Chapter 1. Logic

To reason correctly, we have to follow some rules. These rules of reasoning are what we called *logic.* We will only need a few of these rules, mainly to deal with taking opposite of statements and to handle conditional statements.

We will use the symbol $\sim$ (or $\neg$) to denote the word "not". Also, we will use the symbol $\forall$ to denote "for all", "for any", "for every". Similarly, the symbol $\exists$ will denote "there is (at least one)", "there exists", "there are (some)" and usually followed by "such that". The symbols $\forall$ and $\exists$ are called *quantifiers.*

Negation. Below we will look at rules of negation (i.e. taking opposite). They are needed when we do indirect proofs (or proofs by contradiction). For any expression p, we have $\sim(\sim p) = p$.

Examples. (1)

$$\begin{aligned} \text{expression :}\quad & \overbrace{x > 0}^{p} \ \text{ and } \ \overbrace{x < 1}^{q} \\ \text{opposite expression :}\quad & x \le 0 \ \text{ or } \ x \ge 1 \\ \text{rule :}\quad & \sim(p \ \text{ and } \ q) = (\sim p) \text{ or } (\sim q) \end{aligned}$$

---

(2)

$$\begin{aligned} \text{expression :}\quad & x < 0 \ \text{ or } \ x > 1 \\ \text{opposite expression :}\quad & x \ge 0 \ \text{ and } \ x \le 1 \\ \text{rule :}\quad & \sim(p \ \text{ or } \ q) = (\sim p) \text{ and } (\sim q) \end{aligned}$$

---

(3)

statement : For every $x \ge 0$, x has a square root. (True)

quantified statement : $\forall x \ge 0 \quad (x$ has a square root).

opposite statement : There exists $x \ge 0$ such that x does not have a square root. (False)

quantified opposite statement : $\exists x \ge 0 \quad \sim(x$ has a square root).

---

(4)

statement : For every $x \ge 0$, there is $y \ge 0$ such that $y^2 = x$. (True)

quantified statement : $\forall x \ge 0, \quad \exists y \ge 0 \quad (y^2 = x)$.

opposite statement : There exists $x \ge 0$ such that for every $y \ge 0$, $y^2 \ne x$. (False)

quantified opposite statement : $\exists x \ge 0, \quad \forall y \ge 0 \quad \sim(y^2 = x)$.

From examples (3) and (4), we see that the rule for negating statements with quantifiers is first *switch every* $\forall$ *to* $\exists$ *and every* $\exists$ *to* $\forall$, then *negate the remaining part of the statement.*

If-then Statements. If-then statements occur frequently in mathematics. We will need to know some equivalent ways of expressing an if-then statement to do proofs. The statement "if p, then q" may also be stated as "p implies q", "p only if q", "p is sufficient for q", "q is necessary for p" and is commonly denoted by "$p \implies q$". For example, the statement "if $x = 3$ and $y = 4$, then $x^2 + y^2 = 25$" may also be stated as "$x = 3$ and $y = 4$ are sufficient for $x^2 + y^2 = 25$" or "$x^2 + y^2 = 25$ is necessary for $x = 3$ and $y = 4$".

Example. (5)

$$\begin{aligned} \text{statement :}\quad & \text{If } x \ge 0, \text{ then } |x| = x. \text{ (True)} \\ \text{opposite statement :}\quad & x \ge 0 \ \text{ and } \ |x| \ne x. \text{ (False)} \\ \text{rule :}\quad & \sim(p \implies q) = \ p \ \text{ and } \ (\sim q) \end{aligned}$$

Remark. Note

$$\begin{aligned} p \implies q &= \sim(\sim(p \implies q)) \\ &= \sim(p \ \text{ and } \ (\sim q)) \\ &= (\sim p) \ \text{ or } \ \sim(\sim q) \\ &= (\sim p) \ \text{ or } \ q. \end{aligned}$$

For the statement "if p, then q" ($p \implies q$), there are two related statements: the *converse* of the statement is "if q, then p" ($q \implies p$) and the *contrapositive* of the statement is "if $(\sim q)$, then $(\sim p)$" ($\sim q \implies \sim p$).

Examples. (6)

statement :	If $x = -3$, then $x^2 = 9$. (True)
converse :	If $x^2 = 9$, then $x = -3$. (False, as x may be 3.)
contrapositive :	If $x^2 \neq 9$, then $x \neq -3$. (True)

---

(7)

statement :	$x = -3 \implies 2x = -6$ (True)
converse :	$2x = -6 \implies x = -3$ (True)
contrapositive :	$2x \neq -6 \implies x \neq -3$ (True)

---

(8)

statement :	If $\lvert x\rvert = 3$, then $x = -3$. (False, as x may be 3.)
converse :	If $x = -3$, then $\lvert x\rvert = 3$. (True)
contrapositive :	If $x \neq -3$, then $\lvert x\rvert \neq 3$. (False, as x may be 3.)

Remarks. Examples (6) and (7) showed that the converse of an if-then statement is <u>not</u> the same as the statement nor the opposite of the statement in general. Examples (6), (7) and (8) showed that an if-then statement and its contrapositive are either both true or both false. In fact, this is always the case because by the remark on the last page,

$$\begin{aligned}(\sim q) \implies (\sim p) &= \sim(\sim q) \text{ or } (\sim p)\\ &= q \text{ or } (\sim p)\\ &= (\sim p) \text{ or } q\\ &= p \implies q.\end{aligned}$$

So an if-then statement and its contrapositive statement are <u>equivalent</u>.

Finally, we introduce the terminology "p if and only if q" to mean "if p, then q" and "if q, then p". The statement "p if and only if q" is the same as "p is necessary and sufficient for q". We abbreviate "p if and only if q" by "$p \iff q$". So $p \iff q$ means $p \implies q$ and $q \implies p$". <u>**The phrase "if and only if" is often abbreviated as "iff".**</u>

<u>**Caution!**</u> Note $\forall\alpha\forall\beta = \forall\beta\forall\alpha$ and $\exists\alpha\exists\beta = \exists\beta\exists\alpha$, but $\forall\alpha\exists\beta \neq \exists\beta\forall\alpha$. For example, "every student is assigned a number" is the same as "∀ student, ∃ number such that the student is assigned the number." This statement implies different students may be assigned possibly different numbers. However, if we switch the order of the quantifiers, the statement becomes "∃ number such that ∀ student, the student is assigned the number." This statement implies there is a number and every student is assigned that same number!

Chapter 2. Sets

To read and write mathematical expressions accurately and concisely, we will introduce the language of sets. A *set* is a collection of "objects" (usually numbers, ordered pairs, functions, etc.) If object x is in a set S, then we say x is an *element* (or a *member*) of S and write $x \in S$. If x is not an element of S, then we write $x \notin S$. A set having finitely many elements is called a *finite* set, otherwise it is called an *infinite* set. The *empty set* is the set having no objects and is denoted by $\emptyset$.

A set may be shown by listing its elements enclosed in braces (eg. $\{1, 2, 3\}$ is a set containing the objects 1, 2, 3, the positive integer $\mathbb{N} = \{1, 2, 3, \ldots\}$, the integer $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, the empty set $\emptyset = \{\}$) or by description enclosed in braces (eg. the rational numbers $\mathbb{Q} = \{\frac{m}{n} : m \in \mathbb{Z}, n \in \mathbb{N}\}$, the real numbers $\mathbb{R} = \{x : x \text{ is a real number}\}$ and the complex numbers $\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$.) In describing sets, the usual convention is to put the *form* of the objects on the left side of the colon and to state the *conditions* on the objects on the right side of the colon. The set will consist of **all** elements satisifying all the conditions. It is also common to use a vertical bar in place of colon in set descriptions.

Examples. (i) The closed interval with endpoints a, b is $[a, b] = \{x : \ x \in \mathbb{R} \ \text{ and } \ a \leq x \leq b\}$.

(ii) The set of square numbers is $\{1, 4, 9, 16, 25, \ldots\} = \{n^2 : n \in \mathbb{N}\}$.

(iii) The set of all positive real numbers is $\mathbb{R}^+ = \{x : \ x \in \mathbb{R} \ \text{ and } \ x > 0\}$. (If we want to emphasize this is a subset of $\mathbb{R}$, we may stress x is real in the form of the objects and write $\mathbb{R}^+ = \{x \in \mathbb{R} : \ x > 0\}$. If numbers are always taken to mean real numbers, then we may write simply $\mathbb{R}^+ = \{x : \ x > 0\}$.)

(iv) The set of points (or ordered pairs) on the line ℓ_m with equation $y = mx$ is $\{(x, mx) : x \in \mathbb{R}\}$.

For sets A, B, we say A is a *subset* of B (or B *contains* A) iff every element of A is also an element of B. In that case, we write $A \subseteq B$. (For the case of the empty set, we have $\emptyset \subseteq S$ for every set S.) Two sets A and B are *equal* if and only if they have the same elements (i.e. $A = B$ means $A \subseteq B$ and $B \subseteq A$.) So $A = B$ if and only if ($x \in A \iff x \in B$). If $A \subseteq B$ and $A \neq B$, then we say A is a *proper subset* of B and write $A \subset B$. (For example, if $A = \{1, 2\}$, $B = \{1, 2, 3\}$, $C = \{1, 1, 2, 3\}$, then $A \subset B$ is true, but $B \subset C$ is false. In fact, $B = C$. Repeated elements are counted only one time so that C has 3 elements, not 4 elements.)

For a set S, we can collect all its subsets. This is called the *power set* of S and is denoted by $P(S)$ or 2^S. For examples, $P(\emptyset) = \{\emptyset\}$, $P(\{0\}) = \{\emptyset, \{0\}\}$ and $P(\{0, 1\}) = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. For a set with n elements, its power set will have 2^n element. This is the reason for the alternative notation 2^S for the power set of S. Power set is one operation of a set. There are a few other common operations of sets.

Definitions. For sets $A_1, A_2, \ldots, A_n$,

(i) their *union* is $A_1 \cup A_2 \cup \cdots \cup A_n = \{x : x \in A_1 \text{ or } x \in A_2 \text{ or } \cdots \text{ or } x \in A_n\}$,

(ii) their *intersection* is $A_1 \cap A_2 \cap \cdots \cap A_n = \{x : x \in A_1 \text{ and } x \in A_2 \text{ and } \cdots \text{ and } x \in A_n\}$,

(iii) their *Cartesian product* is

$$A_1 \times A_2 \times \cdots \times A_n = \{(x_1, x_2, \ldots, x_n) : x_1 \in A_1 \text{ and } x_2 \in A_2 \text{ and } \cdots \text{ and } x_n \in A_n\},$$

(iv) the *complement* of A_2 in A_1 is $A_1 \setminus A_2 = \{x : x \in A_1 \text{ and } x \notin A_2\}$.

Examples. (i) $\{1, 2, 3\} \cup \{3, 4\} = \{1, 2, 3, 4\}$, $\{1, 2, 3\} \cap \{2, 3, 4\} = \{2, 3\}$, $\{1, 2, 3\} \setminus \{2, 3, 4\} = \{1\}$.

(ii) $[-2, 4] \cap \mathbb{N} = \{1, 2, 3, 4\}$, $[0, 2] \cup [1, 5] \cup [4, 6] = [0, 6]$.

(iii) $([0, 7] \cap \mathbb{Z}) \setminus \{n^2 : n \in \mathbb{N}\} = \{0, 1, 2, 3, 4, 5, 6, 7\} \setminus \{1, 4, 9, 16, 25, \ldots\} = \{0, 2, 3, 5, 6, 7\}$.

(iv) $\mathbb{R} \times \mathbb{R} \times \mathbb{R} = \{(x, y, z) : x, y, z \in \mathbb{R}\}$, $\mathbb{Q} \times (\mathbb{R} \setminus \mathbb{Q}) = \{(a, b) : a \text{ is rational and } b \text{ is irrational}\}$.

Remarks. (i) For the case of the empty set, we have

$$A \cup \emptyset = A = \emptyset \cup A, \quad A \cap \emptyset = \emptyset = \emptyset \cap A, \quad A \times \emptyset = \emptyset = \emptyset \times A, \quad A \setminus \emptyset = A \ \text{ and } \ \emptyset \setminus A = \emptyset.$$

(ii) The notions of union, intersection and Cartesian product may be extended to *infinitely many* sets similarly. The union is the set of objects in at least one of the sets. The intersection is the set of objects in every one of the sets. The Cartesian product is the set of ordered tuples such that the i-th coordinate must belong to the i-th set.

(iii) The set $A_1 \cup A_2 \cup \cdots \cup A_n$ may be written as $\bigcup_{k=1}^{n} A_k$. If for every positive integer k, there is a set A_k, then the notation $A_1 \cup A_2 \cup A_3 \cup \cdots$ may be abbreviated as $\bigcup_{k=1}^{\infty} A_k$ or $\bigcup_{k\in\mathbb{N}} A_k$. If for every $x \in S$, there is a set A_x, then the union of all the sets A_x's for all $x \in S$ is denoted by $\bigcup_{x\in S} A_x$. Similar abbreviations exist for intersection and Cartesian product.

Examples. (i) $([1,2] \cup [2,3] \cup [3,4] \cup [4,5] \cup \cdots) \cap \mathbb{Z} = [1,+\infty) \cap \mathbb{Z} = \mathbb{N}$.

(ii) $\bigcap_{n\in\mathbb{N}} \Big[0, 1+\frac{1}{n}\Big) = [0,2) \cap \Big[0, 1\frac{1}{2}\Big) \cap \Big[0, 1\frac{1}{3}\Big) \cap \Big[0, 1\frac{1}{4}\Big) \cap \cdots = [0,1]$.

(iii) For every $k \in \mathbb{N}$, let $A_k = \{0, 1\}$, then

$$A_1 \times A_2 \times A_3 \times \cdots = \{(x_1, x_2, x_3, \ldots) : \text{ each } x_k \text{ is 0 or 1 for } k = 1, 2, 3, \ldots\}.$$

(iv) For each $m \in \mathbb{R}$, let ℓ_m be the line with equation $y = mx$ on the plane, then $\bigcup_{m\in\mathbb{R}} \ell_m = \mathbb{R}^2 \setminus \{(0, y) : y \in \mathbb{R}, y \neq 0\}$ and $\bigcap_{m\in\mathbb{R}} \ell_m = \{(0,0)\}$.

(v) Show that if $A \subseteq B$ and $C \subseteq D$, then $A \cap C \subseteq B \cap D$.

Reason. For every $x \in A \cap C$, we have $x \in A$ and $x \in C$. Since $A \subseteq B$ and $C \subseteq D$, we have $x \in B$ and $x \in D$, which imply $x \in B \cap D$. Thus, we see that every element in $A \cap C$ is also in $B \cap D$. Therefore, $A \cap C \subseteq B \cap D$.

(vi) Show that $(A \cup B) \setminus C = (A \setminus C) \cup (B \setminus C)$.

Reason. For every $x \in (A \cup B) \setminus C$, we have $x \in A \cup B$ and $x \notin C$. So either $x \in A$ or $x \in B$. In the former case, $x \in A \setminus C$ or in the latter case, $x \in B \setminus C$. So $x \in (A \setminus C) \cup (B \setminus C)$. Hence, $(A \cup B) \setminus C \subseteq (A \setminus C) \cup (B \setminus C)$.

Conversely, for every $x \in (A \setminus C) \cup (B \setminus C)$, either $x \in A \setminus C$ or $x \in B \setminus C$. In the former case, $x \in A$ and $x \notin C$ or in the latter case, $x \in B$ and $x \notin C$. In both cases, $x \in A \cup B$ and $x \notin C$. So $x \in (A \cup B) \setminus C$. Hence, $(A \setminus C) \cup (B \setminus C) \subseteq (A \cup B) \setminus C$. Combining with the conclusion of the last paragraph, we have $(A \cup B) \setminus C = (A \setminus C) \cup (B \setminus C)$.

We shall say that sets are *disjoint* iff their intersection is the empty set. Also, we say they are *mutually disjoint* iff the intersection of every pair of them is the empty set. A *relation* on a set E is any subset of $E \times E$. The following is an important concept that is needed in almost all branches of mathematics. It is a tool to divide (or partition) the set of objects we like to study into mutually disjoint subsets.

Definition. An *equivalence relation* R on a set E is a subset R of $E \times E$ such that

(a) (*reflexive property*) for every $x \in E$, $(x, x) \in R$,

(b) (*symmetric property*) if $(x, y) \in R$, then $(y, x) \in R$,

(c) (*transitive property*) if $(x, y), (y, z) \in R$, then $(x, z) \in R$.

We write $x \sim y$ if $(x, y) \in R$. For each $x \in E$, let $[x] = \{y : x \sim y\}$. This is called the *equivalence class containing* x. Note that every $x \in [x]$ by (a) so that $\bigcup_{x\in E} [x] = E$. If $x \sim y$, then $[x] = [y]$ because by (b) and (c), $z \in [x] \iff z \sim x \iff z \sim y \iff z \in [y]$. If $x \not\sim y$, then $[x] \cap [y] = \emptyset$ because assuming $z \in [x] \cap [y]$ will lead to $x \sim z$ and $z \sim y$, which imply $x \sim y$, a contradiction. So every pair of equivalence classes are either the same or disjoint. Therefore, R partitions the set E into mutually disjoint equivalence classes.

Examples. (1) (*Geometry*) For triangles T_1 and T_2, define $T_1 \sim T_2$ if and only if T_1 is similar to T_2. This is an equivalence relation on the set of all triangles as the three properties above are satisfied. For a triangle T, $[T]$ is the set of all triangles similar to T.

(2) (*Arithmetic*) For integers m and n, define $m \sim n$ if and only if $m - n$ is even. Again, properties (a), (b), (c) can easily be verified. So this is also an equivalence relation on $\mathbb{Z}$. There are exactly two equivalence classes, namely $[0] = \{\dots, -4, -2, 0, 2, 4, \dots\}$ (even integers) and $[1] = \{\dots, -5, -3, -1, 1, 3, 5, \dots\}$ (odd integers). Two integers in the same equivalence class is said to be *of the same parity.*

(3) Some people think that properties (b) and (c) imply property (a) by using (b), then letting $z = x$ in (c) to conclude $(x, x) \in R$. This is false as shown by the counterexample that $E = \{0, 1\}$ and $R = \{(1, 1)\}$, which satisfies properties (b) and (c), but not property (a). R fails property (a) because $0 \in E$, but $(0, 0) \notin R$ as 0 is not in any ordered pair in R.

A *function* (or *map* or *mapping*) *f from a set A to a set B* (denoted by $f : A \to B$) is a method of assigning to every $a \in A$ exactly one $b \in B$. This b is denoted by $f(a)$ and is called the *value* of f at a. Thus, a function must be *well-defined* in the sense that if $a = a'$, then $f(a) = f(a')$. The set A is called the *domain* of f (denoted by dom f) and the set B is called the *codomain* of f (denoted by codom f). We say f is a B-valued function (eg. if $B = \mathbb{R}$, then we say f is a real-valued function.) When the codomain B is not emphasized, then we may simply say f is a function on A. The *image* or *range* of f (denoted by $f(A)$ or im f or ran f) is the set $\{f(x) : x \in A\}$. (To emphasize this is a subset of B, we also write it as $\{f(x) \in B : x \in A\}$.) The set $G = \{\big(x, f(x)\big) : x \in A\}$ is called the *graph* of f. Two functions are *equal* if and only if they have the same graphs. In particular, the domains of equal functions are the same set.

Examples. The function $f : \mathbb{Z} \to \mathbb{R}$ given by $f(x) = x^2$ has dom $f = \mathbb{Z}$, codom $f = \mathbb{R}$. Also, ran $f = \{0, 1, 4, 9, 16, \dots\}$. This is different from the function $g : \mathbb{R} \to \mathbb{R}$ given by $g(x) = x^2$ because dom $g = \mathbb{R} \neq$ dom f. Also, a function may have more than one parts in its definition, eg. the absolute value function $h : \mathbb{R} \to \mathbb{R}$ defined by $h(x) = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases}$. Be careful in defining functions. The following is bad: let $x_n = (-1)^n$ and $i(x_n) = n$. The rule is not well-defined because $x_1 = -1 = x_3$, but $i(x_1) = 1 \neq 3 = i(x_3)$.

Definitions. (i) The *identity function* on a set S is $I_S : S \to S$ given by $I_S(x) = x$ for all $x \in S$.

(ii) Let $f : A \to B, g : B' \to C$ be functions and $f(A) \subseteq B'$. The *composition* of g by f is the function $g \circ f : A \to C$ defined by $(g \circ f)(x) = g(f(x))$ for all $x \in A$.

(iii) Let $f : A \to B$ be a function and $C \subseteq A$. The function $f|_C : C \to B$ defined by $f|_C(x) = f(x)$ for every $x \in C$ is called the *restriction* of f to C.

(iv) A function $f : A \to B$ is *surjective* (or *onto*) iff $f(A) = B$.

(v) A function $f : A \to B$ is *injective* (or *one-to-one*) iff $f(x) = f(y)$ implies $x = y$.

(vi) A function $f : A \to B$ is a *bijection* (or a *one-to-one correspondence*) iff it is injective and surjective.

(vii) For an injective function $f : A \to B$, the *inverse function* of f is the function $f^{-1} : f(A) \to A$ defined by $f^{-1}(y) = x \iff f(x) = y$.

Remarks. A function $f : A \to B$ is surjective means $f(A) = B$, which is the same as saying every $b \in B$ is an $f(a)$ for at least one $a \in A$. In this sense, the values of f do not omit anything in B. We will loosely say *f does not omit any element of B* for convenience. However, there may possibly be more than one $a \in A$ that are assigned the same $b \in B$. Hence, the range of f may *repeat* some elements of B. If A and B are finite sets, then f surjective implies the number of elements in A is greater than or equal to the number of elements in B.

Next, a function $f : A \to B$ is injective means, in the contrapositive sense, that $x \neq y$ implies $f(x) \neq f(y)$, which we may loosely say *f does not repeat any element of B*. However, f may omit elements of B as there may possibly be elements in B that are not in the range of f. So if A and B are finite sets, then f injective implies the number of elements in A is less than or equal to the number of elements in B.

Therefore, a bijection from A to B is a function whose values do not omit nor repeat any element of B. If A and B are finite sets, then f bijective implies the number of elements in A and B are the same.

Remarks (Exercises). (a) Let $f : A \to B$ be a function. We have f is a bijection if and only if there is a function $g : B \to A$ such that $g \circ f = I_A$ and $f \circ g = I_B$. (In fact, for f bijective, we have $g = f^{-1}$ is bijective.)

(b) If $f : A \to B$ and $h : B \to C$ are bijections, then $h \circ f : A \to C$ is a bijection.

(c) Let A, B be subsets of $\mathbb{R}$ and $f : A \to B$ be a function. If for every $b \in B$, the horizontal line $y = b$ intersects the graph of f exactly once, then f is a bijection.

Example. Show that $f : [0, 1] \to [3, 4]$ defined by $f(x) = x^3 + 3$ is a bijection.

Method 1 If $f(x) = f(y)$, then $x^3 + 3 = y^3 + 3$, which implies $x^3 = y^3$. Taking cube roots of both sides, we get $x = y$. Hence f is injective. Next, for every $y \in [3, 4]$, solving the equation $x^3 + 3 = y$ for x, we get $x = \sqrt[3]{y - 3}$. Since $y \in [3, 4]$ implies $y - 3 \in [0, 1]$, we see $x \in [0, 1]$. Then $f(x) = y$. Hence f is surjective. Therefore, f is bijective.

Method 2. Define $g : [3, 4] \to [0, 1]$ by $g(y) = \sqrt[3]{y - 3}$. For every $x \in [0, 1]$ and $y \in [3, 4]$, $(g \circ f)(x) = g(x^3+3) = \sqrt[3]{(x^3 + 3) - 3} = x$ and $(f \circ g)(y) = f(\sqrt[3]{y - 3}) = (\sqrt[3]{y - 3})^3 + 3 = y$. By remark (a) above, f is a bijection.

To deal with the number of elements in a set, we introduce the following concept. For sets S_1 and S_2, we will define $S_1 \sim S_2$ and say they have the *same cardinality* (or the *same cardinal number*) if and only if there exists a bijection from S_1 to S_2. This is easily checked to be an equivalence relation on the collection of all sets. For a set S, the equivalence class $[S]$ is often called the *cardinal number* of S and is denoted by card S or $|S|$. This is a way to assign a symbol for the number of elements in a set. It is common to denote card $\emptyset = 0$, , for a positive integer n, card $\{1, 2, \dots, n\} = n$, card $\mathbb{N} = \aleph_0$ (read *aleph-naught*) and card $\mathbb{R} = c$ (often called *the cardinality of the continuum*).

Chapter 3. Countability

Often we compare two sets to see if they are different. In case both are infinite sets, then the concept of countable sets may help to distinguish these infinite sets.

Definitions. A set S is *countably infinite* iff there exists a bijection $f : \mathbb{N} \to S$ (i.e. $\mathbb{N}$ and S have the same cardinal number $\aleph_0$.) A set is *countable* iff it is a finite or countably infinite set. A set is *uncountable* iff it is not countable.

