theta$. (This definition is not rigorous because we did not define what it means to travel around the circle for a given distance).

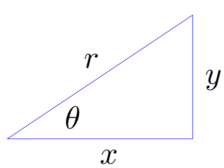


2.15 Definition: We define π, informally, to be the distance along the top half of the unit circle from $(1,0)$ to $(-1,0)$, and so we have $\cos\pi = -1$ and $\sin\pi = 0$. By symmetry, the distance from $(1,0)$ to $(0,1)$ along the circle is equal to $\frac{\pi}{2}$ so we also have $\cos\frac{\pi}{2} = 0$ and $\sin\frac{\pi}{2} = 1$.

2.16 Theorem: *(Basic Trigonometric Properties) For $\theta \in \mathbf{R}$ we have*
(1) $\cos^2\theta + \sin^2\theta = 1$,
(2) $\cos(-\theta) = \cos\theta$ *and* $\sin(-\theta) = -\sin\theta$,
(3) $\cos(\theta + \pi) = -\cos\theta$ *and* $\sin(\theta + \pi) = -\sin\theta$,
(4) $\cos(\theta + 2\pi) = \cos\theta$ *and* $\sin(\theta + 2\pi) = \sin\theta$*).*

Proof: Informally, these properties can all be seen immediately from the above definitions. We omit a rigorous proof.

2.17 Theorem: *(Trigonometric Ratios) Let $\theta \in \left(0, \frac{\pi}{2}\right)$. For a right angle triangle with an angle of size θ and with sides of lengths x, y and r as shown, we have*



$$\cos\theta = \frac{x}{r}\ ,\ \sin\theta = \frac{y}{r}\ \text{ and }\ \tan\theta = \frac{y}{x}.$$

Proof: We can see this informally by scaling the picture in Definition 2.17 by a factor of r.

2.18 Theorem: *(Special Trigonometric Values) We have the following exact trigonometric values.*

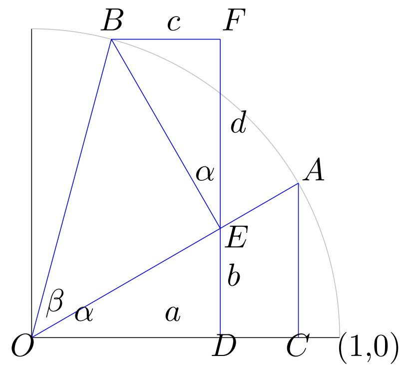
θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\cos\theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\sin\theta$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1

Proof: This follows from the above theorem using certain particular right angled triangles.

2.19 Theorem: *(Trigonometric Sum Formulas) For* $\alpha, \beta \in \mathbf{R}$ *we have*

$$\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \text{ , and}$$
$$\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta.$$

Proof: Informally, we can prove this with the help of a picture. The picture below illustrates the situation when $\alpha, \beta \in \left(0, \frac{\pi}{2}\right)$.



In the picture, O is the origin, A is the point with coordinates $(\cos\alpha, \sin\alpha)$ and B is the point $(x,y) = \big(\cos(\alpha+\beta), \sin(\alpha+\beta)\big)$. In triangle ODE we see that $\cos\alpha = \frac{OD}{OE} = \frac{a}{\cos\beta}$ and $\sin\alpha = \frac{DE}{OE} = \frac{b}{\cos\beta}$, and so $a = \cos\alpha\cos\beta$, $b = \sin\alpha\cos\beta$. In triangle EFB, verify that the angle at E has size α, and so we have $\cos\alpha = \frac{EF}{EB} = \frac{d}{\sin\beta}$ and $\sin\alpha = \frac{BF}{BE} = \frac{c}{\sin\beta}$, and so $c = \sin\alpha\sin\beta$, $d = \cos\alpha\sin\beta$. The x and y-coordinates of the point B are $x = a - c$ and $y = b + d$, and so

$$\cos(\alpha+\beta) = x = a - c = \cos\alpha\cos\beta - \sin\alpha\sin\beta \text{ , and}$$
$$\sin(\alpha+\beta) = y = b + d = \sin\alpha\cos\beta - \cos\alpha\sin\beta.$$

This proves the theorem (informally) in the case that $\alpha, \beta \in \left(0, \frac{\pi}{2}\right)$. One can then show that the theorem holds for all $\alpha, \beta \in \mathbf{R}$ by using the Basic Trigonometric Properties (2), (3) and (4).