Remarks. Suppose $f\colon \mathbb{N} \to S$ is a bijection. Then f is injective means $f(1), f(2), f(3), \ldots$ are all distinct and f is surjective means $\{f(1), f(2), f(3), \ldots\} = S$. So $n \in \mathbb{N} \leftrightarrow f(n) \in S$ is a one-to-one correspondence between $\mathbb{N}$ and S. Therefore, the elements of S can be listed in an "orderly" way (as $f(1), f(2), f(3), \ldots$) without repetition or omission. Conversely, if the elements of S can be listed as $s_1, s_2, \ldots$ without repetition or omission, then $f : \mathbb{N} \to S$ defined by $f(n) = s_n$ will be a bijection as no repetition implies injectivity and no omission implies surjectivity.

Bijection Theorem. *Let $g\colon S \to T$ be a bijection. S is countable if and only if T is countable.*

(**Reasons.** The finite set case is clear. For infinite sets, it is true because S countable implies there is a bijective function $f : \mathbb{N} \to S$, which implies $h = g \circ f : \mathbb{N} \to T$ is bijective, i.e. T is countable. For the converse, h is bijective implies $f = g^{-1} \circ h$ is bijective.)

Remarks. Similarly, taking contrapositive, S is uncountable if and only if T is uncountable.

Basic Examples. (1) $\mathbb{N}$ is countably infinite (because the identity function $I_{\mathbb{N}}(n) = n$ is a bijection).

(2) $\mathbb{Z}$ is countably infinite because the following function is a bijection (one-to-one correspondence):

$$\begin{array}{ccccccccccccc} \mathbb{N} & = & \{ & 1, & 2, & 3, & 4, & 5, & 6, & 7, & 8, & 9, & \ldots \ \} \\ f\ \downarrow & & & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \ldots \\ \mathbb{Z} & = & \{ & 0, & 1, & -1, & 2, & -2, & 3, & -3, & 4, & -4, & \ldots \ \}. \end{array}$$

The function $f : \mathbb{N} \to \mathbb{Z}$ is given by $f(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ -(\frac{n-1}{2}) & \text{if } n \text{ is odd} \end{cases}$ and its inverse function $g : \mathbb{Z} \to \mathbb{N}$ is given by $g(m) = \begin{cases} 2m & \text{if } m > 0 \\ 1 - 2m & \text{if } m \leq 0 \end{cases}$. Just check $g \circ f = I_{\mathbb{N}}$ and $f \circ g = I_{\mathbb{Z}}$.

(3) $\mathbb{N} \times \mathbb{N} = \{(m, n)\colon m, n \in \mathbb{N}\}$ is countably infinite.

(Diagonal Counting Scheme) Using the diagram on the right, define $f\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ by $f(1) = (1, 1)$, $f(2) = (2, 1)$, $f(3) = (1, 2)$, $f(4) = (3, 1)$, $f(5) = (2, 2)$, $f(6) = (1, 3)$, ..., then f is injective because no ordered pair is repeated. Also, f is surjective because

$$(m, n) = f\left(\sum_{k=0}^{m+n-2} k + n\right) = f\left(\frac{(m+n-2)(m+n-1)}{2} + n\right).$$

$$\begin{array}{cccc} (1,1) & (1,2) & (1,3) & (1,4) \\ (2,1) & (2,2) & (2,3) & \vdots \\ (3,1) & (3,2) & \vdots & \vdots \\ (4,1) & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{array}$$

(4) The open interval $(0, 1) = \{x\colon\ x \in \mathbb{R}\ \text{ and }\ 0 < x < 1\}$ is uncountable. Also, $\mathbb{R}$ is uncountable.

$$\begin{array}{rcl} f(1) & = & 0.a_{11}a_{12}a_{13}a_{14}\ldots \\ f(2) & = & 0.a_{21}a_{22}a_{23}a_{24}\ldots \\ f(3) & = & 0.a_{31}a_{32}a_{33}a_{34}\ldots \\ f(4) & = & 0.a_{41}a_{42}a_{43}a_{44}\ldots \\ \vdots & & \vdots \end{array}$$

Suppose $(0, 1)$ is countably infinite and $f\colon \mathbb{N} \to (0, 1)$ is a bijection as shown on the left. Consider the number x whose decimal representation is $0.b_1b_2b_3b_4\ldots$, where $b_n = \begin{cases} 2 & \text{if } a_{nn} = 1 \\ 1 & \text{if } a_{nn} \neq 1 \end{cases}$. Then $0 < x < 1$ and $x \neq f(n)$ for all n because $b_n \neq a_{nn}$. So f cannot be surjective, a contradiction. Next $\mathbb{R}$ is uncountable because $\tan \pi(x - \frac{1}{2})$ provides a bijection from $(0, 1)$ onto $\mathbb{R}$.

To determine the countability of more complicated sets, we will need the theorems below.

Countable Subset Theorem. *Let $A \subseteq B$. If B is countable, then A is countable. (Taking contrapositive, if A is uncountable, then B is uncountable.)*

Countable Union Theorem. *If A_n is countable for every $n \in \mathbb{N}$, then $\bigcup_{n\in\mathbb{N}} A_n$ is countable. In general, if S is countable (say $f : \mathbb{N} \to S$ is a bijection) and A_s is countable for every $s \in S$, then $\bigcup_{s\in S} A_s = \bigcup_{n\in\mathbb{N}} A_{f(n)}$ is countable. (Briefly, countable union of countable sets is countable.)*

Product Theorem. *If A, B are countable, then $A \times B = \{(a, b)\colon\ a \in A, b \in B\}$ is countable. In fact, if $A_1, A_2, \ldots, A_n$ are countable, then $A_1 \times A_2 \times \cdots \times A_n$ is countable (by mathematical induction).*

(**Sketch of Reasons.** For the countable subset theorem, if B is countable, then we can list the elements of B and to count the elements of A, we can skip over those elements of B that are not in A. For the countable union theorem, if we list the elements of A_1 in the first row, the elements of A_2 in the second row, ..., then we can count all the elements by using the diagonal counting scheme. As for the product theorem, we can imitate the example of $\mathbb{N} \times \mathbb{N}$ and also use the diagonal counting scheme.)

Examples. (5) $\mathbb{Q} = \bigcup_{n=1}^{\infty} S_n$, where $S_n = \left\{\dfrac{m}{n} : m \in \mathbb{Z}\right\}$. For every $n \in \mathbb{N}$, the function $f_n : \mathbb{Z} \to S_n$ given by $f_n(m) = \dfrac{m}{n}$ is a bijection (with $f_n^{-1}\left(\frac{m}{n}\right) = m$), so S_n is countable by the bijection theorem. Therefore, $\mathbb{Q}$ is countable by the countable union theorem. (Then subsets of $\mathbb{Q}$ like $\mathbb{Z} \setminus \{0\}$, $\mathbb{N} \cup \{0\}$, $\mathbb{Q} \cap (0, 1)$ are also countable.)

(6) $\mathbb{R} \setminus \mathbb{Q}$ is uncountable. (In fact, if A is uncountable and B is countable, then $A \setminus B$ is uncountable as $A \setminus B$ countable implies $(A \cap B) \cup (A \setminus B) = A$ countable by the countable union theorem, which is a contradiction).

(7) $\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$ contains $\mathbb{R}$ and $\mathbb{R}$ is uncountable, so by the countable subset theorem, $\mathbb{C}$ is uncountable.

(8) Show that the set $A = \{r\sqrt{m} : m \in \mathbb{N}, r \in (0, 1)\}$ is uncountable, but the set $B = \{r\sqrt{m} : m \in \mathbb{N}, r \in \mathbb{Q}\cap(0, 1)\}$ is countable.

Solution. Taking $m = 1$, we see that $(0, 1) \subseteq A$. Since $(0, 1)$ is uncountable, A is uncountable. Next we will observe that $B = \bigcup_{m\in\mathbb{N}} B_m$, where $B_m = \{r\sqrt{m} : r \in \mathbb{Q} \cap (0, 1)\} = \bigcup_{r\in\mathbb{Q}\cap(0,1)} \{r\sqrt{m}\}$ for each $m \in \mathbb{N}$. Since $\mathbb{Q}\cap(0, 1)$ is countable and $\{r\sqrt{m}\}$ has 1 element for every $r \in \mathbb{Q}\cap(0, 1)$, B_m is countable by the countable union theorem. Finally, since $\mathbb{N}$ is countable and B_m is countable for every $m \in \mathbb{N}$, B is countable by the countable union theorem.

(9) Show that the set L of all lines with equation $y = mx + b$, where $m, b \in \mathbb{Q}$, is countable.

Solution. Note that for each pair m, b of rational numbers, there is a unique line $y = mx + b$ in the set L. So the function $f : \mathbb{Q} \times \mathbb{Q} \to L$ defined by letting $f(m, b)$ be the line $y = mx + b$ (with f^{-1} sending the line back to (m, b)) is a bijection. Since $\mathbb{Q} \times \mathbb{Q}$ is countable by the product theorem, so the set L is countable by the bijection theorem.

(10) Show that if $A_n = \{0, 1\}$ for every $n \in \mathbb{N}$, then $A_1 \times A_2 \times A_3 \times \cdots$ is uncountable. (In particular, this shows that the product theorem is not true for infintely many countable sets.)

Solution. Assume $A_1 \times A_2 \times A_3 \times \cdots = \{(a_1, a_2, a_3, \ldots) : \text{each } a_i = 0 \text{ or } 1\}$ is countable and $f : \mathbb{N} \to A_1 \times A_2 \times A_3 \times \cdots$ is a bijection. Following example (4), we can change the n-th coordinate of $f(n)$ (from 0 to 1 or from 1 to 0) to produce an element of $A_1 \times A_2 \times A_3 \times \cdots$ not equal to any $f(n)$, which is a contradiction. So it must be uncountable.

(11) Show that the power set $P(\mathbb{N})$ of all subsets of $\mathbb{N}$ is uncountable.

Solution. As in example (10), let $A_n = \{0, 1\}$ for every $n \in \mathbb{N}$. Define $g : P(\mathbb{N}) \to A_1 \times A_2 \times A_3 \times \cdots$ by $g(S) = (a_1, a_2, a_3, \ldots)$, where $a_m = \begin{cases} 1 & \text{if } m \in S \\ 0 & \text{if } m \notin S \end{cases}$. (For example, $g(\{1, 3, 5, \ldots\}) = (1, 0, 1, 0, 1, \ldots)$.) Note g has the inverse function $g^{-1}\big((a_1, a_2, a_3, \ldots)\big) = \{m : a_m = 1\}$. Hence g is a bijection. Since $A_1 \times A_2 \times A_3 \times \cdots$ is uncountable, so $P(\mathbb{N})$ is uncountable by the bijection theorem.

(12) Show that the set S of all nonconstant polynomials with integer coefficients is countable.

Solution. For $n \in \mathbb{N}$, the set of S_n of all polynomials of degree n with integer coefficients is countable because the function $f : S_n \to (\mathbb{Z} \setminus \{0\}) \times \mathbb{Z} \times \cdots \times \mathbb{Z}$ defined by $f(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0) = (a_n, a_{n-1}, \dots, a_0)$ is a bijection and $(\mathbb{Z} \setminus \{0\}) \times \mathbb{Z} \times \cdots \times \mathbb{Z}$ is countable by the product theorem. So, $S = \bigcup_{n\in\mathbb{N}} S_n$ is countable by the countable union theorem.

(13) Show that there exists a real number, which is not a root of any nonconstant polynomial with integer coefficients.

Solution. For every nonconstant polynomial f with integer coefficients, let R_f denotes the set of roots of f. Then R_f has at most $(\deg f)$ elements, hence R_f is countable. Let S be the set of all nonconstant polynomials with integer coefficients, which is countable by the last example. Then $\bigcup_{f\in S} R_f$ is the set of all roots of nonconstant polynomials with integer coefficients. It is countable by the countable union theorem. Since $\mathbb{R}$ is uncountable, $\mathbb{R} \setminus \bigcup_{f\in S} R_f$ is uncountable by the fact in example (6). So there exist uncountably many real numbers, which are not roots of any nonconstant polynomial with integer coefficients.

Remarks. Any number which is a root of a nonconstant polynomial with integer coefficients is called an *algebraic* number. A number which is not a root of any nonconstant polynomial with integer coefficients is called a *transcendental* number. transcendental numbers? If so, are there finitely many or countably many such numbers? Since every rational number $\frac{a}{b}$ is the root of the polynomial $bx - a$, every rational number is algebraic. There are irrational numbers like $\pm\sqrt{2}$, which are algebraic because they are the roots of $x^2 - 2$. Using the identity $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$, the irrational number $\cos 20^\circ$ is easily seen to be algebraic as it is a root of $8x^3 - 6x - 1$. Example (13) showed that there are only countably many algebraic numbers and there are uncountably many transcendental real numbers. In a number theory course, it will be shown that π and e are transcendental.

Theorem.

(1) (*Injection Theorem*) *Let $f : A \to B$ be injective. If B is countable, then A is countable. (Taking contrapositive, if A is uncountable, then B is uncountable.)*

(2) (*Surjection Theorem*) *Let $g : A \to B$ be surjective. If A is countable, then B is countable. (Taking contrapositive, if B is uncountable, then A is uncountable.)*

(**Reasons.** For the first statement, observe that the function $h : A \to f(A)$ defined by $h(x) = f(x)$ is injective (because f is injective) and surjective (because $h(A) = f(A)$). So h is a bijection. If B is countable, then $f(A)$ is countable by the countable subset theorem, which implies A is countable by the bijection theorem.

For the second statement, observe that $B = g(A) = \bigcup_{x\in A}\{g(x)\}$. If A is countable, then it is a countable union of countable sets. By the countable union theorem, B is countable.)

Examples. (14) Show $\mathbb{Q}$ is countable by using the injection theorem.

Solution. Define $f : \mathbb{Q} \to \mathbb{Z} \times \mathbb{N}$ by $f(x) = (m, n)$, where m/n is the reduced fraction form of x. Then f is injective because $f(x) = f(x') = (m, n)$ implies $x = m/n = x'$. Since $\mathbb{Z} \times \mathbb{N}$ is countable by product theorem, so $\mathbb{Q}$ is countable by the injection theorem.

(15) Let A_1 be uncountable and $A_2, \dots, A_n$ be nonempty sets. Show that $A_1 \times A_2 \times \cdots \times A_n$ is uncountable.

Solution. Define $g : A_1 \times A_2 \times \cdots \times A_n \to A_1$ by $g(x_1, x_2, \dots, x_n) = x_1$. Since $A_2, \dots, A_n$ are nonempty, let $a_2 \in A_2, \dots, a_n \in A_n$. Then for every $a_1 \in A_1$, we have $g(a_1, a_2, \dots, a_n) = a_1$ so that g is surjective. Since A_1 is uncountable, by the surjection theorem, $A_1 \times A_2 \times \cdots \times A_n$ is uncountable.

The following is a famous statement in mathematics.

Continuum Hypothesis. If S is uncountable, then there exists at least one injective function $f : \mathbb{R} \to S$, i.e. every uncountable set has at least as many elements as the real numbers.

In 1940, Kurt Gödel showed that the opposite statement would not lead to any contradiction. In 1966, Paul Cohen won the Fields' Medal for showing the statement also would not lead to any contradiction. So proof by contradiction may not be applied to every statement.

Chapter 4. Series

Definitions. A *series* is the summation of a countable set of numbers in a specific order. If there are finitely many numbers, then the series is a *finite series*, otherwise it is an *infinite series*. The numbers are called *terms*. The sum of the first n terms is called the *n-th partial sum of the series*.

An infinite series is of the form $\underbrace{a_1}_{1^{\text{st}}\text{ term}} + \underbrace{a_2}_{2^{\text{nd}}\text{ term}} + \underbrace{a_3}_{3^{\text{rd}}\text{ term}} + \ldots$ or we may write it as $\sum_{k=1}^{\infty} a_k$.
The first partial sum is $S_1 = a_1$. The second partial sum is $S_2 = a_1 + a_2$. The n^{th} partial sum is $S_n = a_1 + a_2 + \ldots + a_n$.

Series are used frequently in science and engineering to solve problems or approximate solutions. (*E.g.* trigonometric or logarithm tables were computed using series in the old days.)

Examples.(1) $1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots = ?$ ($S_n = 1+\frac{1}{2}+\frac{1}{4}+\ldots+\frac{1}{2^n} = 2-\frac{1}{2^n}$, $1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\ldots = \lim_{n\to\infty}(2-\frac{1}{2^n}) = 2$.)
We say the series *converges* to 2, which is called the *sum* of the series.

(2) $1+1+1+1+1+1+\ldots = \infty$ ($S_n = \underbrace{1+1+\ldots+1}_{n} = n$, $\lim_{n\to\infty} S_n = \infty$.) We say the series *diverges* (to ∞).

(3) $1-1+1-1+1-1+1-1+\ldots$. ($S_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}$, $\lim_{n\to\infty} S_n$ doesn't exist.) We say the series *diverges*.

Definitions. A series $\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + \ldots$ *converges* to a number S iff $\lim_{n\to\infty}(a_1 + a_2 + \ldots + a_n) = \lim_{n\to\infty} S_n = S$. In that case, we may write $\sum_{k=1}^{\infty} a_k = S$ and say S is the *sum* of the series. A series *diverges* to ∞ iff the partial sum S_n tends to infinity as n tends to infinity. A series *diverges* iff it does not converge to any number.

Remarks. (1) For every series $\sum_{k=1}^{\infty} a_k$, there is a sequence (of partial sums) $\{S_n\}$. Conversely, if the partial sum sequence $\{S_n\}$ is given, we can find the terms a_n as follows: $a_1 = S_1$, $a_2 = S_2 - S_1$, ..., $a_k = S_k - S_{k-1}$ for $k > 1$. Then $a_1 + \ldots + a_n = S_1 + (S_2 - S_1) + \ldots + (S_n - S_{n-1}) = S_n$. So $\{S_n\}$ is the partial sum sequence of $\sum_{k=1}^{\infty} a_k$. Conceptually, series and sequences are equivalent. So to study series, we can use facts about sequences.

(2) Let N be a positive integer. $\sum_{k=1}^{\infty} a_k$ converges to A if and only if $\sum_{k=N}^{\infty} a_k$ converges to $B = A - (a_1 + \cdots + a_{N-1})$ because

$$B = \lim_{n\to\infty}(a_N + \cdots + a_n) = \lim_{n\to\infty}(a_1 + a_2 + \cdots + a_n) - (a_1 + \cdots + a_{N-1}) = A - (a_1 + \cdots + a_{N-1}).$$

So to see if a series converges, we may ignore finitely many terms.

Theorem. *If* $\sum_{k=1}^{\infty} a_k$ *converges to* A *and* $\sum_{k=1}^{\infty} b_k$ *converges to* B, *then*

$$\sum_{k=1}^{\infty}(a_k + b_k) = A + B = \sum_{k=1}^{\infty} a_k + \sum_{k=1}^{\infty} b_k, \quad \sum_{k=1}^{\infty}(a_k - b_k) = A - B = \sum_{k=1}^{\infty} a_k - \sum_{k=1}^{\infty} b_k, \quad \sum_{k=1}^{\infty} ca_k = cA = c\sum_{k=1}^{\infty} a_k$$

for any constant c.

For simple series such as geometric or telescoping series, we can find their sums.

Theorem (Geometric Series Test). *We have*

$$\sum_{k=0}^{\infty} r^k = \lim_{n\to\infty}(1+r+r^2+\ldots+r^n) = \lim_{n\to\infty}\frac{1-r^{n+1}}{1-r} = \begin{cases} \dfrac{1}{1-r} & \textit{if } |r|<1 \\ \text{doesn't exist} & \textit{otherwise} \end{cases}.$$

Example. $0.999\cdots = \frac{9}{10}+\frac{9}{100}+\frac{9}{1000}+\cdots = \frac{9}{10}(\frac{1}{1-\frac{1}{10}}) = 1 = 1.000\cdots$. So, this shows that the number 1 has two decimal representations!

Theorem (Telescoping Series Test). *We have* $\sum_{k=1}^{\infty}(b_k-b_{k+1}) = \lim_{n\to\infty}\big((b_1-b_2)+(b_2-b_3)+\cdots+(b_n-b_{n+1})\big)$ $= \lim_{n\to\infty}(b_1-b_{n+1}) = b_1 - \lim_{n\to\infty} b_{n+1}$ *converges if and only if* $\lim_{n\to\infty} b_n$ *is a number.*

Examples. (1) $\sum_{k=1}^{\infty}\frac{1}{k(k+1)} = \sum_{k=1}^{\infty}(\frac{1}{k}-\frac{1}{k+1}) = \big(1-\frac{1}{2}\big)+\big(\frac{1}{2}-\frac{1}{3}\big)+\big(\frac{1}{3}-\frac{1}{4}\big)+\cdots = 1-\lim_{n\to\infty}\frac{1}{n+1} = 1.$

(2) $\sum_{k=1}^{\infty}(5^{1/k}-5^{1/(k+1)}) = (5-\sqrt{5})+(\sqrt{5}-\sqrt[3]{5})+\cdots = 5-\lim_{k\to\infty}5^{1/(k+1)} = 5-5^0 = 4.$

If a series is not geometric or telescoping, we can only determine if it converges or diverges. This can be done most of the time by applying some standard tests. If the series converges, it may be extremely difficult to find the sum!

Theorem (Term Test). *If* $\sum_{k=1}^{\infty} a_k$ *converges, then* $\lim_{k\to\infty} a_k = 0$. *(If* $\lim_{k\to\infty} a_k \neq 0$, *then the series* $\sum_{k=1}^{\infty} a_k$ *diverges.) If* $\lim_{k\to\infty} a_k = 0$, *the series* $\sum_{k=1}^{\infty} a_k$ *may or may not converge.*

(**Reason.** Suppose $\sum_{k=1}^{\infty} a_k$ converges to S. Then $\lim_{n\to\infty} S_n = S$ and $\lim_{k\to\infty} a_k = \lim_{k\to\infty}(S_k - S_{k-1}) = S - S = 0$.)

Term test is only good for series that are suspected to be divergent!

Examples. (1) $1+1+1+1+\ldots$. Here $a_k = 1$ for all k, so $\lim_{k\to\infty} a_k = 1$. Series diverges.

(2) $\sum_{k=1}^{\infty}\cos(\frac{1}{k}) = \cos 1 + \cos\frac{1}{2} + \cos\frac{1}{3} + \ldots$ diverges because $\lim_{k\to\infty}\cos(\frac{1}{k}) = \cos 0 = 1 \neq 0$.

(3) $\sum_{k=1}^{\infty}\cos k = \cos 1 + \cos 2 + \cos 3 + \ldots$ diverges because $\lim_{k\to\infty}\cos k \neq 0$. (Otherwise, $\lim_{k\to\infty}\cos k = 0$. Then $\lim_{k\to\infty}|\sin k| = \lim_{k\to\infty}\sqrt{1-\cos^2 k} = 1$ and $0 = \lim_{k\to\infty}|\cos(k+1)| = \lim_{k\to\infty}|\cos k\cos 1 - \sin k \sin 1| = \sin 1 \neq 0$, a contradiction.)

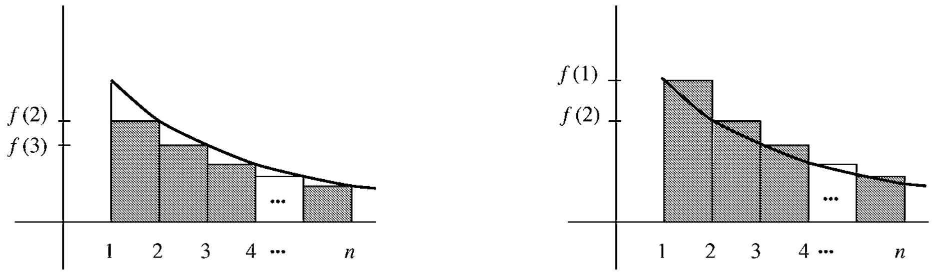
(4) $1-\frac{1}{2}+\frac{1}{4}-\frac{1}{8}+\ldots$. Here $a_k = (-\frac{1}{2})^{k-1}$ for all k, so $\lim_{k\to\infty} a_k = 0$. (Term test doesn't apply!) Series converges by the geometric series test.

(5) $1+\underbrace{\frac{1}{2}+\frac{1}{2}}_{2\text{ times}}+\underbrace{\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}}_{4\text{ times}}+\underbrace{\frac{1}{8}+\ldots+\frac{1}{8}}_{8\text{ times}}+\ldots$. We have $\lim_{k\to\infty} a_k = 0$. (Term test doesn't apply.) Series diverges to ∞ because $S_1 \leq S_2 \leq S_3 \leq \cdots$ and $S_{2^n-1} = n$ has limit ∞.

For a *nonnegative* series $\sum_{k=1}^{\infty} a_k$ (i.e. $a_k \geq 0$ for every k), we have $S_1 \leq S_2 \leq S_3 \leq \ldots$ and $\lim_{n\to\infty} S_n$ must exist as a number or equal to $+\infty$. So either $\sum_{k=1}^{\infty} a_k$ converges to a number or $\sum_{k=1}^{\infty} a_k$ diverges to $+\infty$. (In short, either $\sum_{k=1}^{\infty} a_k = S$ or $\sum_{k=1}^{\infty} a_k = +\infty$.) For nonnegative series, we have the following tests.

Theorem (Integral Test). *Let* $f : [1, +\infty) \to \mathbb{R}$ *decrease to* 0 *as* $x \to +\infty$. *Then* $\sum_{k=1}^{\infty} f(k)$ *converges if and only if* $\int_1^{\infty} f(x)\,dx < \infty$. *(Note in general,* $\sum_{k=1}^{\infty} f(k) \neq \int_1^{\infty} f(x)\,dx$.*)*

(**Reason.** This follows from $f(2) + f(3) + \ldots + f(n) + \cdots \leq \int_1^{\infty} f(x)dx \leq f(1) + f(2) + \ldots + f(n-1) + \cdots$ as shown in the figures below.)



Examples. (1) Consider the convergence or divergence of $\sum_{k=1}^{\infty} \frac{1}{1+k^2}$.

As $x \nearrow \infty$, $1 + x^2 \nearrow \infty$, so $\frac{1}{1+x^2} \searrow 0$. Now $\int_1^{\infty} \frac{1}{1+x^2}dx = \arctan x\Big|_1^{\infty} = \frac{\pi}{2} - \frac{\pi}{4} < \infty$. So $\sum_{k=1}^{\infty} \frac{1}{1+k^2}$ converges.

(2) Consider the convergence or divergence of $\sum_{k=2}^{\infty} \frac{1}{k\ln k}$ and $\sum_{k=2}^{\infty} \frac{1}{k(\ln k)^2}$.

As $x \nearrow \infty$, $x\ln x$ and $x(\ln x)^2 \nearrow \infty$, so their reciprocals decrease to 0. Now $\int_2^{\infty} \frac{dx}{x\ln x} = \ln(\ln x)\Big|_2^{\infty} = \infty$. So $\sum_{k=2}^{\infty} \frac{1}{k\ln k}$ diverges. Next $\int_2^{\infty} \frac{dx}{x(\ln x)^2} = -\frac{1}{\ln x}\Big|_2^{\infty} = \frac{1}{\ln 2} < \infty$. So $\sum_{k=2}^{\infty} \frac{1}{k(\ln k)^2}$ converges.

Theorem (p-test). *For a real number* p, $\zeta(p) = \sum_{k=1}^{\infty} \frac{1}{k^p} = 1 + \frac{1}{2^p} + \frac{1}{3^p} + \frac{1}{4^p} + \ldots$ *converges if and only if* $p > 1$.

(**Reason.** For $p \leq 0$, the terms are at least 1, so the series diverges by term test. For $p > 0$, $f(x) = \frac{1}{x^p}$ decreases to 0 as $x \to +\infty$. Since $\int_1^{\infty} \frac{1}{x^p}\,dx = \frac{x^{-p+1}}{-p+1}\Big|_1^{\infty} = \frac{1}{p-1}$ if $p > 1$, $\int_1^{\infty} \frac{1}{x^p}\,dx = (\ln x)|_1^{\infty} = \infty$ if $p = 1$ and $\int_1^{\infty} \frac{1}{x^p}\,dx = \frac{x^{-p+1}}{-p+1}\Big|_1^{\infty} = \infty$ if $p < 1$, the integral test gives the conclusion.)

Remarks. For even positive integer p, the value of $\zeta(p)$ was computed by Euler back in 1736. He got

$$\zeta(2) = \frac{\pi^2}{6}, \quad \zeta(4) = \frac{\pi^4}{90}, \quad \ldots, \quad \zeta(2n) = (-1)^{n+1}\frac{(2\pi)^{2n}B_{2n}}{2(2n)!}, \quad \ldots,$$

where $B_0 = 1$ and $(k+1)B_k = -\sum_{m=0}^{k-1} \binom{k+1}{m} B_m$ for $k \geq 1$. The values of $\zeta(3), \zeta(5), \ldots$ are unknown. Only in the 1980's, R. Apery was able to show $\zeta(3)$ was irrational.

Theorem (Comparison Test). *Given $v_k \geq u_k \geq 0$ for every k. If $\sum_{k=1}^{\infty} v_k$ converges, then $\sum_{k=1}^{\infty} u_k$ converges. If $\sum_{k=1}^{\infty} u_k$ diverges, then $\sum_{k=1}^{\infty} v_k$ diverges.*

(**Reason.** $v_k \geq u_k \geq 0 \Rightarrow \sum_{k=1}^{\infty} v_k \geq \sum_{k=1}^{\infty} u_k \geq 0$. If $\sum_{k=1}^{\infty} v_k$ is a number, then $\sum_{k=1}^{\infty} u_k$ is a number. If $\sum_{k=1}^{\infty} u_k = +\infty$, then $\sum_{k=1}^{\infty} v_k = +\infty$.)

Theorem (Limit Comparison Test). *Given $u_k, v_k > 0$ for every k. If $\lim_{k\to\infty} \frac{v_k}{u_k}$ is a positive number L, then either (both $\sum_{k=1}^{\infty} u_k$ and $\sum_{k=1}^{\infty} v_k$ converge) or (both diverge to $+\infty$). If $\lim_{k\to\infty} \frac{v_k}{u_k} = 0$, then $\sum_{k=1}^{\infty} u_k$ converges $\Rightarrow \sum_{k=1}^{\infty} v_k$ converges. If $\lim_{k\to\infty} \frac{v_k}{u_k} = \infty$, then $\sum_{k=1}^{\infty} u_k$ diverges $\Rightarrow \sum_{k=1}^{\infty} v_k$ diverges.*

(**Sketch of Reason.** For k large, $\frac{v_k}{u_k} \approx L$. For $L > 0$, $\sum v_k \approx \sum L u_k = L \sum u_k$. If one series converges, then the other also converges. If one diverges (to $+\infty$), so does the other. For $L = 0$, $v_k < u_k$ eventually. For $L = \infty$, $v_k > u_k$ eventually. So the last two statements follow from the comparison test.)

Examples. Consider the convergence or divergence of the following series:

(1) $\sum_{k=1}^{\infty} \frac{1}{k^2}\cos\left(\frac{1}{k}\right)$ (2) $\sum_{k=2}^{\infty} \frac{3^k}{k^2-1}$ (3) $\sum_{k=1}^{\infty} \frac{\sqrt{k+1}}{k^2+5k}$ (4) $\sum_{k=1}^{\infty} \sin\left(\frac{1}{k}\right)$.

Solutions. (1) Since $0 < \frac{1}{k^2}\cos\left(\frac{1}{k}\right) < \frac{1}{k^2}$ and $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges by p-test, $\sum_{k=1}^{\infty} \frac{1}{k^2}\cos\left(\frac{1}{k}\right)$ converges.

(2) Since $0 < \left(\frac{3}{2}\right)^k \leq \frac{3^k}{k^2-1}$ for $k \geq 2$ and $\sum_{k=2}^{\infty} \left(\frac{3}{2}\right)^k$ diverges by the geometric series test, $\sum_{k=2}^{\infty} \frac{3^k}{k^2-1}$ diverges.

(3) When k is large, $\frac{\sqrt{k+1}}{k^2+5k} \approx \frac{\sqrt{k}}{k^2} = \frac{1}{k^{3/2}}$. We compute $\lim_{k\to\infty} \frac{\frac{\sqrt{k+1}}{k^2+5k}}{\frac{\sqrt{k}}{k^2}} = \lim_{k\to\infty} \sqrt{\frac{k+1}{k}} \frac{k^2}{k^2+5k} = 1$. Since $\sum_{k=1}^{\infty} \frac{1}{k^{3/2}}$ converges by p-test, $\sum_{k=1}^{\infty} \frac{\sqrt{k+1}}{k^2+5k}$ converges by the limit comparison test.

(4) When k is large, $\frac{1}{k}$ is close to 0, so $\sin\left(\frac{1}{k}\right)$ is close to $\frac{1}{k}$ because $\lim_{\theta\to 0} \frac{\sin\theta}{\theta} = 1$ (i.e. $\sin\theta \approx \theta$ as $\theta \to 0$). We compute $\lim_{k\to\infty} \frac{\sin(\frac{1}{k})}{\frac{1}{k}} = \lim_{\theta\to 0} \frac{\sin\theta}{\theta} = 1$. Since $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges by p-test, $\sum_{k=1}^{\infty} \sin\left(\frac{1}{k}\right)$ diverges by the limit comparison test.

For series with *alternate* positive and negative terms, we have the following test.

Theorem (Alternating Series Test). *If c_k decreases to 0 as $k \to \infty$ (i.e. $c_1 \geq c_2 \geq c_3 \geq \ldots \geq 0$ and $\lim_{k\to\infty} c_k = 0$), then $\sum_{k=1}^{\infty} (-1)^{k+1} c_k = c_1 - c_2 + c_3 - c_4 + c_5 - \ldots$ converges.*

(**Reason.** Since $c_1 \geq c_2 \geq c_3 \geq \ldots \geq 0$, we have $0 \leq S_2 \leq S_4 \leq S_6 \leq \ldots \leq S_5 \leq S_3 \leq S_1$. Since $\lim_{n\to\infty} |S_n - S_{n-1}| = \lim_{n\to\infty} c_n = 0$, the distances between the partial sums decrease to 0 and so $\lim_{n\to\infty} S_n$ must exist.)

Examples. Both $\sum_{k=2}^{\infty} \frac{(-1)^k}{k \ln k}$ and $\sum_{k=1}^{\infty} e^{-k} \cos k\pi$ converge by the alternating series test because as $k \nearrow \infty$, $k \ln k \nearrow \infty$ and $e^k \nearrow \infty$, so $1/(k \ln k) \searrow 0$ and $e^{-k} \searrow 0$ and $\cos k\pi = (-1)^k$.

For series with *arbitrary* positive or negative term, we have the following tests.

Theorem (Absolute Convergence Test). *If* $\sum_{k=1}^{\infty} |a_k|$ *converges, then* $\sum_{k=1}^{\infty} a_k$ *converges.*

(**Reason.** From $-|a_k| \leq a_k \leq |a_k|$, we get $0 \leq a_k + |a_k| \leq 2|a_k|$. Since $\sum_{k=1}^{\infty} 2|a_k|$ converges, so by the comparison test, $\sum_{k=1}^{\infty}(a_k + |a_k|)$ converges. Then $\sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty}(a_k + |a_k|) - \sum_{k=1}^{\infty} |a_k|$ converges.)

Definition. We say $\sum_{k=1}^{\infty} a_k$ *converges absolutely* iff $\sum_{k=1}^{\infty} |a_k|$ converges. We say $\sum_{k=1}^{\infty} a_k$ *converges conditionally* iff $\sum_{k=1}^{\infty} a_k$ converges, but $\sum_{k=1}^{\infty} |a_k|$ *diverges*.

Examples. Determine if the following series converge absolutely or conditionally

(a) $\sum_{k=1}^{\infty} \frac{\cos k}{k^3}$ (b) $\sum_{k=1}^{\infty} \frac{\cos k\pi}{1+k}$.

Solutions. (a) $\sum_{k=1}^{\infty} \left|\frac{\cos k}{k^3}\right| \leq \sum_{k=1}^{\infty} \frac{1}{k^3}$. Since $\sum_{k=1}^{\infty} \frac{1}{k^3}$ converges by p-test, it follows that $\sum_{k=1}^{\infty} \left|\frac{\cos k}{k^3}\right|$ converges by the comparison test. So $\sum_{k=1}^{\infty} \frac{\cos k}{k^3}$ converges absolutely by the absolute convergence test.

(b) $\sum_{k=1}^{\infty} \left|\frac{\cos k\pi}{1+k}\right| = \sum_{k=1}^{\infty} \frac{1}{1+k}$ because $\cos k\pi = (-1)^k$. $\int_1^{\infty} \frac{dx}{1+x} = \ln(1+x)\Big|_1^{\infty} = \infty \Rightarrow \sum_{k=1}^{\infty} \frac{1}{1+k}$ diverges. However, $\frac{1}{1+k}$ decreases to 0 as $k \to +\infty$. So by the alternating series test, $\sum_{k=1}^{\infty} \frac{\cos k\pi}{1+k} = \sum_{k=1}^{\infty}(-1)^k \frac{1}{1+k}$ converges. Therefore $\sum_{k=1}^{\infty} \frac{\cos k\pi}{1+k}$ converges conditionally.

Theorem (Ratio Test). *If* $a_k \neq 0$ *for every* k *and* $\lim_{k\to\infty} |a_{k+1}/a_k|$ *exists, then*

$$\lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| \begin{cases} < 1 & \Rightarrow \sum_{k=1}^{\infty} a_k \textit{ converges absolutely} \\ = 1 & \Rightarrow \sum_{k=1}^{\infty} a_k \textit{ may converge } \left(\textit{e.g. } \sum_{k=1}^{\infty} \frac{1}{k^2}\right) \textit{ or diverge } \left(\textit{e.g. } \sum_{k=1}^{\infty} \frac{1}{k}\right). \\ > 1 & \Rightarrow \sum_{k=1}^{\infty} a_k \textit{ diverges} \end{cases}$$

(**Sketch of reason.** Let $r = \lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right|$, then for k large, $\left|\frac{a_{k+1}}{a_k}\right|, \left|\frac{a_{k+2}}{a_{k+1}}\right|, \ldots, \left|\frac{a_{k+n}}{a_{k+n-1}}\right| \approx r$, so $|a_{k+n}| \approx |a_k|r^n$ and $|a_k| + |a_{k+1}| + |a_{k+2}| + \ldots \approx |a_k|(1 + r + r^2 + r^3 + \ldots)$ which converges if $r < 1$ by the geometric series test and $\sum a_{k+n} \approx \sum \pm a_k r^n$ diverges if $r > 1$ by the term test.)

Theorem (Root Test). *If* $\lim_{k\to\infty} \sqrt[k]{|a_k|}$ *exists, then*

$$\lim_{k\to\infty} \sqrt[k]{|a_k|} \begin{cases} < 1 & \Rightarrow \sum_{k=1}^{\infty} a_k \textit{ converges absolutely} \\ = 1 & \Rightarrow \sum_{k=1}^{\infty} a_k \textit{ may converge } \left(e.g. \sum_{k=1}^{\infty} \frac{1}{k^2}\right) \textit{ or diverge } \left(e.g. \sum_{k=1}^{\infty} \frac{1}{k}\right). \\ > 1 & \Rightarrow \sum_{k=1}^{\infty} a_k \textit{ diverges} \end{cases}$$

(**Sketch of reason.** Let $r = \lim_{k\to\infty} \sqrt[k]{|a_k|}$, then for k large, $\sqrt[k]{|a_k|} \approx r$. So $|a_k| \approx r^k$, $\sum |a_k| \approx \sum r^k$.)

Examples. Consider the convergence or divergence of the following series:

(1) $\sum_{k=1}^{\infty} \frac{1}{3^k - 2^k}$ (2) $\sum_{k=1}^{\infty} \frac{k!}{k^k}$.

Solutions. (1) Since $\lim_{k\to\infty} \frac{\frac{1}{3^{k+1}-2^{k+1}}}{\frac{1}{3^k-2^k}} = \lim_{k\to\infty} \frac{3^k - 2^k}{3^{k+1} - 2^{k+1}} = \lim_{k\to\infty} \frac{\frac{1}{3} - (\frac{2}{3})^k \frac{1}{3}}{1 - (\frac{2}{3})^{k+1}} = \frac{1}{3} < 1$, by the ratio test, $\sum_{k=1}^{\infty} \frac{1}{3^k - 2^k}$ converges. Alternatively, since $\lim_{k\to\infty} \sqrt[k]{\frac{1}{3^k - 2^k}} = \lim_{k\to\infty} \frac{1}{\sqrt[k]{3^k - 2^k}} = \lim_{k\to\infty} \frac{1}{3\sqrt[k]{1 - (\frac{2}{3})^k}} = \frac{1}{3} < 1$, by the root test, $\sum_{k=1}^{\infty} \frac{1}{3^k - 2^k}$ converges.

(2) Since $\lim_{k\to\infty} \frac{(k+1)!}{(k+1)^{k+1}} \frac{k^k}{k!} = \lim_{k\to\infty} \frac{1}{(1 + \frac{1}{k})^k} = \frac{1}{e} < 1$, by the ratio test, $\sum_{k=1}^{\infty} \frac{k!}{k^k}$ converges.

Remarks. You may have observed that in example (1), the limit you got for applying the root test was the same as the limit you got for applying the ratio test. This was not an accident!

Theorem. *If $a_k > 0$ for all k and $\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = r \in \mathbb{R}$, then $\lim_{k\to\infty} \sqrt[k]{a_k} = r$. (This implies that the root test can be applied to more series than the ratio test.)*

Examples. (1) Let $a_k = k$, then $\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = \lim_{k\to\infty} \frac{k+1}{k} = 1$. So, $\lim_{k\to\infty} \sqrt[k]{k} = 1$.

(2) Let $a_k = \frac{k!}{k^k}$, then $\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = \frac{1}{e}$ as above. So $\lim_{k\to\infty} \sqrt[k]{a_k} = \lim_{k\to\infty} \frac{\sqrt[k]{k!}}{k} = \frac{1}{e}$, i.e. when k is large, $k! \approx \left(\frac{k}{e}\right)^k$, which is a simple version of what is called *Stirling's formula.* It is useful for estimating $n!$ when n is large. For example, since $\log_{10} \frac{100}{e} \approx 1.566$, so $\frac{100}{e} \approx 10^{1.566}$, then we get $100! \approx 10^{156.6}$, which has about 157 digits.

Theorem (Summation by Parts). *Let $S_j = \sum_{k=1}^{j} a_k = a_1 + a_2 + \ldots + a_j$ and $\Delta b_k = \frac{b_{k+1} - b_k}{(k+1) - k} = b_{k+1} - b_k$, then*

$$\sum_{k=1}^{n} a_k b_k = S_n b_n - \sum_{k=1}^{n-1} S_k \Delta b_k.$$

(**Reason.** Note $a_1 = S_1$ and $a_k = S_k - S_{k-1}$ for $k > 1$. So,

$$\begin{aligned}\sum_{k=1}^{n} a_k b_k &= S_1 b_1 + (S_2 - S_1) b_2 + \ldots + (S_n - S_{n-1}) b_n \\ &= S_n b_n - S_1(b_2 - b_1) - \ldots - S_{n-1}(b_n - b_{n-1}).)\end{aligned}$$

Example. Show that $\displaystyle\sum_{k=1}^{\infty} \frac{\sin k}{k}$ converges.

Let $a_k = \sin k$ and $b_k = \dfrac{1}{k}$. Using the identity $\sin m \sin \dfrac{1}{2} = \dfrac{1}{2}\left(\cos(m - \dfrac{1}{2}) - \cos(m + \dfrac{1}{2})\right)$, we have

$$S_k = \sin 1 + \sin 2 + \cdots + \sin k = \frac{\cos \frac{1}{2} - \cos(k + \frac{1}{2})}{2 \sin \frac{1}{2}}.$$

This implies $|S_k| \le \dfrac{1}{\sin(1/2)}$ for every k. Applying summation by parts and noting that $\displaystyle\lim_{n\to\infty} \frac{S_n}{n} = 0$, we get

$$\sum_{k=1}^{\infty} \frac{\sin k}{k} = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{\sin k}{k} = \lim_{n\to\infty} \left(\frac{S_n}{n} - \sum_{k=1}^{n-1} S_k \left(\frac{1}{k+1} - \frac{1}{k} \right) \right) = \sum_{k=1}^{\infty} S_k \left(\frac{1}{k} - \frac{1}{k+1} \right).$$

Now $\displaystyle\sum_{k=1}^{\infty} \left| S_k \left(\frac{1}{k} - \frac{1}{k+1} \right) \right| \le \frac{1}{\sin(1/2)} \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{k+1} \right) = \frac{1}{\sin(1/2)}$ by the telescoping series test. So by the absolute convergence test, $\displaystyle\sum_{k=1}^{\infty} \frac{\sin k}{k} = \sum_{k=1}^{\infty} S_k \left(\frac{1}{k} - \frac{1}{k-1} \right)$ converges.

Inserting Parentheses and Rearrangements of Series.

Definition. We say $\displaystyle\sum_{k=1}^{\infty} b_k$ is obtained from $\displaystyle\sum_{k=1}^{\infty} a_k$ by *inserting parentheses* iff there is a strictly increasing function $p : \mathbb{N}\cup\{0\} \to \mathbb{N}\cup\{0\}$ such that $p(0) = 0$, $b_1 = a_1 + \cdots + a_{p(1)}$, $b_2 = a_{p(1)+1} + \cdots + a_{p(2)}$, $b_3 = a_{p(2)+1} + \cdots + a_{p(3)}, \ldots$. (Note b_n is the sum of $k_n = p(n) - p(n-1)$ terms.)

Grouping Theorem. *Let* $\displaystyle\sum_{k=1}^{\infty} b_k$ *be obtained from* $\displaystyle\sum_{k=1}^{\infty} a_k$ *by inserting parentheses. If* $\displaystyle\sum_{k=1}^{\infty} a_k$ *converges to* s, *then* $\displaystyle\sum_{k=1}^{\infty} b_k$ *will converge to* s. *Next, if* $\displaystyle\lim_{n\to\infty} a_n = 0$, k_n *is bounded and* $\displaystyle\sum_{k=1}^{\infty} b_k$ *converges to* s, *then* $\displaystyle\sum_{k=1}^{\infty} a_k$ *will converge to* s.

(**Reason.** Let $s_n = \displaystyle\sum_{k=1}^{n} a_k$ and $t_n = \displaystyle\sum_{k=1}^{n} b_k$. For the first part, $\displaystyle\sum_{k=1}^{\infty} b_k = \lim_{n\to\infty} t_n = \lim_{n\to\infty} \sum_{k=1}^{p(n)} a_k = s$. For the second part, let $p(n) - p(n-1)$ be bounded by M. For a positive integer j, let $p(i) \le j < p(i+1)$. For $r = 1, 2, \ldots, M$, define $c_{r,j} = \begin{cases} a_{p(i)+r} & \text{if } p(i) + r \le j \\ 0 & \text{if } p(i) + r > j \end{cases}$. Then $\displaystyle\sum_{k=1}^{\infty} a_k = \lim_{j\to\infty} s_j = \lim_{i\to\infty} t_i + \lim_{j\to\infty} (c_{1,j} + \cdots + c_{M,j}) = s + 0 + \cdots + 0 = s$.)

Examples. (1) Since $\displaystyle\sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots$ converges to 1, so by the theorem,

$$\frac{1}{2} + \left(\frac{1}{4} + \frac{1}{8} \right) + \left(\frac{1}{16} + \frac{1}{32} + \frac{1}{64} \right) + \left(\frac{1}{128} + \frac{1}{256} + \frac{1}{512} + \frac{1}{1024} \right) + \cdots = 1.$$

(2) $(1-1)+(1-1)+\cdots$ converges to 0, but $1-1+1-1+\cdots$ diverges by term test. So $\lim_{n\to\infty} a_n = 0$ is important. Also, $(1-1)+\left(\frac{1}{2}+\frac{1}{2}-\frac{1}{2}-\frac{1}{2}\right)+\left(\frac{1}{3}+\frac{1}{3}+\frac{1}{3}-\frac{1}{3}-\frac{1}{3}-\frac{1}{3}\right)+\cdots$ converges to 0. However, the series without parentheses diverges (as $S_{n^2}=1$ and $S_{n^2+n}=0$) even though the terms have limit 0. So k_n bounded is important.

(3) Since $\left(1-\frac{1}{2}\right)+\left(\frac{1}{3}-\frac{1}{4}\right)+\cdots=\sum_{j=1}^{\infty}\left(\frac{1}{2j-1}-\frac{1}{2j}\right)=\sum_{j=1}^{\infty}\frac{1}{2j(2j-1)}$ converges (by the limit comparison test with $\sum_{j=1}^{\infty}\frac{1}{j^2}$), so by the theorem, $1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\cdots=\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}$ converges to the same sum.

Definition. $\sum_{k=1}^{\infty} b_k$ is a *rearrangement* of $\sum_{k=1}^{\infty} a_k$ iff there is a bijection $\sigma:\mathbb{N}\to\mathbb{N}$ such that $b_k=a_{\sigma(k)}$.