2.20 Theorem: *(Double Angle Formulas) For all* $x, y \in \mathbf{R}$ *we have*

(1) $\sin 2x = 2\sin x\cos x$ *and* $\cos 2x = \cos^2 - \sin^2 x = 2\cos^2 x - 1 = 1 - 2\sin^2 x$*, and*

(2) $\cos^2 x = \dfrac{1+\cos 2x}{2}$ *and* $\sin^2 x = \dfrac{1-\cos 2x}{2}$.

Proof: The proof is left as an exercise.

2.21 Theorem: *(Trigonometric Functions)*

(1) The function $f : [0,\pi] \to [-1,1]$ *defined by* $f(x) = \cos x$ *is decreasing and bijective.*
(2) The function $g : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to [-1,1]$ *given by* $g(x) = \sin x$ *is increasing and bijective.*
(3) The function $h : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ *given by* $h(x) = \tan x$ *is increasing and bijective.*

Proof: We omit the proof.

2.22 Definition: The inverses of the functions f, g and h in the above theorem are called the **inverse cosine**, the **inverse sine**, and the **inverse tangent** functions. We write $f^{-1}(x) = \cos^{-1} x$, $g^{-1} = \sin^{-1} x$ and $h^{-1}(x) = \tan^{-1} x$. By the definition of the inverse function, we have

2.23 Definition: Let A and B be sets, let F be a field, let $c \in F$. Let $f : A \to F$ and $g : B \to F$. We define the functions cf , $f + g$, $f - g$, $f \cdot g$: $A \cap B \to F$ by

$$\begin{aligned}(cf)(x) &= c\,f(x)\\ (f+g)(x) &= f(x)+g(x)\\ (f-g)(x) &= f(x)-g(x)\\ (f\cdot g)(x) &= f(x)g(x)\end{aligned}$$

for all $x \in A \cap B$, and for $C = \{x \in A \cap B \,|\, g(x) \neq 0\}$ we define $f/g : C \to F$ by

$$(f/g)(x) = f(x)/g(x)$$

for all $x \in C$.

2.24 Definition: A **polynomial function** over a field F is a function $f : F \to F$ which can be obtained from the functions 1 and x using (finitely many applications of) the operations cf, $f + g$, $f - g$, $f \cdot g$ and $f \circ g$. In other words, a polynomial is a function of the form

$$f(x) = \textstyle\sum_{i=0}^{n} c_i x^i = c_0 + c_1 x + c_2 x^2 + \cdots + c_n x^n$$

for some $n \in \mathbf{N}$ and some $c_i \in F$. The numbers c_i are called the **coefficients** of the polynomial and when $c_n \neq 0$ the number n is called the **degree** of the polynomial.

2.25 Definition: A **rational function** over a field F is a function $f : A \subseteq F \to F$ which can be obtained from the functions 1 and x using (finitely many applications of) the operations cf, $f + g$, $f - g$, $f \cdot g$, f/g and $f \circ g$. In other words, a rational function is a function of the form

$$f(x) = p(x)/q(x)$$

for some polynomials p and q.

2.26 Definition: The functions 1, x, $x^{1/n}$ with $0 < n \in \mathbf{Z}$, e^x, $\ln x$, $\sin x$ and $\sin^{-1} x$, are called the **basic elementary functions**. An **elementary function** is any function $f : A \subseteq \mathbf{R} \to \mathbf{R}$ which can be obtained from the basic elementary functions using (finitely many applications of) the operations cf, $f + g$, $f - g$, $f \cdot g$, f/g and $f \circ g$.

2.27 Example: The following functions are elementary

$$\begin{aligned}|x| &= \sqrt{x^2},\\ \cos x &= \sin\left(x + \tfrac{\pi}{2}\right),\\ \tan^{-1} x &= \sin^{-1}\left(\frac{x}{\sqrt{1+x^2}}\right),\\ f(x) &= \frac{e^{\sqrt{x}+\sin x}}{\tan^{-1}(\ln x)}\end{aligned}$$

We shall see later that every elementary function is continuous in its domain, so any function which is discontinuos at a point in its domain cannot be elementary.