Example. Given $\ln 2 = 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\ldots$ (which converges conditionally). Consider the rearrangement $\underbrace{1+\frac{1}{3}}_{2+}\underbrace{-\frac{1}{2}}_{1-}+\underbrace{\frac{1}{5}+\frac{1}{7}}_{2+}\underbrace{-\frac{1}{4}}_{1-}+\underbrace{\frac{1}{9}+\frac{1}{11}}_{2+}\underbrace{-\frac{1}{6}}_{1-}+\ldots$. Observe that

$$
\begin{array}{rl}
 & (1-\frac{1}{2})+(\frac{1}{3}-\frac{1}{4})+(\frac{1}{5}-\frac{1}{6})+(\frac{1}{7}-\frac{1}{8})+\ldots=\ln 2\\
+ & \quad\frac{1}{2}\qquad\quad -\frac{1}{4}\qquad\quad +\frac{1}{6}\qquad\quad -\frac{1}{8}\qquad \ldots=\frac{1}{2}\ln 2\\
\hline
 & 1\quad +(\frac{1}{3}-\frac{1}{2})+\quad\frac{1}{5}\quad+(\frac{1}{7}-\frac{1}{4})+\ldots=\frac{3}{2}\ln 2.
\end{array}
$$

Riemann's Rearrangement Theorem. *Let $a_k\in\mathbb{R}$ and $\sum_{k=1}^{\infty} a_k$ converge conditionally. For any $x\in\mathbb{R}$ or $x=\pm\infty$, there is a rearrangement $\sum_{k=1}^{\infty} a_{\sigma(k)}$ of $\sum_{k=1}^{\infty} a_k$ such that $\sum_{k=1}^{\infty} a_{\sigma(k)}=x$.*

(**Sketch of reason.** Let $p_k=\begin{cases}a_k & \text{if } a_k\ge 0\\ 0 & \text{if } a_k<0\end{cases}$ and $q_k=\begin{cases}0 & \text{if } a_k\ge 0\\ |a_k| & \text{if } a_k<0\end{cases}$. Then $a_k=p_k-q_k$ and $|a_k|=p_k+q_k$. Now both $\sum_{k=1}^{\infty} p_k$, $\sum_{k=1}^{\infty} q_k$ must diverge to $+\infty$. (If both converges, then their sum $\sum_{k=1}^{\infty}|a_k|$ will be finite, a contradiction. If one converges and the other diverges to $+\infty$, then $\sum_{k=1}^{\infty} a_k=\sum_{k=1}^{\infty} p_k-\sum_{k=1}^{\infty} q_k$ will diverges to $\pm\infty$, a contradiction also.) Let u_n, v_n be sequences of real numbers having limits x and $u_n<v_n,\ u_n<v_{n+1},\ v_1>0$. Now let $P_1, P_2, \ldots$ be the nonnegative terms of $\sum_{k=1}^{\infty} a_k$ in the order they occur and $Q_1, Q_2, \ldots$ be the absolute value of the negative terms in the order they occur. Since $\sum_{k=1}^{\infty} P_k$, $\sum_{k=1}^{\infty} Q_k$ differ from $\sum_{k=1}^{\infty} p_k$, $\sum_{k=1}^{\infty} q_k$ only by zero terms, they also diverges to $+\infty$. Let m_1, k_1 be the smallest integers such that $P_1+\cdots+P_{m_1}>v_1$ and $P_1+\cdots+P_{m_1}-Q_1-\cdots-Q_{k_1}<u_1$. Let m_2, k_2 be the smallest integers such that $P_1+\cdots+P_{m_1}-Q_1-\cdots-Q_{k_1}+P_{m_1+1}+\cdots+P_{m_2}>v_2$ and $P_1+\cdots+P_{m_1}-Q_1-\cdots-Q_{k_1}+P_{m_1+1}+\cdots+P_{m_2}-Q_{k_1+1}-\cdots-Q_{k_2}<u_2$ and continue this way. This is possible since the sums of P_k and Q_k are $+\infty$. Now if s_n, t_n are the partial sums of this series $P_1+\cdots+P_{m_1}-Q_1-\cdots-Q_{k_1}+\cdots$ whose last terms are P_{m_n}, Q_{k_n}, respectively, then $|s_n-v_n|\le P_{m_n}$ and $|t_n-u_n|\le Q_{k_n}$ by the choices of m_n, k_n. Since P_n, Q_n have limit 0, so s_n, t_n must have limit x. As all other partial sums are squeezed by s_n and t_n, the series we constructed must have limit x.)

Dirichlet's Rearrangement Theorem. *If $a_k\in\mathbb{R}$ and $\sum_{k=1}^{\infty} a_k$ converges absolutely, then every rearrangement $\sum_{k=1}^{\infty} a_{\sigma(k)}$ converges to the same sum as $\sum_{k=1}^{\infty} a_k$.*

(**Reason.** Define p_k, q_k as in the last proof. Since $p_k, q_k \le |a_k|$, $\sum_{k=1}^{\infty} p_k$, $\sum_{k=1}^{\infty} q_k$ converge, say to p and q, respectively. Since $a_{\sigma(k)} = p_{\sigma(k)} - q_{\sigma(k)}$, we may view $\sum_{k=1}^{\infty} p_{\sigma(k)}$ as a rearrangement of the nonnegative terms of $\sum_{k=1}^{\infty} a_k$ and inserting zeros where $a_{\sigma(k)} < 0$. For any positive integer m, the partial sum $s_m = \sum_{k=1}^{m} p_{\sigma(k)} \le \sum_{k=1}^{\infty} p_k = p$. Since $p_k \ge 0$, the partial sum s_m is also increasing, hence $\sum_{k=1}^{\infty} p_{\sigma(k)}$ converges. Now, for every positive integer n, $\sum_{k=1}^{n} p_k \le \sum_{k=1}^{\infty} p_{\sigma(k)} \le p$. As $n \to \infty$, we get $\sum_{k=1}^{\infty} p_{\sigma(k)} = p$. Similarly, $\sum_{k=1}^{\infty} q_{\sigma(k)} = q$. Then $\sum_{k=1}^{\infty} a_{\sigma(k)} = p - q = \sum_{k=1}^{\infty} a_k$.)

Example. $\sum_{k=1}^{\infty}(-\frac{1}{2})^k = -\frac{1}{2} + \frac{1}{2^2} - \frac{1}{2^3} + \frac{1}{2^4} - \frac{1}{2^5} + \dots$ converges (absolutely) to $\dfrac{-\frac{1}{2}}{1-(-\frac{1}{2})} = -\dfrac{1}{3}$.

$$-\frac{1}{2} + \frac{1}{2^2} + \underbrace{\frac{1}{2^4} - \frac{1}{2^3}}_{\text{2 terms}} + \underbrace{\frac{1}{2^8} - \frac{1}{2^7} + \frac{1}{2^6} - \frac{1}{2^5}}_{\text{4 terms}} + \underbrace{\frac{1}{2^{16}} - \frac{1}{2^{15}} + \frac{1}{2^{14}} - \frac{1}{2^{13}} + \frac{1}{2^{12}} - \frac{1}{2^{11}} + \frac{1}{2^{10}} - \frac{1}{2^9}}_{\text{8 terms}} + \dots$$

is a rearrangement of $\sum_{k=1}^{\infty}(-\frac{1}{2})^k$, so it also converges to $-\frac{1}{3}$.

Remarks. As a consequence of the rearrangement theorem, the sum of a nonnegative series is the same no matter how the terms are rearranged.

Complex Series

Complex numbers $S_1, S_2, S_3, \dots$ with $S_n = u_n + i v_n$ are said to have limit $\lim_{n\to\infty} S_n = u + iv$ iff $\lim_{n\to\infty} u_n = u$ and $\lim_{n\to\infty} v_n = v$. A *complex series* is a series where the terms are complex numbers. The definitions of convergent, absolutely convergent and conditional convergent are the same. The remarks and the basic properties following the definitions of convergent and divergent series are also true for complex series.

The geometric series test, telescoping series test, term test, absolute convergence test, ratio test and root test are also true for complex series. For $z_k = x_k + i y_k$, we have $\sum_{k=1}^{\infty} z_k$ converges to $z = x + iy$ if and only if $\sum_{k=1}^{\infty} x_k$ converges to x and $\sum_{k=1}^{\infty} y_k$ converges to y. So complex series can be reduced to real series for study if necessary.

Examples. (1) Note $\lim_{n\to\infty} i^n \ne 0$ (otherwise $0 = \lim_{n\to\infty} |i^n| = \lim_{n\to\infty} 1$ is a contradiction). So $\sum_{k=1}^{\infty} i^k$ diverges by term test.

(2) If $|z| \le 1$, then $\left|\frac{z^k}{k^2}\right| \le \frac{1}{k^2}$ and $\sum_{k=1}^{\infty} \frac{1}{k^2}$ converges by p-test implies $\sum_{k=1}^{\infty} \frac{z^k}{k^2}$ converges absolutely. However, if $|z| > 1$, then $\lim_{k\to\infty}\left|\frac{z^{k+1}}{(k+1)^2}\frac{k^2}{z^k}\right| = \lim_{k\to\infty} \frac{k^2}{(k+1)^2}|z| = |z| > 1$ implies $\sum_{k=1}^{\infty} \frac{z^k}{k^2}$ diverges by the ratio test.

Chapter 5. Real Numbers

Decimal representations and points on a line are possible ways of introducing real numbers, but they are not too convenient for proving many theorems. Instead we will introduce real numbers by its important properties.

Axiomatic Formulation. There exists a set $\mathbb{R}$ (called *real numbers*) satisfying the following four axioms:

(1) (*Field Axiom*) $\mathbb{R}$ is a *field* (i.e. $\mathbb{R}$ has two operations $+$ and $\cdot$ such that for any $a, b, c \in \mathbb{R}$,

  (i) $a+b, a\cdot b \in \mathbb{R}$, (ii) $a+b = b+a, a\cdot b = b\cdot a$, (iii) $(a+b)+c = a+(b+c), (a\cdot b)\cdot c = a\cdot(b\cdot c)$,

  (iv) there are unique elements $0, 1 \in \mathbb{R}$ with $1 \neq 0$ such that $a+0 = a, a\cdot 1 = a$,

  (v) there is a unique element $-a \in \mathbb{R}$ such that $a + (-a) = 0$; if $a \neq 0$, then there is a unique element a^{-1} such that $a \cdot (a^{-1}) = 1$.

  (vi) $a \cdot (b + c) = a \cdot b + a \cdot c$.)

  (This axiom allows us to do algebra with equations. Define $a - b$ to mean $a + (-b)$; ab to mean $a \cdot b$; $\frac{a}{b}$ to mean $a \cdot (b^{-1})$. Also, define $2 = 1 + 1, 3 = 2 + 1, \ldots$.)

(2) (*Order Axiom*) $\mathbb{R}$ has an (ordering) relation $<$ such that for any $a, b \in \mathbb{R}$

  (i) exactly one of the following $a < b, a = b, b < a$ is true,

  (ii) if $a < b, b < c$, then $a < c$,

  (iii) if $a < b$, then $a + c < b + c$,

  (iv) if $a < b$ and $0 < c$, then $ac < bc$.

  (This axiom allows us to work with inequalities. For example, using (ii) and (iii), we can see that if $a < b$ and $c < d$, then $a+c < b+d$ because $a+c < b+c < b+d$. Also, we can get $0 < 1$ (for otherwise $1 < 0$ would imply by (iii) that $0 = 1 + (-1) < 0 + (-1) = -1$, which implies by (iv) that $0 < (-1)(-1) = 1$, a contradiction). Now define $a > b$ to mean $b < a$; $a \leq b$ to mean $a < b$ or $a = b$; etc. Also, define closed interval $[a, b] = \{x : a \leq x \leq b\}$; open interval $(a, b) = \{x : a < x < b\}$; etc. Part (i) of the order axiom implies any two real numbers can be compared. We define $\max(a_1, \ldots, a_n)$ to be the maximum of $a_1, \ldots, a_n$ and similarly for minimum. Also, define $|x| = \max(x, -x)$. Then $x \leq |x|$ and $-x \leq |x|$, i.e. $-|x| \leq x \leq |x|$. Next $|x| \leq a$ if and only if $x \leq a$ and $-x \leq a$, i.e. $-a \leq x \leq a$. Finally, adding $-|x| \leq x \leq |x|$ and $-|y| \leq y \leq |y|$, we get $-|x| - |y| \leq x + y \leq |x| + |y|$, which is the triangle inequality $|x + y| \leq |x| + |y|$.)

(3) (*Well-ordering Axiom*) $\mathbb{N} = \{1, 2, 3, \ldots\}$ is *well-ordered* (i.e. for any nonempty subset S of $\mathbb{N}$, there is $m \in S$ such that $m \leq x$ for all $x \in S$. This m is the *least element* (or the *minimum*) of S).

  (This axiom allows us to formulate the principle of mathematical induction later.)

Definitions. For a nonempty subset S of $\mathbb{R}$, S is *bounded above* iff there is some $M \in \mathbb{R}$ such that $x \leq M$ for all $x \in S$. Such an M is called an *upper bound* of S. The *supremum* or *least upper bound* of S (denoted by $\sup S$ or lub S) is an upper bound $\tilde{M}$ of S such that $\tilde{M} \leq M$ for all upper bounds M of S.

(4) (*Completeness Axiom*) Every nonempty subset of $\mathbb{R}$ which is bounded above has a supremum in $\mathbb{R}$.

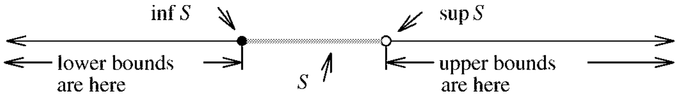
  (This axiom allows us to prove results that have to do with the existence of certain numbers with specific properties, as in the intermediate value theorem.)

Examples. (1) For $S = \left\{\frac{1}{n} : n \in \mathbb{N}\right\} = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$, the upper bounds of S are all $M \geq 1$. So $\sup S = 1 \in S$.
(2) For $S = \{x \in \mathbb{R} : x < 0\}$, the upper bounds of S are all $M \geq 0$. So $\sup S = 0 \notin S$.

Definitions. $\mathbb{N} = \{1, 2, 3, 4, \ldots\}$ is the *natural numbers* (or *positive integers*), $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ is the *integers*, $\mathbb{Q} = \{\frac{m}{n} : m \in \mathbb{Z}$ and $n \in \mathbb{N}\}$ is the *rational numbers* and $\mathbb{R} \setminus \mathbb{Q} = \{x \in \mathbb{R} : x \notin \mathbb{Q}\}$ is the *irrational numbers.*

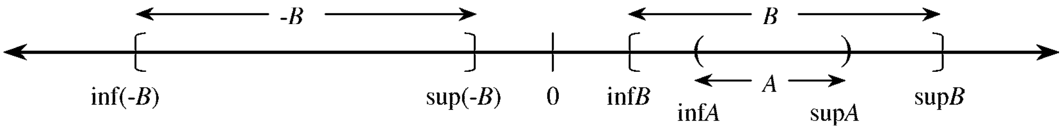
Remarks (Exercises). The first three axioms are also true if $\mathbb{R}$ is replaced by $\mathbb{Q}$. However, the completeness axiom is false for $\mathbb{Q}$. For example, $S = \{x : x \in \mathbb{Q}, x > 0, x^2 < 2\}$ is bounded above by 3 in $\mathbb{Q}$, but it does not have a supremum in $\mathbb{Q}$.

As above, we define S to be *bounded below* if there is some $m \in \mathbb{R}$ such that $m \le x$ for all $x \in S$. Such an m is called a *lower bound* of S. The *infimum* or *greatest lower bound* (denoted by $\inf S$ or $\operatorname{glb} S$) of S is a lower bound $\tilde{m}$ of S such that $m \le \tilde{m}$ for all lower bounds m of S.

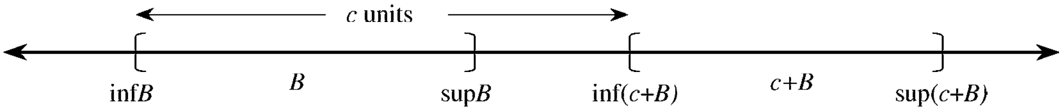


Remarks (Exercises). (1) Let $-B = \{-x : x \in B\}$. (This is the *reflection* of B about 0.) If B is bounded below, then $-B$ is bounded above and $\inf B = -\sup(-B)$. Similarly, if B is bounded above, then $-B$ is bounded below and $\sup B = -\inf(-B)$. From these and the completeness axiom, we get the following statement.

(*Completeness Axiom for Infimum*) Every nonempty subset of $\mathbb{R}$ which is bounded below has an infimum in $\mathbb{R}$.



(2) For a set B, if it is bounded above and $c \ge 0$, then let $cB = \{cx : x \in B\}$. (This is the *scaling* of B by a factor of c.) We have $\sup cB = c \sup B$. If $\emptyset \ne A \subseteq B$, then $\inf B \le \inf A$ whenever B is bounded below and $\sup A \le \sup B$ whenever B is bounded above.



(3) For $c \in \mathbb{R}$, let $c + B = \{c + x : x \in B\}$. (This is a *translation* of B by c units.) It follows that B has a supremum if and only if $c + B$ has a supremum, in which case $\sup(c + B) = c + \sup B$. The infimum statement is similar, i.e. $\inf(c + B) = c + \inf B$. More generally, if A and B are bounded, then letting $A + B = \{x + y : x \in A, y \in B\}$, we have $\sup(A + B) = \sup A + \sup B$ and $\inf(A + B) = \inf A + \inf B$.

If S is bounded above and below, then S is *bounded*. Note $\sup S$, $\inf S$ may or may not be in S. Also, if S is bounded, then for all $x \in S$, $|x| = \max(x, -x) \le \max(\sup S, -\inf S)$ (because $x \le \sup S$ and $-x \le -\inf S$.) Conversely, if there is $c \in \mathbb{R}$ such that for all $x \in S$, $|x| \le c$, then $-c \le x \le c$ so that S is bounded (above by c and below by $-c$.)

Simple Consequences of the Axioms.

Theorem (Infinitesimal Principle). *For $x, y \in \mathbb{R}$, $x < y + \varepsilon$ for all $\varepsilon > 0$ if and only if $x \le y$. (Similarly, $y - \varepsilon < x$ for all $\varepsilon > 0$ if and only if $y \le x$.)*

Proof. If $x \le y$, then for all $\varepsilon > 0$, $x \le y = y + 0 < y + \varepsilon$ by (iv) of the field axiom and (iii) of the order axiom.

Conversely, if $x < y + \varepsilon$ for all $\varepsilon > 0$, then assuming $x > y$, we get $x - y > 0$ by (iii) of the order axiom. Let $\varepsilon_0 = x - y$, then $x = y + \varepsilon_0$. Since $\varepsilon_0 > 0$, we also have $x < y + \varepsilon_0$. These contradict (i) of the order axiom. So $x \le y$. The other statement follows from the first statement since $y - \varepsilon < x$ is the same as $y < x + \varepsilon$.

Remarks. Taking $y = 0$, we see that $|x| < \varepsilon$ for all $\varepsilon > 0$ if and only if $x = 0$. This is used when it is difficult to show two expressions a, b are equal, but it may be easier to show $|a - b| < \varepsilon$ for every $\varepsilon > 0$.

Theorem (Mathematical Induction). *For every $n \in \mathbb{N}$, $A(n)$ is a (true or false) statement such that $A(1)$ is true and for every $k \in \mathbb{N}$, $A(k)$ is true implies $A(k + 1)$ is also true. Then $A(n)$ is true for all $n \in \mathbb{N}$.*

Proof. Suppose $A(n)$ is false for some $n \in \mathbb{N}$. Then $S = \{n \in \mathbb{N}: A(n) \text{ is false}\}$ is a nonempty subset of $\mathbb{N}$. By the well-ordering axiom, there is a least element m in S. Then $A(m)$ is false. Also, if $A(n)$ is false, then $m \le n$. Taking contrapositive, this means that if $n < m$, then $A(n)$ is true.

Now $A(1)$ is true, so $m \ne 1$ and $m \in \mathbb{N}$ imply $m \ge 2$. So $m - 1 \ge 1$. Let $k = m - 1 \in \mathbb{N}$, then $k = m - 1 < m$ implies $A(k)$ is true. By hypothesis, $A(k+1) = A(m)$ is true, a contradiction.

Theorem (Supremum Property). *If a set S has a supremum in $\mathbb{R}$ and $\varepsilon > 0$, then there is $x \in S$ such that* $\sup S - \varepsilon < x \le \sup S$.

Proof. Since $\sup S - \varepsilon < \sup S$, $\sup S - \varepsilon$ is not an upper bound of S. Then there is $x \in S$ such that $\sup S - \varepsilon < x$. Since $\sup S$ is an upper bound of S, $x \le \sup S$. Therefore $\sup S - \varepsilon < x \le \sup S$.

Theorem (Infimum Property). *If a set S has an infimum in $\mathbb{R}$ and $\varepsilon > 0$, then there is $x \in S$ such that* $\inf S + \varepsilon > x \ge \inf S$.

Proof. Since $\inf S + \varepsilon > \inf S$, $\inf S + \varepsilon$ is not a lower bound of S. Then there is $x \in S$ such that $\inf S + \varepsilon > x$. Since $\inf S$ is a lower bound of S, $x \ge \inf S$. Therefore $\inf S + \varepsilon > x \ge \inf S$.

Theorem (Archimedean Principle). *For any $x \in \mathbb{R}$, there is $n \in \mathbb{N}$ such that $n > x$.*

Proof. Assume there exists $x \in \mathbb{R}$ such that for all $n \in \mathbb{N}$, we have $n \le x$. Then $\mathbb{N} = \{n : n \in \mathbb{N}\}$ has an upper bound x. By the completeness axiom, $\mathbb{N}$ has a supremum in $\mathbb{R}$. By the supremum property, there is $n \in \mathbb{N}$ such that $\sup \mathbb{N} - 1 < n$, which yields the contradiction $\sup \mathbb{N} < n + 1 \in \mathbb{N}$.

Question. How is $\mathbb{Q}$ contained in $\mathbb{R}$? How is $\mathbb{R} \setminus \mathbb{Q}$ contained in $\mathbb{R}$?

Below we will show that $\mathbb{Q}$ is "dense" in $\mathbb{R}$ in the sense that between any two distinct real numbers x, y, no matter how close, there is a rational number. Similarly, $\mathbb{R} \setminus \mathbb{Q}$ is "dense" in $\mathbb{R}$. First we need a lemma.

Lemma. *For every $x \in \mathbb{R}$, there exists a least integer greater than or equal to x. (In computer science, this is called the ceiling of x and is denoted by $\lceil x \rceil$.) Similarly, there exists a greatest integer less than or equal to x. (This is denoted by $[x]$. In computer science, this is also called the floor of x and is denoted by $\lfloor x \rfloor$.)*

Proof. By the Archimedean principle, there is $n \in \mathbb{N}$ such that $n > |x|$. Then $-n < x < n$. By (iii) of the order axiom, $0 < x + n < 2n$. The set $S = \{k \in \mathbb{N} : k \ge x + n\}$ is a nonempty subset of $\mathbb{N}$ because $2n \in S$. By the well-ordering axiom, there is a least positive integer $m \ge x + n$. Then $m - n$ is the least integer greater than or equal to x. So the ceiling of every real number always exist.

Next, to find the floor of x, let k be the least integer greater than or equal to $-x$, then $-k$ is the greatest integer less than or equal to x.

Theorem (Density of Rational Numbers). *If $x < y$, then there is $\frac{m}{n} \in \mathbb{Q}$ such that $x < \frac{m}{n} < y$.*

Proof. By the Archimedean principle, there is $n \in \mathbb{N}$ such that $n > 1/(y - x)$. So $ny - nx > 1$ and hence $nx + 1 < ny$. Let $m = [nx] + 1$, then $m - 1 = [nx] \le nx < [nx] + 1 = m$. So $nx < m \le nx + 1 < ny$, i.e. $x < \frac{m}{n} < y$.

Theorem (Density of Irrational Numbers). *If $x < y$, then there is $w \in \mathbb{R} \setminus \mathbb{Q}$ such that $x < w < y$.*

Proof. Let $w_0 \in \mathbb{R} \setminus \mathbb{Q}$. By the density of rational numbers, there is $\frac{m}{n} \in \mathbb{Q}$ such that $\frac{x}{|w_0|} < \frac{m}{n} < \frac{y}{|w_0|}$. (If $\frac{m}{n} = 0$, then pick another rational number between 0 and $\frac{y}{|w_0|}$. So we may take $\frac{m}{n} \ne 0$.) Let $w = \frac{m}{n}|w_0|$, then $w \in \mathbb{R} \setminus \mathbb{Q}$ and $x < w < y$.

Examples. (1) Let $S = (-\infty, 3) \cup (4, 5]$, then S is not bounded below and so S has no infimum. On the other hand, S is bounded above by 5 and every upper bound of S is greater than or equal to $5 \in S$. So $\sup S = 5$.

(2) Let $S = \{\frac{1}{n} : n \in \mathbb{N}\} = \{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\}$. In the examples following the definition of supremum, we saw $\sup S = 1$. Here we will show $\inf S = 0$. (Note $0 \notin S$.) Since $\frac{1}{n} > 0$ for all $n \in \mathbb{N}$, 0 is a lower bound of S. So by the completeness axiom for infimum, $\inf S$ must exist. Assume S has a lower bound $t > 0$. By the Archimedean principle, there is $n \in \mathbb{N}$ such that $n > 1/t$. Then $t > 1/n \in S$, a contradiction to t being a lower bound of S. So 0 is the greatest lower bound of S.

(3) Let $S = [2, 6) \cap \mathbb{Q}$. Since $2 \leq x < 6$ for every $x \in S$, S has 2 as a lower bound and 6 as an upper bound. We will show $\inf S = 2$ and $\sup S = 6$. (Note $2 \in S$ and $6 \notin S$.) Since $2 \in S$, so every lower bound t satisfy $t \leq 2$. Therefore $\inf S = 2$. For supremum, assume there is an upper bound $u < 6$. Since $2 \in S$, so $2 \leq u$. By the density of rational numbers, there is a $r \in \mathbb{Q}$ such that $u < r < 6$. Then $r \in [2, 6) \cap \mathbb{Q} = S$. As $u < r$ contradicts u being an upper bound of S, so every upper bound $u \geq 6$. Therefore, $\sup S = 6$.

Chapter 6. Limits

Limit is the most important concept in analysis. We will first discuss limits of sequences, then limits of functions.

Definitions. An (infinite) *sequence* in a set S (*e.g.* $S = \mathbb{R}$ or $S = [0, 1]$) is a list $x_1, x_2, x_3, \ldots$ of elements of S in a specific order. Briefly it is denoted by $\{x_n\}$. (Mathematically it may be viewed as a function $x\colon \mathbb{N} \to S$ with $x(n) = x_n$ for $n \in \mathbb{N}$.) We say the sequence $\{x_n\}$ is *bounded above* iff the set $\{x_1, x_2, x_3, \ldots\}$ is bounded above. (Bounded below and bounded sequences are defined similarly.) We will also write $\sup\{x_n\}$ for the supremum of the set $\{x_1, x_2, x_3, \ldots\}$ and $\inf\{x_n\}$ for the infimum of the set $\{x_1, x_2, x_3, \ldots\}$.

CAUTION: Since we seldom talk about a set with one element from now on, so notations like $\{x_n\}$ will denote sequences unless explicitly stated otherwise.

For $x, y \in \mathbb{R}$, the distance between x and y is commonly denoted by $d(x, y)$, which equals $|x - y|$. Below we will need a quantitative measure of what it means to be "close" for a discussion of the concept of limit. For $\varepsilon > 0$, the open interval $(c - \varepsilon, c + \varepsilon)$ is called the *ε-neighborhood* of c. Note $x \in (c - \varepsilon, c + \varepsilon)$ if and only if $d(x, c) = |x - c| < \varepsilon$, i.e. every number in $(c - \varepsilon, c + \varepsilon)$ has distance less than ε from c.

Limit of a sequence $\{x_n\}$ is often explained by saying it is the number the x_n's are *closer and closer* to as n gets *larger and larger.* There are two bad points about this explanations.

(1) Being close or large is a feeling! It is not a fact. It cannot be proved by a logical argument.

(2) The effect of being close can accumulate to yield large separation! If two numbers having a distance less than or equal to 1 are considered close, then 0 is close to 1 and 1 is close to 2 and 2 is close to 3, ..., 99 is close to 100, but 0 is quite far from 100.

So what is the meaning of close? How can limit be defined so it can be checked? Intuitively, a sequence $\{x_n\}$ gets close to a number x if and only if the distance $d(x_n, x)$ goes to 0. This happens if and only if for every positive ε, the distance $d(x_n, x)$ *eventually* becomes less than ε. The following example will try to make this more precise.

Example. As n gets large, intuitively we may think $x_n = \dfrac{2n^2 - 1}{n^2 + 1}$ gets close to 2. For $\varepsilon = 0.1$, how soon (that is, for what n) will the distance $d(x_n, 2)$ be less than ε? (What if $\varepsilon = 0.01$? What if $\varepsilon = 0.001$? What if ε is an arbitrary positive number?)

Solution. Consider $d(x_n, 2) = \left|\dfrac{2n^2 - 1}{n^2 + 1} - 2\right| = \dfrac{3}{n^2 + 1} < \varepsilon$. Solving for n, we get $n^2 > (3/\varepsilon) - 1$. If $\varepsilon = 0.1$, then $n > \sqrt{29}$. So as soon as $n \geq 6$, the distance between x_n and 2 will be less than $\varepsilon = 0.1$.

(If $\varepsilon = 0.01$, then $n > \sqrt{299}$. So $n \geq 18$ will do. If $\varepsilon = 0.001$, then $n > \sqrt{2999}$. So $n \geq 55$ will do. If $0 < \varepsilon \leq 3$, then $n \geq [\sqrt{(3/\varepsilon) - 1}] + 1$ will do. If $\varepsilon > 3$, then since $\frac{3}{n^2+1} < 3 < \varepsilon$ for every $n \in \mathbb{N}$, so $n \geq 1$ will do. So for every $\varepsilon > 0$, there is a $K \in \mathbb{N}$ so that as soon as $n \geq K$, the distance $d(x_n, 2)$ will be less than ε.) Note the value of K depends on the value of ε; the smaller ε is, the larger K will be. (Some people write K_ε to indicate K depends on ε.)

Definition. A sequence $\{x_n\}$ *converges* to a number x (or has *limit* x) iff for every $\varepsilon > 0$, there is $K \in \mathbb{N}$ such that for every $n \geq K$, it implies $d(x_n, x) = |x_n - x| < \varepsilon$ (which means $x_K, x_{K+1}, x_{K+2}, \ldots \in (x - \varepsilon, x + \varepsilon)$.)

Remarks. (i) From the definition, we see that $\{x_n\}$ converges to x, $\{x_n - x\}$ converges to 0 and $\{|x_n - x|\}$ converges to 0 are equivalent because in the definition, $|x_n - x|$ is the same as $|(x_n - x) - 0| = ||x_n - x| - 0|$.

(ii) To show $\{x_n\}$ converges to x means for every $\varepsilon > 0$, we have to *find* a K as in the definition or *show* such a K exists. On the other hand, if we are given that $\{x_n\}$ converges to x, then for every $\varepsilon > 0$, (which we can even choose for our convenience,) there is a K as in the definition for us to use.

Let us now do a few more examples to illustrate how to show a sequence converges by checking the definition. Later, we will prove some theorems that will help in establishing convergence of sequences.

Examples. (1) Let $v_n = c$. For every $\varepsilon > 0$, let $K = 1$, then $n \geq K$ implies $|v_n - c| = 0 < \varepsilon$. So $\{v_n\}$ converges to c.

---

(2) Let $w_n = c - \frac{1}{n}$. For every $\varepsilon > 0$, there exists an integer $K > \frac{1}{\varepsilon}$ (by the Archimedean principle). Then $n \geq K$ implies $|w_n - c| = \frac{1}{n} \leq \frac{1}{K} < \varepsilon$. So $\{w_n\}$ converges to c.

---

(3) Let $x_n = \dfrac{n}{(\cos n) - n}$. Show that $\{x_n\}$ converges to -1 by checking the definition.

Solution. For every $\varepsilon > 0$, there exists an integer $K > 1 + \frac{1}{\varepsilon}$ by the Archimedian principle. Then $n \geq K$ implies $\left|\dfrac{n}{(\cos n) - n} - (-1)\right| = \left|\dfrac{\cos n}{(\cos n) - n}\right| \leq \dfrac{1}{n-1} \leq \dfrac{1}{K-1} < \varepsilon$. So $\{x_n\}$ converges to -1.

---

(4) Let $y_n = (-1)^n$. Show that $\{y_n\}$ does not converge.

Solution. Assume $\{y_n\}$ converges, say to y. Let $\varepsilon = 0.1$. Then there exists $K \in \mathbb{N}$ such that $n \geq K$ implies $|(-1)^n - y| < \varepsilon = 0.1$. Taking an odd integer $n \geq k$, we get $|-1-y| < 0.1$, which implies $y \in (-1.1, -0.9)$. Taking a even integer $n \geq K$, we get $|1 - y| < 0.1$, which implies $y \in (0.9, 1.1)$. Since no y is in both $(-1.1, -0.9)$ and $(0.9, 1.1)$, we have a contradiction.

---

(5) Let $z_n = n^{1/n}$. Show that $\{z_n\}$ converges to 1 by checking the definition.

Solution. (Let $u_n = |z_n - 1| = z_n - 1$. By the binomial theorem,

$$n = z_n^n = (1 + u_n)^n = 1 + nu_n + \frac{n(n-1)}{2}u_n^2 + \cdots + u_n^n \geq \frac{n(n-1)}{2}u_n^2$$

so that $u_n \leq \sqrt{\dfrac{2}{n-1}}$.) For every $\varepsilon > 0$, there exists integer $K > 1 + \frac{2}{\varepsilon^2}$ (by the Archimedean principle). Then $n \geq K$ implies $|z_n - 1| = u_n \leq \sqrt{\dfrac{2}{n-1}} \leq \sqrt{\dfrac{2}{K-1}} < \varepsilon$. So $\{z_n\}$ converges to 1.

Theorem (Uniqueness of Limit). *If $\{x_n\}$ converges to x and y, then $x = y$ (and so we may write $\lim_{n\to\infty} x_n = x$).*

Proof. For every $\varepsilon > 0$, we will show $|x-y| < \varepsilon$. (By the infinitesimal principle, we will get $x = y$.) Let $\varepsilon_0 = \varepsilon/2 > 0$. By the definition of convergence, there are $K_1, K_2 \in \mathbb{N}$ such that $n \geq K_1 \Rightarrow |x_n - x| < \varepsilon_0$ and $n \geq K_2 \Rightarrow |x_n - y| < \varepsilon_0$. Let $K = \max(K_1, K_2)$. By the triangle inequality, $|x-y| = |(x - x_K) + (x_K - y)| \leq |x - x_K| + |x_K - y| < \varepsilon_0 + \varepsilon_0 = \varepsilon$.

Boundedness Theorem. *If $\{x_n\}$ converges, then $\{x_n\}$ is bounded.*

Proof. Let $\lim_{n\to\infty} x_n = x$. For $\varepsilon = 1$, there is $K \in \mathbb{N}$ such that $n \geq K \Rightarrow |x_n - x| < 1 \Rightarrow |x_n| = |(x_n - x) + x| < 1 + |x|$. Let $M = \max(|x_1|, \ldots, |x_{K-1}|, 1 + |x|)$, then for every $n \in \mathbb{N}$, $|x_n| \leq M$ (i.e. $x_n \in [-M, M]$).

Remarks. The converse is false. The sequence $\{(-1)^n\}$ is bounded, but not convergent by example (4). In general, bounded sequences may or may not converge.

Theorem (Computation Formulas for Limits). *If $\{x_n\}$ converges to x and $\{y_n\}$ converges to y, then*

(i) *$\{x_n \pm y_n\}$ converges to $x \pm y$, respectively, i.e.* $\lim_{n\to\infty}(x_n \pm y_n) = \lim_{n\to\infty} x_n \pm \lim_{n\to\infty} y_n$,

(ii) *$\{x_n y_n\}$ converges to xy, i.e.* $\lim_{n\to\infty}(x_n y_n) = \left(\lim_{n\to\infty} x_n\right)\left(\lim_{n\to\infty} y_n\right)$,

(iii) *$\{x_n/y_n\}$ converges to x/y, provided $y_n \neq 0$ for all n and $y \neq 0$.*

Proof. (i) For every $\varepsilon > 0$, there are $K_1, K_2 \in \mathbb{N}$ such that $n \ge K_1 \Rightarrow |x_n - x| < \varepsilon/2$ and $n \ge K_2 \Rightarrow |y_n - y| < \varepsilon/2$. Let $K = \max(K_1, K_2)$. Then $n \ge K$ implies $n \ge K_1$ and $n \ge K_2$. So for these n's,

$$|(x_n \pm y_n) - (x \pm y)| = |(x_n - x) \pm (y_n - y)| \le |x_n - x| + |y_n - y| < \varepsilon/2 + \varepsilon/2 = \varepsilon.$$

(ii) We prove a lemma first.

Lemma. *If $\{a_n\}$ is bounded and $\lim_{n\to\infty} b_n = 0$, then $\lim_{n\to\infty} a_n b_n = 0$.*

Proof. Since $\{a_n\}$ is bounded, there is M such that $|a_n| < M$ for all n. For every $\varepsilon > 0$, since $\varepsilon/M > 0$ and $\{b_n\}$ converges to 0, there is $K \in \mathbb{N}$ such that $n \ge K \Rightarrow |b_n - 0| < \varepsilon/M \Rightarrow |a_n b_n - 0| \le M|b_n| < \varepsilon$.

To prove (ii), we write $x_n y_n - xy = x_n y_n - x_n y + x_n y - xy = x_n(y_n - y) + y(x_n - x)$. Since $\{x_n\}$ converges, $\{x_n\}$ is bounded by the boundedness theorem. So by (i) and the lemma,

$$\lim_{n\to\infty} x_n y_n = \lim_{n\to\infty} (x_n y_n - xy) + \lim_{n\to\infty} xy = \lim_{n\to\infty} x_n(y_n - y) + \lim_{n\to\infty} y(x_n - x) + xy = 0 + 0 + xy = xy.$$

(iii) Note $\frac{1}{2}|y| > 0$. Since $\{y_n\}$ converges to y, there is $K_0 \in \mathbb{N}$ such that $n \ge K_0$ implies $|y_n - y| < \frac{1}{2}|y|$. By the triangle inequality, $|y| - |y_n| \le |y_n - y| < \frac{1}{2}|y| \Rightarrow \frac{1}{2}|y| < |y_n|$ for $n \ge K_0$. Then for every $n \in \mathbb{N}$, $|y_n| \ge m = \min(|y_1|, \ldots, |y_{K_0-1}|, \frac{1}{2}|y|) > 0$.

Next we will show $\lim_{n\to\infty} \dfrac{1}{y_n} = \dfrac{1}{y}$ (then by (ii), $\lim_{n\to\infty} \dfrac{x_n}{y_n} = \lim_{n\to\infty} (x_n \dfrac{1}{y_n}) = x\dfrac{1}{y} = \dfrac{x}{y}$). For every $\varepsilon > 0$, let $\varepsilon_0 = m|y|\varepsilon > 0$. Since $\lim_{n\to\infty} y_n = y \ne 0$, there is $K \in \mathbb{N}$ such that $n \ge K \Rightarrow |y_n - y| < \varepsilon_0$. Then

$$n \ge K \Rightarrow \left|\frac{1}{y_n} - \frac{1}{y}\right| = \frac{|y - y_n|}{|y_n||y|} < \frac{\varepsilon_0}{m|y|} = \varepsilon.$$

Remarks. (1) As in the proofs of the uniqueness of limit and part (i) of the computation formulas, when we have $n \ge K_1 \Rightarrow |a_n - a| < \varepsilon_1$ and $n \ge K_2 \Rightarrow |b_n - b| < \varepsilon_2$, we may as well take $K = \max(K_1, K_2)$ to say $n \ge K \Rightarrow |a_n - a| < \varepsilon_1$ and $|b_n - b| < \varepsilon_2$ from now on.

(2) By mathematical induction, we can show that the computation formulas also hold for finitely many sequences. However, the number of sequences must stay constant as the following example shows

$$1 = \lim_{n\to\infty} \underbrace{\left(\frac{1}{n} + \cdots + \frac{1}{n}\right)}_{n \text{ terms}} \ne \lim_{n\to\infty} \frac{1}{n} + \cdots + \lim_{n\to\infty} \frac{1}{n} = 0 + \cdots + 0 = 0.$$

Sandwich Theorem (or Squeeze Limit Theorem). *If $x_n \le w_n \le y_n$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n = z$, then $\lim_{n\to\infty} w_n = z$.*

Proof. For any $\varepsilon > 0$, there is K such that $n \ge K \Rightarrow |x_n - z| < \varepsilon$ and $|y_n - z| < \varepsilon$, i.e. $x_n, y_n \in (z - \varepsilon, z + \varepsilon)$. Since $x_n \le w_n \le y_n$, so $w_n \in (z - \varepsilon, z + \varepsilon)$, i.e. $|w_n - z| < \varepsilon$.

Example. Let $w_n = \dfrac{[10^n\sqrt{2}]}{10^n} \in \mathbb{Q}$ for every $n \in \mathbb{N}$. (Note $w_1 = 1.4, w_2 = 1.41, w_3 = 1.414, w_4 = 1.4142, \ldots$.) Then $\dfrac{10^n\sqrt{2} - 1}{10^n} < w_n \le \dfrac{10^n\sqrt{2}}{10^n} = \sqrt{2}$. Since $\lim_{n\to\infty} \dfrac{10^n\sqrt{2} - 1}{10^n} = \sqrt{2}$, by the sandwich theorem, $\lim_{n\to\infty} w_n = \sqrt{2}$.

Theorem (Limit Inequality). *If $a_n \ge 0$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} a_n = a$, then $a \ge 0$.*

Proof. Assume $a < 0$. Then for $\varepsilon = |a|$, there is $K \in \mathbb{N}$ such that $n \ge K \Rightarrow |a_n - a| < \varepsilon = |a|$, which implies $a - \varepsilon < a_n < a + \varepsilon = a + (-a) = 0$, a contradiction.

Remarks. By the limit inequality above, if $x_n \le y_n$, $\lim_{n\to\infty} x_n = x$, $\lim_{n\to\infty} y_n = y$, then taking $a_n = y_n - x_n \ge 0$, we get the limit $y - x \ge 0$, i.e. $x \le y$. Also, if $a \le x_n \le b$ and $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} a = a \le \lim_{n\to\infty} x_n = x \le \lim_{n\to\infty} b = b$, i.e. $x_n \in [a, b]$ and $\lim_{n\to\infty} x_n = x$ imply $x \in [a, b]$. (This is false for open intervals as $\frac{1}{n} \in (0, 2)$, $\lim_{n\to\infty} \dfrac{1}{n} = 0 \notin (0, 2)$.)

Supremum Limit Theorem. *Let S be a nonempty set with an upper bound c. There is a sequence $\{w_n\}$ in S converging to c if and only if $c = \sup S$.*

Proof. If $c = \sup S$, then for $n \in \mathbb{N}$, by the supremum property, there is $w_n \in S$ such that $c - \frac{1}{n} = \sup S - \frac{1}{n} < w_n \leq \sup S = c$. Since $\lim_{n\to\infty} (c - \dfrac{1}{n}) = \lim_{n\to\infty} c = c$, the sandwich theorem implies $\lim_{n\to\infty} w_n = c = \sup S$.

Conversely, if a sequence $\{w_n\}$ in S converges to c, then $w_n \leq \sup S$ implies $c = \lim_{n\to\infty} w_n \leq \sup S$. Since c is an upper bound of S, so $\sup S \leq c$. Therefore $c = \sup S$.

Infimum Limit Theorem. *Let S be a nonempty set with a lower bound c. There is a sequence $\{w_n\}$ in S converging to c if and only if $c = \inf S$.*

Examples. (1) Let $S = \{\frac{1}{n} : n \in \mathbb{N}\} = \{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$. Since $0 \leq \frac{1}{n}$ for all $n \in \mathbb{N}$, 0 is a lower bound of the set S. Now the sequence $\{\frac{1}{n}\}$ in S converges to 0. By the infimum limit theorem, $\inf S = 0$.

(2) Let $S = \{x\pi + \frac{1}{y} : x \in \mathbb{Q} \cap (0, 1], y \in [1, 2]\}$. Since $\frac{1}{2} \leq x\pi + \frac{1}{y}$ for all $x \in \mathbb{Q} \cap (0, 1]$ and $y \in [1, 2]$, $\frac{1}{2}$ is a lower bound of S. Now the sequence $\{\frac{1}{n}\pi + \frac{1}{2}\}$ in S converges to $\frac{1}{2}$. By the infimum limit theorem, $\inf S = \frac{1}{2}$.

(3) Let A and B be bounded in $\mathbb{R}$. Prove that if $A - 2B = \{a - 2b : a \in A, b \in B\}$, then $\sup(A - 2B) = \sup A - 2 \inf B$.

Solution. Since A and B are bounded, $\sup A$ and $\inf B$ exist by the completeness axiom. For $x \in A - 2B$, we have $x = a - 2b$ for some $a \in A$ and $b \in B$. So $x = a - 2b \leq \sup A - 2 \inf B$. Hence $\sup A - 2 \inf B$ is an upper bound for $A - 2B$. By the supremum limit theorem, there is a sequence $a_n \in A$ such that $\{a_n\}$ converges to $\sup A$. By the infimum limit theorem, there is a sequence $b_n \in B$ such that $\{b_n\}$ converges to $\inf B$. Then $\{a_n - 2b_n\}$ is a sequence in $A - 2B$ and $\{a_n - 2b_n\}$ converges to $\sup A - 2 \inf B$ by the computation formulas for limits. By the supremum limit theorem, therefore $\sup(A - 2B) = \sup A - 2 \inf B$.

Definition. A *subsequence* of $\{x_n\}$ is a sequence $\{x_{n_j}\}$, where $n_j \in \mathbb{N}$ and $n_1 < n_2 < n_3 < \ldots$.

Examples. For the sequence $x_1, x_2, x_3, x_4, x_5, x_6, \ldots$, if we set $n_j = j^2$, we get the subsequence $x_1, x_4, x_9, x_{16}, \ldots$. If $n_j = 2j + 1$, then we get the subsequence $x_3, x_5, x_7, x_9, \ldots$. If n_j is the j-th prime number, then we get the subsequence $x_2, x_3, x_5, x_7, \ldots$.

Remarks. (1) Taking $n_j = j$, we see that every sequence is a subsequence of itself. A subsequence can also be thought of as obtained from the original sequence by throwing away possibly some terms. Also, a subsequence of a subsequence of $\{x_n\}$ is a subsequence of $\{x_n\}$.

(2) By mathematical induction, we have $n_j \geq j$ for all $j \in \mathbb{N}$ because $n_1 \geq 1$ and $n_{j+1} > n_j \geq j$ implies $n_{j+1} \geq j + 1$.

Subsequence Theorem. *If $\lim_{n\to\infty} x_n = x$, then $\lim_{j\to\infty} x_{n_j} = x$ for every subsequence $\{x_{n_j}\}$ of $\{x_n\}$. (The converse is trivially true because every sequence is a subsequence of itself.)*

Proof. For every $\varepsilon > 0$, there is $K \in \mathbb{N}$ such that $n \geq K \Rightarrow |x_n - x| < \varepsilon$. Then $j \geq K \Rightarrow n_j \geq K \Rightarrow |x_{n_j} - x| < \varepsilon$.

Question. How can we tell if a sequence converges without knowing the limit (especially if the sequence is given by a recurrence relation)?

For certain types of sequences, the question has an easy answer.

Definitions. $\{x_n\}$ is $\begin{Bmatrix} \textit{increasing} \\ \textit{decreasing} \\ \textit{strictly increasing} \\ \textit{strictly decreasing} \end{Bmatrix}$ iff $\begin{Bmatrix} x_1 \leq x_2 \leq x_3 \leq \ldots \\ x_1 \geq x_2 \geq x_3 \geq \ldots \\ x_1 < x_2 < x_3 < \ldots \\ x_1 > x_2 > x_3 > \ldots \end{Bmatrix}$, respectively. $\{x_n\}$ is $\begin{Bmatrix} \textit{monotone} \\ \textit{strictly monotone} \end{Bmatrix}$ iff $\{x_n\}$ is $\begin{Bmatrix} \text{increasing or decreasing} \\ \text{strictly increasing or decreasing} \end{Bmatrix}$, respectively.

Monotone Sequence Theorem. *If $\{x_n\}$ is increasing and bounded above, then $\lim_{n\to\infty} x_n = \sup\{x_n\}$. (Similarly, if $\{x_n\}$ is decreasing and bounded below, then $\lim_{n\to\infty} x_n = \inf\{x_n\}$.)*

Proof. Let $M = \sup\{x_n\}$, which exists by the completeness axiom. By the supremum property, for any $\varepsilon > 0$, there is x_K such that $M - \varepsilon < x_K \leq M$. Then $j \geq K \Rightarrow M - \varepsilon < x_K \leq x_j \leq M \Rightarrow |x_j - M| = M - x_j < \varepsilon$.

Remark. Note the completeness axiom was used to show the limit of x_n exists (without giving the value).

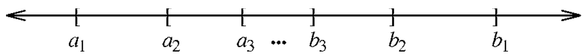
Examples. (1) Let $0 < c < 1$ and $x_n = c^{1/n}$. Then $x_n < 1$ and $c^{n+1} < c^n \Rightarrow x_n = c^{1/n} < c^{1/(n+1)} = x_{n+1}$. So by the monotone sequence theorem, $\{x_n\}$ has a limit x. Now $x_{2n}^2 = (c^{1/2n})^2 = c^{1/n} = x_n$. Taking limits and using the subsequence theorem, we get $x^2 = x$. So $x = 0$ or 1. Since $0 < c = x_1 \leq x$, the limit x is 1. Similarly, if $c \geq 1$, then $c^{1/n}$ will decrease to the limit 1.

(2) Does $\sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots}}}$ represent a real number?

Here we have a nested radical defined by $x_1 = \sqrt{2}$ and $x_{n+1} = \sqrt{2 + x_n}$. The question is whether $\{x_n\}$ converges to a real number x. (Computing a few terms, we suspect that $\{x_n\}$ is increasing. To find an upper bound, observe that if $\lim_{n\to\infty} x_n = x$, then $x = \sqrt{2 + x}$ implies $x = 2$.) Now by mathematical induction, we can show that $x_n < x_{n+1} < 2$. (If $x_n < x_{n+1} < 2$, then $2 + x_n < 2 + x_{n+1} < 4$, so taking square roots, we get $x_{n+1} < x_{n+2} < 2$.) By the monotone sequence theorem, $\{x_n\}$ has a limit x. We have $x^2 = \lim_{n\to\infty} x_{n+1}^2 = \lim_{n\to\infty} 2 + x_n = 2 + x$. Then $x = -1$ or 2. Since $\sqrt{2} = x_1 \leq x$, so $x = 2$.

Another common type of sequences is obtained by mixing a decreasing sequence and an increasing sequence into one of the form $a_1, b_1, a_2, b_2, a_3, b_3, \ldots$. In the next example, we will have such a situation and we need two theorems to handle these kind of sequences.

Nested Interval Theorem. *If $I_n = [a_n, b_n]$ is such that $I_1 \supseteq I_2 \supseteq I_3 \supseteq \ldots$, then $\bigcap_{n=1}^{\infty} I_n = [a, b]$, where $a = \lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n = b$. If $\lim_{n\to\infty}(b_n - a_n) = 0$, then $\bigcap_{n=1}^{\infty} I_n$ contains exactly one number.*



Proof. $I_1 \supseteq I_2 \supseteq I_3 \supseteq \ldots$ implies $\{a_n\}$ is increasing and bounded above by b_1 and $\{b_n\}$ is decreasing and bounded below by a_1. By the monotone sequence theorem, $\{a_n\}$ converges to $a = \sup\{a_n\}$ and $\{b_n\}$ converges to $b = \inf\{b_n\}$. Since $a_n \leq b_n$ for every $n \in \mathbb{N}$, taking limits, we have $a_n \leq a \leq b \leq b_n$. Consequently, $x \in [a_n, b_n]$ (i.e. $a_n \leq x \leq b_n$) for all n if and only if $\lim_{n\to\infty} a_n = a \leq x \leq b = \lim_{n\to\infty} b_n$. So $\bigcap_{n=1}^{\infty} I_n = [a, b]$. If $0 = \lim_{n\to\infty}(b_n - a_n) = b - a$, then $a = b$ and $\bigcap_{n=1}^{\infty} I_n = \{a\}$.

Remarks. Note in the proof, the monotone sequence theorem was used. So the nested interval theorem also implicitly depended on the completeness axiom.

Intertwining Sequence Theorem. *If $\{x_{2m}\}$ and $\{x_{2m-1}\}$ converge to x, then $\{x_n\}$ also converges to x.*

Proof. For every $\varepsilon > 0$, since $\{x_{2m}\}$ converges to x, there is $K_0 \in \mathbb{N}$ such that $m \geq K_0 \Rightarrow |x_{2m} - x| < \varepsilon$. Since $\{x_{2m-1}\}$ also converges to x, there is $K_1 \in \mathbb{N}$ such that $m \geq K_1 \Rightarrow |x_{2m-1} - x| < \varepsilon$. Now if $n \geq K = \max(2K_0, 2K_1 - 1)$, then either $n = 2m \geq 2K_0 \Rightarrow |x_n - x| = |x_{2m} - x| < \varepsilon$ or $n = 2m - 1 \geq 2K_1 - 1 \Rightarrow |x_n - x| = |x_{2m-1} - x| < \varepsilon$.

Example. Does $\dfrac{1}{1 + \dfrac{1}{1 + \cdots}}$ represent a number?

Here we have a continued fraction defined by $x_1 = 1$ and $x_{n+1} = 1/(1 + x_n)$. We have $x_1 = 1, x_2 = 1/2, x_3 = 2/3, x_4 = 3/5, \ldots$. Plotting these on the real line suggests $1/2 \leq x_{2n} < x_{2n+2} < x_{2n+1} < x_{2n-1} \leq 1$ for all $n \in \mathbb{N}$. This can be easily established by mathematical induction. (If $1/2 \leq x_{2n} < x_{2n+2} < x_{2n+1} < x_{2n-1} \leq 1$, then $1 + x_{2n} < 1 + x_{2n+2} < 1 + x_{2n+1} < 1 + x_{2n-1}$. Taking reciprocal and applying the recurrence relation, we have $x_{2n+1} > x_{2n+3} > x_{2n+2} > x_{2n}$. Repeating these steps once more, we get $x_{2n+2} < x_{2n+4} < x_{2n+3} < x_{2n+1}$.)

Let $I_n = [x_{2n}, x_{2n-1}]$, then $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$. Now

$$|x_m - x_{m+1}| = \left|\frac{1}{1 + x_{m-1}} - \frac{1}{1 + x_m}\right| = \frac{|x_{m-1} - x_m|}{(1 + x_{m-1})(1 + x_m)} < \frac{|x_{m-1} - x_m|}{(1 + \frac{1}{2})(1 + \frac{1}{2})} = \frac{4}{9}|x_{m-1} - x_m|.$$

Using this, we get $|x_{2n-1} - x_{2n}| < \frac{4}{9}|x_{2n-2} - x_{2n-1}| < \frac{4}{9}\frac{4}{9}|x_{2n-3} - x_{2n-2}| < \cdots < \underbrace{\frac{4}{9}\cdots\frac{4}{9}}_{2n-2}|x_1 - x_2| = \left(\frac{4}{9}\right)^{2n-2}\frac{1}{2}$.

By the sandwich theorem, $\lim_{n\to\infty}(x_{2n-1} - x_{2n}) = 0$. So by the nested interval theorem, $\bigcap_{n=1}^{\infty} I_n = \{x\}$ for some x and $\lim_{n\to\infty} x_{2n} = x = \lim_{n\to\infty} x_{2n-1}$. By the intertwining sequence theorem, $\lim_{n\to\infty} x_n = x$. So $x = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} 1/(1 + x_n) = 1/(1 + x)$. Then $x = (-1 \pm \sqrt{5})/2$. Since $x \in I_1$, $x = (-1 + \sqrt{5})/2$.

(Instead of estimating the lengths of the I_n's and squeezing them to 0 to see their intersection is a single point, we can also do the following. Let $I_n = [x_{2n}, x_{2n-1}]$ be as above so that $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$. By the nested interval theorem, $\bigcap_{n=1}^{\infty} I_n = [a, b]$, where $a = \lim_{n\to\infty} x_{2n}$ and $b = \lim_{n\to\infty} x_{2n-1}$. Taking limits on both sides of $x_{2n+1} = \frac{1}{1 + x_{2n}}$ and $x_{2n} = \frac{1}{1 + x_{2n-1}}$, we get $b = \frac{1}{1 + a}$ and $a = \frac{1}{1 + b}$. Then $b(1 + a) = 1 = a(1 + b)$, which yields $b + ab = a + ab$, so $a = b$. Hence $\bigcap_{n=1}^{\infty} I_n$ is a single point.)

Back to answering the question above in general, French mathematician Augustine Cauchy (1789–1857) introduced the following condition.

Definition. $\{x_n\}$ is a *Cauchy sequence* iff for every $\varepsilon > 0$, there is $K \in \mathbb{N}$ such that $m, n \geq K$ implies $|x_m - x_n| < \varepsilon$.

Remark. Roughly, the condition says that the terms of these sequences are getting closer and closer to each other.

Example. Let $x_n = \frac{1}{n^2}$. (Note that if $m, n \geq K$, say $m \geq n$, then we have $|x_m - x_n| = \frac{1}{n^2} - \frac{1}{m^2} < \frac{1}{n^2} \leq \frac{1}{K^2}$.) For every $\varepsilon > 0$, we can take an integer $K > \frac{1}{\sqrt{\varepsilon}}$ (by the Archimedean principle). Then $m, n \geq K$ implies $|x_m - x_n| \leq \frac{1}{K^2} < \varepsilon$. So $\{x_n\}$ is a Cauchy sequence.

Theorem. *If $\{x_n\}$ converges, then $\{x_n\}$ is a Cauchy sequence.*

Proof. For every $\varepsilon > 0$, since $\lim_{n\to\infty} x_n = x$, there is $K \in \mathbb{N}$ such that $j \geq K \Rightarrow |x_j - x| < \varepsilon/2$. For $m, n \geq K$, we have $|x_m - x_n| \leq |x_m - x| + |x - x_n| < \varepsilon/2 + \varepsilon/2 = \varepsilon$. So $\{x_n\}$ is a Cauchy sequence.

The converse of the previous theorem is true, but it takes some work to prove that. The difficulty lies primarily on how to come up with a limit of the sequence. The strategy of showing every Cauchy sequence in $\mathbb{R}$ must converge is first to find a subsequence that converges, then show that the original sequence also converge to the same limit.

Theorem. *If $\{x_n\}$ is a Cauchy sequence, then $\{x_n\}$ is bounded.*

Proof. Let $\varepsilon = 1$. Since $\{x_n\}$ is a Cauchy sequence, there is $K \in \mathbb{N}$ such that $m, n \geq K \Rightarrow |x_m - x_n| < \varepsilon = 1$. In particular, for $n \geq K$, $|x_K - x_n| < 1 \Rightarrow |x_n| = |(x_n - x_K) + x_K| \leq |x_n - x_K| + |x_K| < 1 + |x_K|$. Let $M = \max(|x_1|, \ldots, |x_{K-1}|, 1 + |x_K|)$, then for all $n \in \mathbb{N}$, $|x_n| \leq M$ (i.e. $x_n \in [-M, M]$).

Bolzano-Weierstrass Theorem. *If $\{x_n\}$ is bounded, then $\{x_n\}$ has a subsequence $\{x_{n_j}\}$ that converges.*

Proof. (Bisection Method) Let $a_1 = \inf\{x_n\}$, $b_1 = \sup\{x_n\}$ and $I_1 = [a_1, b_1]$. Let m_1 be the midpoint of I_1. If there are infinitely many terms of $\{x_n\}$ in $[a, m_1]$, then let $a_2 = a_1$, $b_2 = m_1$ and $I_2 = [a_2, b_2]$. Otherwise, there will be infinitely many terms of $\{x_n\}$ in $[m_1, b_1]$, then let $a_2 = m_1$, $b_2 = b_1$ and $I_2 = [a_2, b_2]$. For $k = 2, 3, 4, \ldots$, repeat this bisection on I_k to get I_{k+1}. We have $I_j = [a_j, b_j]$ and $I_1 \supseteq I_2 \supseteq I_3 \supseteq \ldots$. By the nested interval theorem, since $\lim_{j\to\infty}(b_j - a_j) = \lim_{j\to\infty} \frac{b_1 - a_1}{2^{j-1}} = 0$, $\bigcap_{n=1}^{\infty} I_n$ contains exactly one number x.

Take $n_1 = 1$, then $x_{n_1} = x_1 \in I_1$. Suppose n_j is chosen with $x_{n_j} \in I_j$. Since there are infinitely many terms x_n in I_{j+1}, choose $n_{j+1} > n_j$ and $x_{n_{j+1}} \in I_{j+1}$. Then $\lim_{j\to\infty} |x_{n_j} - x| \leq \lim_{j\to\infty}(b_j - a_j) = 0$. Therefore, $\lim_{j\to\infty} x_{n_j} = x$.

Remarks. In the proof, the nested interval theorem was used, so the Bolzano-Weierstrass theorem depended on the completeness axiom.

Alternate Proof. We will show every sequence $\{x_n\}$ has a monotone subsequence. (If $\{x_n\}$ is bounded, then the monotone sequence theorem will imply the subsequence converges.)

Call x_m a *peak* of $\{x_n\}$ if $x_m \geq x_k$ for all $k > m$. If $\{x_n\}$ has infinitely many peaks, then we order the peaks by strictly increasing subscripts $m_1 < m_2 < m_3 < \cdots$. By the definition of a peak, $x_{m_1} \geq x_{m_2} \geq x_{m_3} \geq \cdots$. So $\{x_{m_j}\}$ is a decreasing subsequence of $\{x_n\}$. On the other hand, if $\{x_n\}$ has only finitely many peaks $x_{m_1}, \cdots, x_{m_k}$, then let $n_1 = \max\{m_1, \cdots, m_k\} + 1$. Since x_{n_1} is not a peak, there is $n_2 > n_1$ such that $x_{n_2} > x_{n_1}$. Inductively, if x_{n_j} is not a peak, there is $n_{j+1} > n_j$ with $x_{n_{j+1}} > x_{n_j}$. So $\{x_{n_j}\}$ is a strictly increasing subsequence of $\{x_n\}$.

Remarks. This alternate proof used the monotone sequence theorem, so it also depended on the completeness axiom.

Cauchy's Theorem. *$\{x_n\}$ converges if and only if $\{x_n\}$ is a Cauchy sequence.*

Proof. The 'only if' part was proved. For the 'if' part, since $\{x_n\}$ is a Cauchy sequence, $\{x_n\}$ is bounded. By the Bolzano-Weierstrass theorem, $\{x_n\}$ has a subsequence $\{x_{n_j}\}$ that converges in $\mathbb{R}$, say $\lim_{j\to\infty} x_{n_j} = x$.

We will show $\lim_{n\to\infty} x_n = x$. For every $\varepsilon > 0$, since $\{x_n\}$ is a Cauchy sequence, there is $K_1 \in \mathbb{N}$ such that $m, n \geq K_1 \Rightarrow |x_m - x_n| < \varepsilon/2$. Since $\lim_{j\to\infty} x_{n_j} = x$, there is $K_2 \in \mathbb{N}$ such that $j \geq K_2 \Rightarrow |x_{n_j} - x| < \varepsilon/2$. If $n \geq J = \max(K_1, K_2)$, then $n_J \geq J \geq K_1$, $J \geq K_2$ and $|x_n - x| \leq |x_n - x_{n_J}| + |x_{n_J} - x| < \varepsilon/2 + \varepsilon/2 = \varepsilon$.

Example. Does the sequence $\{x_n\}$ converge, where $x_1 = \sin 1$ and $x_k = x_{k-1} + \dfrac{\sin k}{k^2}$ for $k = 2, 3, 4, \ldots$?

We will check the Cauchy condition. For $m > n$, $x_m - x_n = \sum_{k=n+1}^{m} (x_k - x_{k-1}) = \sum_{k=n+1}^{m} \dfrac{\sin k}{k^2}$ and

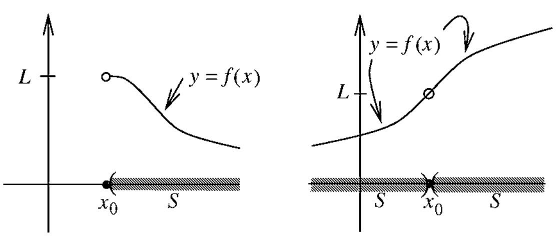
$$|x_m - x_n| \leq \frac{1}{(n+1)^2} + \cdots + \frac{1}{m^2} < \frac{1}{n(n+1)} + \cdots + \frac{1}{(m-1)m} = \left(\frac{1}{n} - \frac{1}{n+1}\right) + \cdots + \left(\frac{1}{m-1} - \frac{1}{m}\right) = \frac{1}{n} - \frac{1}{m} < \frac{1}{n}.$$

So for $\varepsilon > 0$, by the Archimedean principle, there is $K > 1/\varepsilon$. Then $m, n \geq K \Rightarrow |x_m - x_n| < \frac{1}{K} < \varepsilon$. Therefore, by the Cauchy theorem, $\{x_n\}$ converges.

Limits of Functions

Let S be an interval (or more generally a set). Suppose $f\colon S \to \mathbb{R}$ is a function. We would like to define $\lim_{x\to x_0} f(x)$. For this to be a meaningful expression, in the interval case, x_0 must be a point on the interval or an endpoint. In the general case, x_0 must be "approachable" by the points of S.

CONVENTION. In discussing the limit of a function $f : S \to \mathbb{R}$ at x_0, x_0 is always assumed to be the limit of some sequence in $S \setminus \{x_0\}$ so that x_0 can be *approached* by points of S. (We say x_0 is an *accumulation* (or *limit* or *cluster*) *point* of S iff x_0 is the limit of a sequence $\{x_n\}$ in $S \setminus \{x_0\}$.)



Let $f : S \to \mathbb{R}$ be a function. To say $\lim_{x\to x_0} f(x) = L$ roughly means the distance between $f(x)$ and L is *as small as we please* when $x \in S$ is *sufficiently close* to x_0.(Again, *small* and *close* need to be clarified.)

Example. Let $f(x) = (x^3 - 3x^2)/(x - 3)$. If x gets close to 3 and $x \neq 3$, then $f(x) = x^2$ should be close to 9. In other words, the distance $d\big(f(x), 9\big) = |f(x) - 9|$ goes to 0 when the distance $d(x, 3) = |x - 3|$ gets small. So for every positive ε, the distance $|f(x)-9|$ will soon or later be less than ε when the distance $|x-3|$ becomes small enough. For $\varepsilon = 0.1$, how small $d(x,3)$ be (that is, for what x) will make $d\big(f(x),9\big) = |f(x)-9| < \varepsilon$? Equivalently, we are seeking a positive δ so that $0 < |x-3| < \delta \Rightarrow |f(x)-9| < \varepsilon$.

Now $|f(x)-9| < \varepsilon \Leftrightarrow 8.9 < f(x) < 9.1$. So $2.9833 \approx \sqrt{8.9} < x < \sqrt{9.1} \approx 3.0166$ and $-0.0167 < x-3 < 0.0166$. If we take $\delta = 0.016$, then $0 < |x-3| < \delta \Rightarrow |f(x)-9| < \varepsilon$.

(For small $\varepsilon > 0$, $|f(x)-9| < \varepsilon \Leftrightarrow 9-\varepsilon < f(x) < 9+\varepsilon \Rightarrow \sqrt{9-\varepsilon}-3 < x-3 < \sqrt{9+\varepsilon}-3$. So we may take $\delta = \min(\sqrt{9+\varepsilon}-3, 3-\sqrt{9-\varepsilon})$, then $0 < |x-3| < \delta \Rightarrow |f(x)-9| < \varepsilon$.)

Following the example, we are ready to state the precise definition of the limit of a function.

Definition. Let $f : S \to \mathbb{R}$ be a function. We say $f(x)$ *converges to* L (or has *limit* L) as x tends to x_0 in S iff for every $\varepsilon > 0$, there is $\delta > 0$ such that for every $x \in S$, $0 < |x-x_0| < \delta$ implies $|f(x)-L| < \varepsilon$. This is denoted by $\lim\limits_{\substack{x\to x_0\\ x\in S}} f(x) = L$ (or $\lim\limits_{x\to x_0} f(x) = L$ in short.)

In the definition, δ depends on ε and x_0. For different ε (or different x_0), δ will be different. If a limit value exists, then it is unique. The proof is similar to the sequential case and is left as an exercise for the readers.

Examples. (1) For $g : [0,\infty) \to \mathbb{R}$ defined by $g(x) = \sqrt{x}$, show that $\lim\limits_{x\to 0} g(x) = 0$ and $\lim\limits_{x\to 4} g(x) = 2$ by checking the ε-δ definition.

Solution. For every $\varepsilon > 0$, let $\delta = \varepsilon^2$. Then for every $x \in [0,\infty)$, $0 < |x-0| < \delta$ implies $|g(x)-0| = \sqrt{x} < \sqrt{\delta} = \varepsilon$. This takes care of the checking for the first limit.

For the second limit, note that $|\sqrt{x}-2| = \dfrac{|x-4|}{\sqrt{x}+2} \le \dfrac{|x-4|}{2}$. For every $\varepsilon > 0$, let $\delta = 2\varepsilon$. Then for every $x \in [0,\infty)$, $0 < |x-4| < \delta$ implies $|g(x)-2| \le \dfrac{|x-4|}{2} < \dfrac{\delta}{2} = \varepsilon$.

---

(2) Let $f : \mathbb{R}\setminus\{0\} \to \mathbb{R}$ be defined by $f(x) = \dfrac{1}{5x}$. Show that $\lim\limits_{x\to 2} f(x) = \dfrac{1}{10}$ by checking the ε-δ definition.

Solution. (Note that $\left|\dfrac{1}{5x} - \dfrac{1}{10}\right| = \dfrac{|x-2|}{10x} \le \dfrac{|x-2|}{10}$ for $x \in (1,3)$.) For every $\varepsilon > 0$, take $\delta = \min(1, 10\varepsilon)$. Then $\delta \le 1$ and $\delta \le 10\varepsilon$. For every $x \in \mathbb{R}\setminus\{0\}$, $0 < |x-2| < \delta$ implies $x \in (1,3)$. So $\left|f(x) - \dfrac{1}{10}\right| \le \dfrac{|x-2|}{10} < \dfrac{\delta}{10} \le \varepsilon$ and we are done.

Notation: We will write $x_n \to x_0$ in $S\setminus\{x_0\}$ to mean sequence $\{x_n\}$ in $S\setminus\{x_0\}$ converges to x_0.

Sequential Limit Theorem. $\lim\limits_{\substack{x\to x_0\\ x\in S}} f(x) = L$ *if and only if for every* $x_n \to x_0$ *in* $S\setminus\{x_0\}$, $\lim\limits_{n\to\infty} f(x_n) = L$.

In quantifier symbols ($\forall$ = for any, for every, for all, $\exists$ = there is, there exists),

$$\lim_{\substack{x\to x_0\\ x\in S}} f(x) = L \iff \forall\varepsilon > 0,\ \exists\delta > 0 \text{ such that } \forall x \in S, 0 < |x-x_0| < \delta \Rightarrow |f(x)-L| < \varepsilon,$$

$$\lim_{\substack{x\to x_0\\ x\in S}} f(x) \ne L \iff \exists\varepsilon > 0 \text{ such that } \forall\delta > 0,\ \exists x \in S, 0 < |x-x_0| < \delta \text{ and } |f(x)-L| \ge \varepsilon.$$

Proof. If $\lim\limits_{\substack{x\to x_0\\ x\in S}} f(x) = L$, then for every $\varepsilon > 0$, there is $\delta > 0$ such that for $x \in S$, $0 < |x-x_0| < \delta \Rightarrow |f(x)-L| < \varepsilon$. If $\lim\limits_{n\to\infty} x_n = x_0$ and $x_n \ne x_0$, then there is $K \in \mathbb{N}$ such that $n \ge K \Rightarrow 0 < |x_n - x_0| < \delta (\Rightarrow |f(x_n)-L| < \varepsilon)$. So $\lim\limits_{n\to\infty} f(x_n) = L$.

Conversely, if $\lim_{\substack{x\to x_0\\ x\in S}} f(x) \neq L$, then there is $\varepsilon > 0$ such that for every $\delta > 0$, there is $x \in S$ with $0 < |x - x_0| < \delta$ and $|f(x) - L| \geq \varepsilon$. Now, setting $\delta = \frac{1}{n}$, there is $x_n \in S$ with $0 < |x_n - x_0| < \delta = \frac{1}{n}$ and $|f(x_n) - L| \geq \varepsilon$. By the sandwich theorem, $x_n \to x_0$ in $S \setminus \{x_0\}$. So $\lim_{n\to\infty} f(x_n) = L$. Then $0 = \lim_{n\to\infty} |f(x_n) - L| \geq \varepsilon$, a contradiction.

Remarks. If $\lim_{n\to\infty} f(x_n)$ exists for *every* $x_n \to x_0$ in $S \setminus \{x_0\}$, then all the limit values are the same. To see this, suppose $x_n \to x_0$ and $w_n \to x_0$ in $S \setminus \{x_0\}$. Then the intertwining sequence $\{z_n\} = \{x_1, w_1, x_2, w_2, x_3, w_3, \ldots\}$ converges to x_0 in $S \setminus \{x_0\}$. Since $\{f(x_n)\}$ and $\{f(w_n)\}$ are subsequences of the convergent sequence $\{f(z_n)\}$, they have the same limit.

Consequences of Sequential Limit Theorem.

(1) (*Computation Formulas*) If $f, g : S \to \mathbb{R}$ are functions, $\lim_{\substack{x\to x_0\\ x\in S}} f(x) = L_1$ and $\lim_{\substack{x\to x_0\\ x\in S}} g(x) = L_2$, then

$$\lim_{\substack{x\to x_0\\ x\in S}} \Big(f(x) \begin{Bmatrix} + \\ - \\ \cdot \\ / \end{Bmatrix} g(x)\Big) = L_1 \begin{Bmatrix} + \\ - \\ \cdot \\ / \end{Bmatrix} L_2 = \lim_{\substack{x\to x_0\\ x\in S}} f(x) \begin{Bmatrix} + \\ - \\ \cdot \\ / \end{Bmatrix} \lim_{\substack{x\to x_0\\ x\in S}} g(x)$$

respectively (in the case of division, provided $g(x) \neq 0$ for every $x \in S$ and $L_2 \neq 0$).

Proof. Since $f(x)$ and $g(x)$ have limits L_1 and L_2, respectively, as x tends to x_0 in S, by the sequential limit theorem, $f(x_n)$ and $g(x_n)$ will have limits L_1 and L_2, respectively, for every $x_n \to x_0$ in $S \setminus \{x_0\}$. By the computation formulas for sequences, the limit of $f(x_n) + g(x_n)$ is $L_1 + L_2$ for every $x_n \to x_0$ in $S \setminus \{x_0\}$. By the sequential limit theorem, $f(x) + g(x)$ has limit $L_1 + L_2$ as x tends to x_0 in S. Replacing $+$ by $-, \cdot, /$, we get the proofs for the other parts.

Alternate Proof. If $\lim_{\substack{x\to x_0\\ x\in S}} f(x) = L_1$ and $\lim_{\substack{x\to x_0\\ x\in S}} g(x) = L_2$, then for every $\varepsilon > 0$, there are $\delta_1 > 0$ such that for every $x \in S$, $0 < |x - x_0| < \delta_1$ implies $|f(x) - L_1| < \dfrac{\varepsilon}{2}$ and $\delta_2 > 0$ such that for every $x \in S$, $0 < |x - x_0| < \delta_2$ implies $|g(x) - L_2| < \dfrac{\varepsilon}{2}$. Now we take $\delta = \min(\delta_1, \delta_2) > 0$ so that $\delta \leq \delta_1$ and $\delta \leq \delta_2$. Then, for every $x \in S$, $0 < |x - x_0| < \delta$ implies both $0 < |x - x_0| < \delta_1$ and $0 < |x - x_0| < \delta_2$ so that

$$\big|\big(f(x) + g(x)\big) - (L_1 + L_2)\big| = |(f(x) - L_1) + (g(x) - L_2)| \leq |f(x) - L_1| + |g(x) - L_2| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

The other parts of the computation formulas can be proved by adapting the arguments for the sequential case.

(2) (*Sandwich Theorem or Squeeze Limit Theorem*) If $f(x) \leq g(x) \leq h(x)$ for every $x \in S$ and $\lim_{\substack{x\to x_0\\ x\in S}} f(x) = L = \lim_{\substack{x\to x_0\\ x\in S}} h(x)$, then $\lim_{\substack{x\to x_0\\ x\in S}} g(x) = L$.

(3) (*Limit Inequality*) If $f(x) \geq 0$ for all $x \in S$ and $\lim_{\substack{x\to x_0\\ x\in S}} f(x) = L$, then $L \geq 0$.

The proofs of (2) and (3) can be done by switching to sequences like the first proof of (1) or by adapting the arguments of the sequential cases and checking the ε-δ definition like the alternate proof of (1).

Next we will discuss one-sided limits.

Definitions. For $f : (a, b) \to \mathbb{R}$ and $x_0 \in (a, b)$, the *left hand limit* of f at x_0 is $f(x_0-) = \lim_{x\to x_0^-} f(x) = \lim_{\substack{x\to x_0\\ x\in(a,x_0)}} f(x)$. The *right hand limit* of f at x_0 is $f(x_0+) = \lim_{x\to x_0^+} f(x) = \lim_{\substack{x\to x_0\\ x\in(x_0,b)}} f(x)$.

Theorem. *For $x_0 \in (a, b)$, $\lim_{\substack{x\to x_0\\ x\in(a,b)}} f(x) = L$ if and only if $f(x_0-) = L = f(x_0+)$.*

Proof. If $\lim_{x\to x_0} f(x) = L$, then for every $\varepsilon > 0$, there is $\delta > 0$ such that for $x \in (a,b)$, $0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$. In particular, for $x \in (a, x_0)$, $0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$. So $f(x_0-) = L$. Similarly, for $x \in (x_0, b)$, $0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$. So $f(x_0+) = L$.

Conversely, if $f(x_0-) = L = f(x_0+)$, then for every $\varepsilon > 0$, there is $\delta_1 > 0$ such that for $x \in (a, x_0)$, $0 < |x-x_0| < \delta_1 \Rightarrow |f(x)-L| < \varepsilon$ and there is $\delta_2 > 0$ such that for $x \in (x_0, b)$, $0 < |x-x_0| < \delta_2 \Rightarrow |f(x)-L| < \varepsilon$. As $\delta = \min(\delta_1, \delta_2) \le \delta_1$ and δ_2, we have for $x \in (a,b)$, $0 < |x - x_0| < \delta \Rightarrow |f(x) - L| < \varepsilon$. So $\lim_{x\to x_0} f(x) = L$.

Definitions. A function $f\colon S \to \mathbb{R}$ is $\left\{\begin{array}{c} \textit{increasing} \\ \textit{decreasing} \\ \textit{strictly increasing} \\ \textit{strictly decreasing} \end{array}\right\}$ on S iff for every $x, y \in S$, $x < y \Rightarrow \left\{\begin{array}{c} f(x) \le f(y) \\ f(x) \ge f(y) \\ f(x) < f(y) \\ f(x) > f(y) \end{array}\right\}$.
Also, f is $\left\{\begin{array}{c} \textit{monotone} \\ \textit{strictly monotone} \end{array}\right\}$ on S iff f is $\left\{\begin{array}{c} \text{increasing or decreasing} \\ \text{strictly increasing or strictly decreasing} \end{array}\right\}$ on S, respectively.

For a nonempty subset S_0 of S, we say f is $\left\{\begin{array}{c} \textit{bounded above} \\ \textit{bounded below} \\ \textit{bounded} \end{array}\right\}$ on S_0 iff $\{f(x) : x \in S_0\}$ is $\left\{\begin{array}{c} \text{bounded above} \\ \text{bounded below} \\ \text{bounded} \end{array}\right\}$, respectively. If The set S_0 is not mentioned, then it is the domain S.

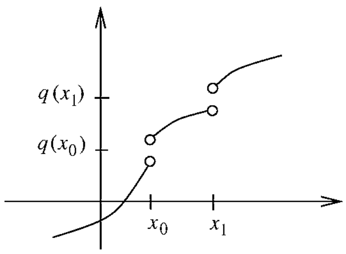
For monotone functions, the following theorem is analogous to the monotone sequence theorem. It will be used in the next chapter to prove the continuous inverse theorem, which will be used to prove the $\dfrac{dx}{dy} = 1\Big/\dfrac{dy}{dx}$ rule later. Also, it will be used again in the chapter on integration.

Monotone Function Theorem. *If f is increasing on (a,b), then for every $x_0 \in (a,b)$, $f(x_0-) = \sup\{f(x) : a < x < x_0\}$ and $f(x_0+) = \inf\{f(x) : x_0 < x < b\}$ and $f(x_0-) \le f(x_0) \le f(x_0+)$. If f is bounded below, then $f(a+) = \inf\{f(x) : a < x < b\}$. If f is bounded above, then $f(b-) = \sup\{f(x) : a < x < b\}$. Also f has countably many discontinuous points on (a,b), i.e. $J = \{x_0 : x_0 \in (a,b), f(x_0-) \ne f(x_0+)\}$ is countable. (The theorem is similarly true for decreasing functions and all other kinds of intervals.)*

Proof. If $a < x < x_0 < b$, then $f(x) \le f(x_0)$. So $M = \sup\{f(x) : a < x < x_0\} \le f(x_0)$. By the supremum property, for every $\varepsilon > 0$, there is $c \in (a, x_0)$ such that $M - \varepsilon < f(c) \le M$. If we let $\delta = x_0 - c$, then

$$\forall x \in (a, x_0), 0 < |x - x_0| < \delta \Rightarrow c = x_0 - \delta < x < x_0 \Rightarrow f(c) \le f(x) \le M \Rightarrow |f(x) - M| \le M - f(c) < \varepsilon.$$

So $\lim_{x\to x_0^-} f(x) = M \le f(x_0)$. Similarly, $f(x_0) \le \lim_{x\to x_0^+} f(x) = \inf\{f(x) : x_0 < x < b\}$. In the case f is bounded below or above, the proof of the existence of $f(a+)$ or $f(b-)$ is similar.



Next let $x_0 \in (a,b)$ with $f(x_0-) < f(x_0+)$. By the density of rational numbers, we may choose a $q(x_0) \in \mathbb{Q}$ between $f(x_0-)$ and $f(x_0+)$. The function $q\colon J \to \mathbb{Q}$ is injective because if f is discontinuous at x_0, x_1 with $x_0 < x_1$, then

$$q(x_0) < f(x_0+) \le f\Big(\frac{x_0 + x_1}{2}\Big) \le f(x_1-) < q(x_1).$$

By the injection theorem, J is countable, i.e. f has countably many discontinuous points on (a,b). The cases of decreasing functions or other kinds of intervals are similar.

Appendix: Infinite Limits and Limit at Infinity

We begin with a definition of a sequence having $+\infty$ as limit. In this case, the sequence does not have any upper bound in $\mathbb{R}$, i.e. the sequence will pass any fixed $r \in \mathbb{R}$ eventually and keep on going. Sequences with $-\infty$ as limit and functions with $\pm\infty$ limit are defined similarly.

Definitions. (1) A sequence $\{x_n\}$ *diverges* to $+\infty$ (or has *limit* $+\infty$) iff for every $r \in \mathbb{R}$, there is $K \in \mathbb{N}$ (depending on r) such that $n \geq K$ implies $x_n > r$. Similarly, a sequence $\{x_n\}$ *diverges* to $-\infty$ (or has *limit* $-\infty$) iff for every $r \in \mathbb{R}$, there is $K \in \mathbb{N}$ (depending on r) such that $n \geq K$ implies $x_n < r$.

(2) A function $f : S \to \mathbb{R}$ *diverges* to $+\infty$ (or has *limit* $+\infty$) as x tends to x_0 in S iff for every $r \in \mathbb{R}$, there is $\delta > 0$ (depending on r and x_0) such that for every $x \in S$, $0 < |x - x_0| < \delta$ implies $f(x) > r$. Similarly, a function $f : S \to \mathbb{R}$ *diverges* to $-\infty$ (or has *limit* $-\infty$) as x tends to x_0 in S iff for every $r \in \mathbb{R}$, there is $\delta > 0$ (depending on r and x_0) such that for every $x \in S$, $0 < |x - x_0| < \delta$ implies $f(x) < r$.

Exercise. Prove that if $\{x_n\}$ is an increasing sequence, then either the limit of $\{x_n\}$ exists (as a real number) or the limit is $+\infty$. (The decreasing case is similar with $+\infty$ replaced by $-\infty$.)

Limit at infinity for functions are defined similarly as for sequences as follow.

Definitions. (1) Let $f : S \to \mathbb{R}$ be a function such that $+\infty$ is an accumulation point of S (i.e. there is a sequence in S diverges to $+\infty$.) We say $f(x)$ *converges to* L (or has *limit* L) as x tends to $+\infty$ in S iff for every $\varepsilon > 0$, there is $K \in \mathbb{R}$ such that for every $x \in S$, $x \geq K$ implies $|f(x) - L| < \varepsilon$. Similarly, let $f : S \to \mathbb{R}$ be a function such that $-\infty$ is an accumulation point of S (i.e. there is a sequence in S diverges to $-\infty$.) We say $f(x)$ *converges to* L (or has *limit* L) as x tends to $-\infty$ in S iff for every $\varepsilon > 0$, there is $K \in \mathbb{R}$ such that for every $x \in S$, $x \leq K$ implies $|f(x) - L| < \varepsilon$.

(2) Let $f : S \to \mathbb{R}$ be a function such that $+\infty$ is an accumulation point of S (i.e. there is a sequence in S diverges to $+\infty$.) We say $f(x)$ *diverges to* $+\infty$ (or has *limit* $+\infty$) as x tends to $+\infty$ in S iff for every $r \in \mathbb{R}$, there is $K \in \mathbb{R}$ such that for every $x \in S$, $x \geq K$ implies $f(x) > r$. Similarly, let $f : S \to \mathbb{R}$ be a function such that $-\infty$ is an accumulation point of S (i.e. there is a sequence in S diverges to $-\infty$.) We say $f(x)$ *diverges to* $+\infty$ (or has *limit* $+\infty$) as x tends to $-\infty$ in S iff for every $r \in \mathbb{R}$, there is $K \in \mathbb{R}$ such that for every $x \in S$, $x \leq K$ implies $f(x) > r$. By replacing $f(x) > r$ with $f(x) < r$, we get the definitions of limit equal to $-\infty$ as x tends to $\pm\infty$ in S.

It is good exercises for the readers to formulate the computation formulas and properties for these limits, which can be proved by checking these definitions just like the finite cases.

Chapter 7. Continuity

Definitions. A function $f\colon S \to \mathbb{R}$ is *continuous* at $x_0 \in S$ iff $\lim_{\substack{x\to x_0\\ x\in S}} f(x) = f(x_0)$, i.e. for every $\varepsilon > 0$, there is $\delta > 0$ such that for all $x \in S$, $|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon$. For $E \subseteq S$, we say f is *continuous* on E iff f is continuous at every element of E. Also, we say f is *continuous* iff f is continuous on the domain S.

Sequential Continuity Theorem. *$f : S \to \mathbb{R}$ is continuous at $x_0 \in S$ if and only if for every $x_n \to x_0$ in S, $\lim_{n\to\infty} f(x_n) = f(x_0) = f(\lim_{n\to\infty} x_n)$.*

Proof. Just replace L by $f(x_0)$, $0 < |x - x_0| < \delta$ by $|x - x_0| < \delta$ and $x_n \to x_0$ in $S \setminus \{x_0\}$ by $x_n \to x_0$ in S (i.e. delete the $x_n \neq x_0$ requirement) in the proof of the sequential limit theorem.

Example. It is easy to give examples of continuous functions, such as polynomials. Here is an example of a function *not* continuous at any point. Let $f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$, then f is discontinuous at every $x \in \mathbb{R}$!

Reason. For every $x_0 \in \mathbb{R}$, $n \in \mathbb{N}$, by the density of rational numbers and irrational numbers, there are $r_n \in \mathbb{Q}$, $s_n \notin \mathbb{Q}$ in $(x_0 - \frac{1}{n}, x_0)$. Now $r_n \to x_0$, $s_n \to x_0$, but $\lim_{n\to\infty} f(r_n) = 1$ and $\lim_{n\to\infty} f(s_n) = 0$. So $\lim_{x\to x_0} f(x)$ cannot exist.

Theorem. *If $f, g : S \to \mathbb{R}$ are continuous at $x_0 \in S$, then $f \pm g$, fg, f/g (provided $g(x_0) \neq 0$) are continuous at x_0.*

Proof. Since f, g are continuous at x_0, $\lim_{x\to x_0} (f + g)(x) = \lim_{x\to x_0} f(x) + \lim_{x\to x_0} g(x) = f(x_0) + g(x_0) = (f + g)(x_0)$. So $f + g$ is continous at x_0 by definition. The subtraction, multiplication and division cases are similar.

Theorem. *If $f : S \to \mathbb{R}$ is continuous at x_0, $f(S) \subseteq S'$, $g : S' \to \mathbb{R}$ is continuous at $f(x_0)$, then $g \circ f$ is continuous at x_0.*

Proof. By the sequential continuity theorem, all we need to show is that $\lim_{n\to\infty} (g \circ f)(x_n) = (g \circ f)(x_0)$ for every sequence $\{x_n\}$ converging to x_0. Since f is continuous at x_0, by the sequential continuity theorem, the limit of $f(x_n)$ is $f(x_0)$. Since g is continuous at $f(x_0)$, so $\lim_{n\to\infty} g(f(x_n)) = g(\lim_{n\to\infty} f(x_n)) = g(f(x_0))$.

In the discussions below, S will denote an interval of positive length.

Theorem (Sign Preserving Property). *If $g\colon S \to \mathbb{R}$ is continuous and $g(x_0) > 0$, then there is an interval $I = (x_0 - \delta, x_0 + \delta)$ with $\delta > 0$ such that $g(x) > 0$ for every $x \in S \cap I$. (The case $g(x_0) < 0$ is similar by considering $-g$.)*

Proof. Let $\varepsilon = g(x_0) > 0$. Since g is continuous at x_0, there is $\delta > 0$ such that for $x \in S$, $|x - x_0| < \delta \Rightarrow |g(x) - g(x_0)| < \varepsilon$. So $x \in S \cap (x_0 - \delta, x_0 + \delta) \Rightarrow 0 = g(x_0) - \varepsilon < g(x)$.

Intermediate Value Theorem. *If $f\colon [a, b] \to \mathbb{R}$ is continuous and y_0 is between $f(a)$ and $f(b)$ inclusive, then there is (at least one) $x_0 \in [a, b]$ such that $f(x_0) = y_0$.*

Proof. The cases $y_0 = f(a)$ or $y_0 = f(b)$ are trivial as $x_0 = a$ or b will do. We assume $f(a) < y_0 < f(b)$. (The other possibility $f(a) > y_0 > f(b)$ is similar.) The set $S = \{x \in [a, b]\colon f(x) \le y_0\}$ is nonempty ($a \in S$) and has b as an upper bound. Then $x_0 = \sup S \in [a, b]$.

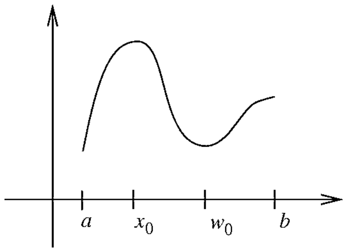
By the supremum limit theorem, there is a sequence $\{x_n\}$ in S such that $\lim_{n\to\infty} x_n = x_0$. Note $x_n \in [a, b]$ implies $x_0 \in [a, b]$. Then $f(x_0) = \lim_{n\to\infty} f(x_n) \le y_0$ by continuity of f at x_0. Assume $f(x_0) < y_0$. Then $x_0 \neq b$ as $y_0 < f(b)$. Define $g(x) = y_0 - f(x)$ on $[a, b]$. Since $g(x_0) > 0$, by the sign preserving property, there is $x_1 > x_0$ such that $g(x_1) > 0$. Then $f(x_1) < y_0$. So $x_0 < x_1 \in S$, which contradicts $x_0 = \sup S$. Therefore, $f(x_0) = y_0$.

Examples. (1) The equation $x^5 + 3x + \sin x = \cos x + 10$ has a solution. To see this, let $f(x) = x^5 + 3x + \sin x - \cos x - 10$, then $f(0) = -11$ and $26 \le f(2) \le 30$. Since f is continuous, the intermediate value theorem implies $f(x) = 0$ for some $x \in (0, 2)$.

(2) Suppose $p(x) = x^n + a_1x^{n-1} + \cdots + a_n$ with n odd. Let $x_0 = 1 + |a_1| + \cdots + |a_n| \geq 1$, then we have $p(x_0) = x_0^n + a_1x_0^{n-1} + \cdots + a_n$ and $p(-x_0) = -x_0^n + a_1x_0^{n-1} - \cdots + a_n$. So $x_0^n - p(x_0) = -a_1x_0^{n-1} - \cdots - a_n$ and $x_0^n + p(-x_0) = a_1x_0^{n-1} - \cdots + a_n$. By the triangle inequality,

$$\left.\begin{matrix} x_0^n - p(x_0) \\ x_0^n + p(-x_0) \end{matrix}\right\} \leq |a_1x_0^{n-1}| + \cdots + |a_n| \leq |a_1||x_0^{n-1} + \cdots + |a_n||x_0^{n-1} = (|a_1| + \cdots + |a_n|)x_0^{n-1} < x_0^n.$$

Then $p(x_0) > 0$ and $p(-x_0) < 0$. By the intermediate value theorem, there is a root of $p(x)$ between $-x_0$ and x_0.



Extreme Value Theorem. *If $f\colon [a,b] \to \mathbb{R}$ is continuous, then there are x_0, $w_0 \in [a,b]$ such that $f(w_0) \leq f(x) \leq f(x_0)$ for every $x \in [a,b]$. So the range of f is $f\big([a,b]\big) = [f(w_0), f(x_0)]$. In particular, f is bounded on $[a,b]$*

In this case, we write $f(x_0) = \sup\{f(x) : x \in [a,b]\} = \max\limits_{x\in[a.b]} f(x)$ and $f(w_0) = \inf\{f(x) : x \in [a,b]\} = \min\limits_{x\in[a.b]} f(x)$.

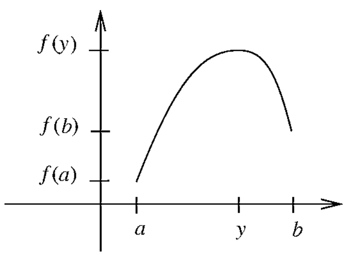
Proof. We first show $f([a,b]) = \{f(x) : x \in [a,b]\}$ is bounded above. Assume it is not bounded above. Then each $n \in \mathbb{N}$ is not an upper bound. So there is $z_n \in [a,b]$ such that $f(z_n) > n$. By the Bolzano-Weierstrass theorem, $\{z_n\}$ in $[a,b]$ has a subsequence $\{z_{n_j}\}$ converging to some $z_0 \in [a,b]$. Since f is continuous at z_0, $\lim\limits_{j\to\infty} f(z_{n_j}) = f(z_0)$, which implies $\{f(z_{n_j})\}$ is bounded by the boundedness theorem. However, $f(z_{n_j}) > n_j \geq j$ implies $\{f(z_{n_j})\}$ is not bounded, a contradiction.

By the completeness axiom, $M = \sup\{f(x) : x \in [a,b]\}$ exists. By the supremum limit theorem, there is a sequence $\{x_n\}$ in $[a,b]$ such that $\lim\limits_{n\to\infty} f(x_n) = M$. By the Bolzano-Weierstrass theorem, $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ such that $\lim\limits_{k\to\infty} x_{n_k} = x_0$ in $[a,b]$. Since f is continuous at $x_0 \in [a,b]$, $M = \lim\limits_{k\to\infty} f(x_{n_k}) = f(x_0)$ by the subsequence theorem and the sequential continuity theorem, respectively.

Similarly, $\inf\{f(x) : x \in [a,b]\} = f(w_0)$ for some $w_0 \in [a,b]$.

The following two theorems are for explaining the $\dfrac{dx}{dy} = 1\Big/\dfrac{dy}{dx}$ rule in the next chapter.

Continuous Injection Theorem. *If f is continuous and injective on $[a,b]$, then f is strictly monotone on $[a,b]$ and $f\big([a,b]\big) = [f(a), f(b)]$ or $[f(b), f(a)]$. (This is true for any nonempty interval in place of $[a,b]$. The range is an interval with $f(a+)$, $f(b-)$ as endpoints.)*



Proof. Since f is injective, either $f(a) < f(b)$ or $f(a) > f(b)$. Suppose $f(a) < f(b)$. Let $y \in (a,b)$. Then $f(y)$ cannot be greater than $f(b)$, otherwise by the intermediate value theorem, there is $w \in (a,y)$ such that $f(w) = f(b)$, contradicting injectivity. Similarly, $f(y)$ cannot be less than $f(a)$. Therefore, $f(a) < f(y) < f(b)$. If $a \leq x < y \leq b$, then similarly $f(a) \leq f(x) < f(y) \leq f(b)$, i.e. f is strictly increasing on $[a,b]$ and $f\big([a,b]\big) = [f(a), f(b)]$ by the intermediate value theorem.

The case $f(a) > f(b)$ leads to f strictly decreasing on $[a,b]$.

Continuous Inverse Theorem. *If f is continuous and injective on $[a,b]$, then f^{-1} is continuous on $f\big([a,b]\big)$.*

Proof. By the last theorem, f is strictly monotone on $[a,b]$. We first suppose f is strictly increasing. Then f^{-1} is also strictly increasing on $f\big([a,b]\big) = [f(a), f(b)]$.

For $y_0 \in (f(a), f(b)]$, let $r = f^{-1}(y_0-) = \lim\limits_{y\to y_0^-} f^{-1}(y)$, which exists by the monotone function theorem. Since $f^{-1}(y) \in [a,b]$, $r \in [a,b]$. Let $\{y_n\}$ be a sequence in $(f(a), y_0)$ converging to y_0, then $w_n = f^{-1}(y_n)$ will converge to r by sequential limit theorem. Since f is continuous at r, by the sequential continuity theorem, $f(r) = \lim\limits_{n\to\infty} f(w_n) = \lim\limits_{n\to\infty} f(f^{-1}(y_n)) = y_0$. Then $f^{-1}(y_0) = r = f^{-1}(y_0-)$. Similarly, if $y_0 \in [f(a), f(b))$, then $f^{-1}(y_0+) = f^{-1}(y_0)$. So f^{-1} is continuous on $[f(a), f(b)] = f\big([a,b]\big)$. The case f is strictly decreasing is similar.

Appendix: Fundamental Theorem of Algebra

The extreme value theorem is often used in estimating integrals. More precisely, if $f : [a, b] \to \mathbb{R}$ is continuous, then for $m = \min_{x\in[a,b]} f(x)$ and $M = \max_{x\in[a,b]} f(x)$, we have $m \le f(x) \le M$ and $m(b-a) \le \int_a^b f(x)\,dx \le M(b-a)$.

Here we will mention that there is a version of the extreme value theorem for continuous functions defined on closed disks of finite radius on the plane. As an application of this fact, we can sketch a proof of *the fundamental theorem of algebra*, which asserts that every nonconstant polynomial with complex coefficients must have a root.

Let $p(z) = z^n + a_1 z^{n-1} + \cdots + a_n$ and $m = \inf\{|p(z)| : z \in \mathbb{C}\}$. We have

$$|z^n| = \Big|p(z) - \sum_{k=1}^{n} a_k z^{n-k}\Big| \le |p(z)| + \sum_{k=1}^{n} |a_k||z|^{n-k} \Rightarrow |p(z)| \ge |z|^n - \sum_{k=1}^{n} |a_k||z|^{n-k} = |z|^n \underbrace{\Big(1 - \sum_{k=1}^{n} |a_k||z|^{-k}\Big)}_{\to 1- \text{ as } |z|\to\infty}.$$

So for $|z|$ large, $1 - \sum_{k=1}^{n} |a_k||z|^{-k} \ge \frac{1}{2}$. For a very large $K > \sqrt[n]{2m}$, we have $|z| > K$ implies

$$|p(z)| \ge |z|^n\Big(1 - \sum_{k=1}^{n} |a_k||z|^{-k}\Big) \ge \frac{1}{2}|z|^n > \frac{1}{2}K^n > m.$$

Let D_K be the closed disk $\{z : |z| \le K\}$. We have $m = \inf\{|p(z)| : z \in \mathbb{C}\} = \inf\{|p(z)| : z \in D_K\} = |p(z_0)|$ for some $z_0 \in D_K$ by the extreme value theorem, since $|p(z)|$ is continuous on D_K.

We claim $m = |p(z_0)| = 0$. Suppose $m \ne 0$. Then $f(z) = p(z + z_0)/p(z_0)$ is a polynomial of degree n and $f(0) = 1$. So $f(z) = 1 + b_1 z + \cdots + b_n z^n$ for some $b_1, \ldots, b_n$ with $b_n \ne 0$. Let k be the smallest positive integer such that $b_k \ne 0$. Then $f(z) = 1 + b_k z^k + \cdots + b_n z^n$. Since $|p(z + z_0)| \ge m = |p(z_0)|$, we get $(*)$ $|f(z)| \ge 1$ for all z.

Introduce the notation $e^{i\alpha} = \cos\alpha + i \sin\alpha$, which describes the points on the unit circle. Since the absolute value of $-|b_k|/b_k$ is 1, so $-|b_k|/b_k = e^{i\alpha}$ for some α. Let $\theta = \alpha/k$, then $e^{ik\theta} b_k = e^{i\alpha} b_k = -|b_k|$. Considering $w = re^{i\theta}$ with $r < |b_k|^{-1/k} (\Rightarrow 1 - r^k|b_k| > 0)$, we have $|1 + b_k w^k| = |1 + b_k r^k e^{ik\theta}| = \big|1 - r^k|b_k|\big| = 1 - r^k|b_k|$ and

$$\begin{aligned} |f(w)| = |1 + b_k w^k + \cdots + b_n w^n| &\le |1 + b_k w^k| + |b_{k+1} w^{k+1}| + \cdots + |b_n w^n| \\ &= 1 - r^k|b_k| + |b_{k+1}|r^{k+1} + \cdots + |b_n|r^n \\ &= 1 - r^k \underbrace{\big(|b_k| - |b_{k+1}|r - \cdots - |b_n|r^{n-k}\big)}_{\to |b_k| > \frac{1}{2}|b_k| > 0 \text{ as } r\to 0+}. \end{aligned}$$

Taking $w = re^{i\theta}$ with a very small positive r, we have $|f(w)| \le 1 - r^k(\frac{1}{2}|b_k|) < 1$, a contradiction to $(*)$. Therefore $m = |p(z_0)| = 0$ and z_0 is a root of $p(z)$.

Chapter 8. Differentiation

Definitions. Let S be an interval of positive length. A function $f\colon S \to \mathbb{R}$ is *differentiable* at $x_0 \in S$ iff $f'(x_0) = \lim\limits_{\substack{x\to x_0\\ x\in S}} \dfrac{f(x)-f(x_0)}{x-x_0}$ exists in $\mathbb{R}$. Also, f is *differentiable* iff f is differentiable at every element of S.

Theorem. *If $f : S \to \mathbb{R}$ is differentiable at x_0, then f is continuous at x_0.*
Proof. By the computation formulas for limits,

$$\lim_{x\to x_0} f(x) = \lim_{x\to x_0}\left[\left(\frac{f(x)-f(x_0)}{x-x_0}\right)(x-x_0)+f(x_0)\right] = f'(x_0)\cdot 0 + f(x_0) = f(x_0).$$

Theorem (Differentiation Formulas). *If $f, g : S \to \mathbb{R}$ are differentiable at x_0, then $f+g$, $f-g$, fg, f/g (when $g(x_0)\neq 0$) are differentiable at x_0. In fact, $(f\pm g)'(x_0) = f'(x_0)\pm g'(x_0)$, $(fg)'(x_0) = f'(x_0)g(x_0)+f(x_0)g'(x_0)$ and $(\frac{f}{g})'(x_0) = \dfrac{f'(x_0)g(x_0)-f(x_0)g'(x_0)}{g(x_0)^2}$.*
Proof. By the computation formulas for limits,

$$\lim_{x\to x_0}\frac{(f\pm g)(x)-(f\pm g)(x_0)}{x-x_0} = \lim_{x\to x_0}\left[\frac{f(x)-f(x_0)}{x-x_0}\pm\frac{g(x)-g(x_0)}{x-x_0}\right] = f'(x_0)\pm g'(x_0),$$

$$\lim_{x\to x_0}\frac{(fg)(x)-(fg)(x_0)}{x-x_0} = \lim_{x\to x_0}\left[\frac{f(x)-f(x_0)}{x-x_0}g(x)+f(x_0)\frac{g(x)-g(x_0)}{x-x_0}\right] = f'(x_0)g(x_0)+f(x_0)g'(x_0),$$

$$\begin{aligned}\lim_{x\to x_0}\frac{(\frac{f}{g})(x)-(\frac{f}{g})(x_0)}{x-x_0} &= \lim_{x\to x_0}\frac{1}{g(x)g(x_0)}\left[\frac{f(x)-f(x_0)}{x-x_0}g(x_0)-\frac{g(x)-g(x_0)}{x-x_0}f(x_0)\right]\\ &= \frac{f'(x_0)g(x_0)-f(x_0)g'(x_0)}{g(x_0)^2}.\end{aligned}$$

Theorem (Chain Rule). *If $f : S \to \mathbb{R}$ is differentiable at x_0, $f(S)\subseteq S'$ and $g : S' \to \mathbb{R}$ is differentiable at $f(x_0)$, then $g\circ f$ is differentiable at x_0 and $(g\circ f)'(x_0) = g'(f(x_0))f'(x_0)$.*
Proof. The function

$$h(y) = \begin{cases} \dfrac{g(y)-g(f(x_0))}{y-f(x_0)} & \text{if } y\neq f(x_0)\\ g'(f(x_0)) & \text{if } y = f(x_0)\end{cases}$$

is continuous at $f(x_0)$ because $\lim\limits_{y\to f(x_0)} h(y) = g'(f(x_0)) = h(f(x_0))$. So $g(y)-g(f(x_0)) = h(y)(y-f(x_0))$ holds for every y in the domain of g. Then

$$\lim_{x\to x_0}\frac{g\circ f(x)-g\circ f(x_0)}{x-x_0} = \lim_{x\to x_0} h(f(x))\frac{f(x)-f(x_0)}{x-x_0} = h(f(x_0))f'(x_0) = g'(f(x_0))f'(x_0).$$

Remarks. Note f differentiable at x_0 does not imply f' is continuous at x_0. In fact, the function $f(x) = x^2\sin\frac{1}{x}$ for $x\neq 0$ and $f(0) = \lim\limits_{x\to 0} x^2\sin\dfrac{1}{x} = 0$ is continuous and differentiable with $f'(x) = 2x\sin\frac{1}{x}-\cos\frac{1}{x}$ for $x\neq 0$ and $f'(0) = \lim\limits_{x\to 0}(x^2\sin\dfrac{1}{x}-0)/x = 0$. However, f' is not continuous at 0 because $\lim\limits_{x\to 0} f'(x) = \lim\limits_{x\to 0} 2x\sin\dfrac{1}{x}-\cos\dfrac{1}{x}$ does not exist, hence not equal to $f'(0)$. In particular, f is not twice differentiable. (Also, the function $g(x) = x^2\sin\dfrac{1}{x^2}$ for $x\neq 0$ and $g(0)=0$ is differentiable on $\mathbb{R}$, but $g'(x)$ is not continuous at 0 and $g'(x)$ is unbounded on every open interval containing 0.)

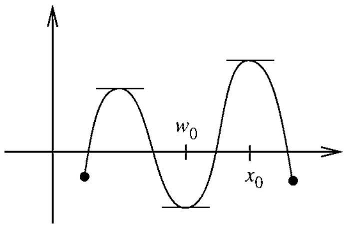
Notations. $C^0(S) = C(S)$ is the set of all continuous functions on S. For $n \in \mathbb{N}$, $C^n(S)$ is the set of all functions $f: S \to \mathbb{R}$ such that the n-th derivative $f^{(n)}$ is continuous on S. $C^\infty(S)$ is the set of all functions having n-th derivatives for all $n \in \mathbb{N}$. Functions in $C^1(S)$ are said to be *continuously differentiable* on S.

Inverse Function Theorem. *If f is continuous and injective on (a,b) and $f'(x_0) \neq 0$ for some $x_0 \in (a,b)$, then f^{-1} is differentiable at $y_0 = f(x_0)$ and $(f^{-1})'(y_0) = \dfrac{1}{f'(x_0)}$, i.e. $\dfrac{dx}{dy} = 1 \Big/ \dfrac{dy}{dx}$.*

Proof. The function $g(x) = \begin{cases} \dfrac{x - x_0}{f(x) - f(x_0)} & \text{if } x \neq x_0 \\ \dfrac{1}{f'(x_0)} & \text{if } x = x_0 \end{cases}$ is continuous at x_0 because $\lim_{x \to x_0} g(x) = \dfrac{1}{f'(x_0)} = g(x_0)$. Since f is continuous and injective on (a,b), f^{-1} is continuous by the continuous inverse theorem. So $\lim_{y \to y_0} f^{-1}(y) = f^{-1}(y_0) = x_0$ and $(f^{-1})'(y_0) = \lim_{y \to y_0} \dfrac{f^{-1}(y) - f^{-1}(y_0)}{y - y_0} = \lim_{y \to y_0} g\big(f^{-1}(y)\big) = g(x_0) = \dfrac{1}{f'(x_0)}$.

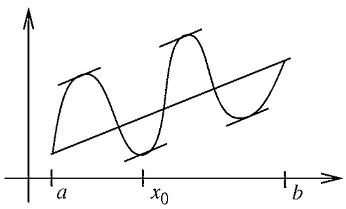
Local Extremum Theorem. *Let $f : (a,b) \to \mathbb{R}$ be differentiable. If $f(x_0) = \min_{x \in (a,b)} f(x)$ or $f(x_0) = \max_{x \in (a,b)} f(x)$, then $f'(x_0) = 0$.*

Proof. If $f(x_0) = \min_{x \in (a,b)} f(x)$, then $0 \leq \lim_{x \to x_0^+} \dfrac{f(x) - f(x_0)}{x - x_0} = f'(x_0) = \lim_{x \to x_0^-} \dfrac{f(x) - f(x_0)}{x - x_0} \leq 0$. So $f'(x_0) = 0$. The other case is similar.



Rolle's Theorem. *Let f be continuous on $[a,b]$ and differentiable on (a,b). If $f(a) = f(b)$, then there is (at least one) $z_0 \in (a,b)$ such that $f'(z_0) = 0$.*

Proof. This is trivial if f is constant on $[a,b]$. Otherwise, by the extreme value theorem, there are $x_0, w_0 \in [a,b]$ such that $f(x_0) = \max_{x \in [a,b]} f(x) > \min_{x \in [a,b]} f(x) = f(w_0)$. Either $f(x_0) \neq f(a)$ or $f(w_0) \neq f(a)$. So either x_0 or w_0 is in (a,b). By the last theorem, $f'(x_0) = 0$ or $f'(w_0) = 0$.



Mean-Value Theorem. *Let f be continuous on $[a,b]$ and differentiable on (a,b). Then there exists $x_0 \in (a,b)$ such that $f(b) - f(a) = f'(x_0)(b-a)$.*

Proof. Define $F(x) = f(x) - \big(\dfrac{f(b) - f(a)}{b - a}(x - a) + f(a)\big)$. Then $F(a) = 0 = F(b)$. By Rolle's theorem, there is $x_0 \in (a,b)$ such that $0 = F'(x_0) = f'(x_0) - \dfrac{f(b) - f(a)}{b - a}$, which is equivalent to $f(b) - f(a) = f'(x_0)(b - a)$.

Examples. (1) For $a < b$, show $|\sin b - \sin a| \leq |b - a|$. (By the mean-value theorem, there is $x_0 \in (a,b)$ such that $\sin b - \sin a = (\cos x_0)(b - a)$, so $|\sin b - \sin a| \leq |b - a|$.)

(2) Show $(1+x)^\alpha \geq 1 + \alpha x$ for $x \geq -1$ and $\alpha \geq 1$. (Let $f(x) = (1+x)^\alpha - 1 - \alpha x$, then there is $x_0 \in (0,x)$ if $x > 0$ or $x_0 \in (x,0)$ if $-1 < x < 0$ such that

$$(1+x)^\alpha - 1 - \alpha x = f(x) - f(0) = f'(x_0)(x - 0) = \alpha((1+x_0)^{\alpha-1} - 1)x \geq 0.)$$

(3) Show $\ln x \leq x - 1$ for $x > 0$. (Let $f(x) = \ln x - x + 1$, then $f(1) = 0$. If $x > 1$, then there is $x_0 \in (1,x)$ such that

$$\ln x - x + 1 = f(x) = f(x) - f(1) = f'(x_0)(x - 1) = \left(\frac{1}{x_0} - 1\right)(x - 1) \leq 0.$$

The case $0 < x < 1$ is similar.)

(4) To approximate $\sqrt{16.1}$, we can let $f(x) = \sqrt{x}$. Then $f(16.1) - f(16) = f'(c)(16.1 - 16)$ for some c between 16 and 16.1. Now $c \approx 16$ and $f(16.1) - f(16) \approx f'(16)(16.1 - 16) = 0.0125$, which gives $\sqrt{16.1} = f(16.1) \approx 4.0125$.

Curve Tracing Theorem. *If $f' \geq 0$ (respectively $f' > 0$, $f' \leq 0$, $f' < 0$, $f' \neq 0$, $f' \equiv 0$) everywhere on (a, b), then f is increasing (respectively strictly increasing, decreasing, strictly decreasing, injective, constant) on (a, b).*

Proof. If $x, y \in (a, b)$ and $x < y$, then by the mean value theorem, there is $x_0 \in (x, y)$ such that $f(y) - f(x) = f'(x_0)(y - x) \geq 0$ (respectively > 0, ≤ 0, < 0, $\neq 0$, $= 0$). Therefore, $f(x) \leq f(y)$ (respectively $f(x) < f(y)$, $f(x) \geq f(y)$, $f(x) > f(y)$, $f(x) \neq f(y)$, $f(x) = f(y)$).

Remarks. For differentiable function f, the converse of the strictly increasing (respectively strictly decreasing, injective) parts of the curve tracing theorem are false. To see an example, consider $f(x) = x^3$, which is strictly increasing on $\mathbb{R}$, but $f'(0) = 0$. The converse of the increasing (respectively decreasing, constant) parts are true because $(f(x) - f(x_0))/(x - x_0)$ is nonnegative (respectively nonpositive, zero) for $x, x_0 \in (a, b)$ and hence the same for the limit as x tends to x_0 on (a, b).

Local Tracing Theorem. *If $f : [a, b] \to \mathbb{R}$ is continuous and $f'(c) > 0$ for some $c \in [a, b]$, then there are $c_0, c_1 \in \mathbb{R}$ such that $c_0 < c < c_1$ and $f(x) < f(c) < f(y)$ for all $x \in (c_0, c) \cap [a, b]$ and all $y \in (c, c_1) \cap [a, b]$. A similar result for the case $f'(c) < 0$ holds and the inequality becomes $f(x) > f(c) > f(y)$.*

Proof. Let $f'(c) > 0$. Assume there is no such c_0. Then for every $n \in \mathbb{N}$, there is $x_n \in (c - \frac{1}{n}, c) \cap [a, b]$ such that $f(x_n) \geq f(c)$. This leads to $f'(c) = \lim\limits_{n\to\infty} \dfrac{f(x_n) - f(c)}{x_n - c} \leq 0$, a contradiction. The other parts of the theorem are similar.

Remarks. If $f'(c) \geq 0$, we may not have $f(x) \leq f(c) \leq f(y)$ similar to above as the function $f(x) = x^2 \sin \dfrac{1}{x}$ with $f(0) = 0$ satisfies $f'(0) = 0$, but f takes positive and negative values on every open interval about 0.

Exercise. A function $f : [a, b] \to \mathbb{R}$ is said to have the *intermediate value property* iff for every y_0 in the open interval with endpoints $f(a)$ and $f(b)$, there exists at least one $x_0 \in (a, b)$ such that $f(x_0) = y_0$. We have already showed that continuous functions on $[a, b]$ satisfied this property. Prove that if g is differentiable on $[a, b]$, then g' has the intermediate value property. In particular, if $g'(x) \neq 0$ for all $x \in [a, b]$, then $g' > 0$ or $g' < 0$ everywhere on $[a, b]$.

Next we will introduce the generalized mean-value theorem, which has two very important applications, namely Taylor's theorem and L'Hôpital's rule.

Generalized Mean-Value Theorem. *Let f, g be continuous on $[a, b]$ and differentiable on (a, b). Then there is $x_0 \in (a, b)$ such that $g'(x_0)(f(b) - f(a)) = f'(x_0)(g(b) - g(a))$. (Note the case $g(x) = x$ is the mean-value theorem.)*

Proof. Let $F(x) = f(x)(g(b) - g(a)) - (f(b) - f(a))(g(x) - g(a))$, then $F(a) = f(a)(g(b) - g(a)) = F(b)$. By Rolle's theorem, there is $x_0 \in (a, b)$ such that $0 = F'(x_0) = f'(x_0)(g(b) - g(a)) - g'(x_0)(f(b) - f(a))$.

Taylor's Theorem. *Let $f: (a, b) \to \mathbb{R}$ be n times differentiable on (a, b). For every $x, c \in (a, b)$, there is x_0 between x and c such that*

$$f(x) = f(c) + \frac{f'(c)}{1!}(x - c) + \frac{f''(c)}{2!}(x - c)^2 + \ldots + \frac{f^{(n-1)}(c)}{(n-1)!}(x - c)^{n-1} + \frac{f^{(n)}(x_0)}{n!}(x - c)^n.$$

(This is called the n-th Taylor expansion of f about c. The term $R_n(x) = \dfrac{f^{(n)}(x_0)}{n!}(x - c)^n$ is called the Lagrange form of the remainder.)

Proof. Let I be the closed interval with endpoints x and c. For $t \in I$, define $g(t) = (n - 1)!\sum_{k=0}^{n-1} \dfrac{f^{(k)}(t)}{k!}(x - t)^k$, where $f^{(0)} = f$, $(x - x)^0 = 1$. Let $p(t) = -\dfrac{(x - t)^n}{n}$. We get $g'(t) = f^{(n)}(t)(x - t)^{n-1}$ and $p'(t) = (x - t)^{n-1}$. By the generalized mean value theorem, there is x_0 between x and c such that

$$\underbrace{g'(x_0)}_{f^{(n)}(x_0)(x-x_0)^{n-1}} \underbrace{[p(x) - p(c)]}_{(x-c)^n/n} = \underbrace{p'(x_0)}_{(x-x_0)^{n-1}} [\underbrace{g(x)}_{(n-1)!f(x)} - g(c)].$$

Then $f(x) = \frac{g(c)}{(n-1)!} + \frac{f^{(n)}(x_0)}{n!}(x-c)^n = \sum_{k=0}^{n-1} \frac{f^{(k)}(c)}{k!}(x-c)^k + \frac{f^{(n)}(x_0)}{n!}(x-c)^n$.

Lemma. *Let $h : (a,b) \to [0,+\infty)$ be a bounded function, where a may be real or $-\infty$ and b may be real or $+\infty$. We have $\lim_{x\to a^+} h(x) = 0$ if and only if $\lim_{x\to a^+} \sup\{h(t) : a < t < x\} = 0$. Similarly, $\lim_{x\to b^-} h(x) = 0$ if and only if $\lim_{x\to b^-} \sup\{h(t) : x < t < b\} = 0$.*

Proof. For right hand limits at a, $\lim_{x\to a^+} h(x) = 0$ is the same as for every $\varepsilon > 0$, there exists $r \in \mathbb{R}$ such that for all $a < x < r$, $h(x) = |h(x) - 0| < \varepsilon$. The later part is the same as for all $a < x < r$, $\sup\{h(t) : a < t < x\} \le \varepsilon$, which is $\lim_{x\to a^+} \sup\{h(t) : a < t < x\} = 0$. The statement for the left hand limits at b is similar.

L'Hôpital's Rule ($\frac{0}{0}$ Version). *Let f, g be differentiable on (a,b) and $g(x), g'(x) \neq 0$ for all $x \in (a,b)$, where $-\infty \le a < b \le +\infty$. If*

(a) $\lim_{x\to a^+} \frac{f'(x)}{g'(x)} = L$, *where* $-\infty \le L \le +\infty$, *and*

(b) $\lim_{x\to a^+} f(x) = 0 = \lim_{x\to a^+} g(x)$,

then $\lim_{x\to a^+} \frac{f(x)}{g(x)} = L$. *(Similarly, the rule is also true if $x \to b^-$.)*

Proof. For $f'/g' \to L \in \mathbb{R}$, replacing f/g by $(f - Lg)/g = (f/g) - L$, we may assume $L = 0$.

Since $\lim_{x\to a^+} \frac{f'(x)}{g'(x)} = 0$, we may assume $\frac{f'(x)}{g'(x)}$ is bounded on $(a,c) \subseteq (a,b)$. For every $x, y \in (a,c)$ with $y < x$, by generalized mean-value theorem, there is $w \in (y,x)$ such that

$$\left|\frac{f(x)-f(y)}{g(x)-g(y)}\right| = \left|\frac{f'(w)}{g'(w)}\right| \le \sup\left\{\left|\frac{f'(t)}{g'(t)}\right| : a < t < x\right\}. \qquad (*)$$

Taking limit of $y \to a^+$ on both sides, we get $\left|\frac{f(x)}{g(x)}\right| \le \sup\left\{\left|\frac{f'(t)}{g'(t)}\right| : a < t < x\right\}$. By condition (a) and the lemma, the right side goes to 0. So by the sandwich theorem, we get the left side go to 0, which is the conclusion.

For $f'/g' \to L = +\infty$, there is $r \in \mathbb{R}$ such that $a < t < r$ implies $f'(t)/g'(t) > 1$ so that f' and g' are both positive or both negative when $x \in (a,r)$. Next for $a < y < x < r$, we have $\frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f'(t)}{g'(t)} > 1$. As $y \to a^+$, we see $\frac{f(x)}{g(x)} \ge 1$. So f and g are both positive or both negative when $x \in (a,r)$. Now $g'/f' \to 0$, $g \to 0$ and $f \to 0$ as $x \to a^+$. So by above, $g/f \to 0$. Hence $f/g \to +\infty$. The cases $L = -\infty$ and $x \to b^-$ are similar.

Example. (Even if $\lim_{x\to a} f(x) = 0 = \lim_{x\to a} g(x)$ and $\lim_{x\to a} \frac{f'(x)}{g'(x)}$ does not exist, $\lim_{x\to a} \frac{f(x)}{g(x)}$ may still exist as the following example will show.) Let $f(x) = x^2 \sin\frac{1}{x}$, $g(x) = \sin x$, $a = 0$, then $-x^2 \le x^2 \sin\frac{1}{x} \le x^2$ implies $\lim_{x\to 0} f(x) = 0$ by sandwich theorem, $\lim_{x\to 0} g(x) = 0$, $\lim_{x\to 0} \frac{f'(x)}{g'(x)} = \lim_{x\to 0} \frac{2x\sin\frac{1}{x} - \cos\frac{1}{x}}{\cos x} = \lim_{x\to 0} \frac{0 - \cos\frac{1}{x}}{1}$ does not exist. However, $\lim_{x\to 0} \frac{f(x)}{g(x)} = \lim_{x\to 0}\left(\frac{x}{\sin x}\right)(x\sin\frac{1}{x}) = 1 \cdot 0 = 0$.

L'Hôpital's Rule ($\frac{*}{\infty}$ Version). *Let f, g be differentiable on (a,b) and $g(x), g'(x) \neq 0$ for all $x \in (a,b)$, where $-\infty \le a < b \le +\infty$. If*

(a) $\lim_{x\to a^+} \frac{f'(x)}{g'(x)} = L$, *where* $-\infty \le L \le +\infty$, *and*

(b) $\lim_{x\to a^+} g(x) = +\infty$

then $\lim_{x\to a^+} \dfrac{f(x)}{g(x)} = L$. *(Similarly, the rule is also true if* $x \to b^-$ *or* $g(x) \to -\infty$*).*

Proof. We will modify the proof above as follow. If $L \in \mathbb{R}$, then we can reduce to $L = 0$ as above. In the case $f'/g' \to L = +\infty$, for $x, s \in (a, b)$, by generalized mean-value theorem, there is t between x and s such that $f(x) = f(s) + \dfrac{f'(t)}{g'(t)}(g(x) - g(s)) \to +\infty$ as $x \to a^+$ by (a) and (b). So we may replace f/g by g/f to assume $L = 0$. Multiplying the left and right sides of (*) by $\left|\dfrac{g(x) - g(y)}{g(y)}\right|$, we get for $a < y < x < c$,

$$\left|\frac{f(x) - f(y)}{g(y)}\right| \le \underbrace{\sup\left\{\left|\frac{f'(t)}{g'(t)}\right| : a < t < x\right\}}_{\to\ L=0 \text{ as } x\to a^+ \text{ by lemma}} \cdot \underbrace{\left|\frac{g(x) - g(y)}{g(y)}\right|}_{\to\ 1 \text{ as } y\to a^+}.$$

For every $\varepsilon > 0$, there is $x \in (a, b)$ such that $\sup\left\{\left|\dfrac{f'(t)}{g'(t)}\right| : a < t < x\right\} < \dfrac{\varepsilon}{4}$. Also, $\lim_{y\to a^+} \dfrac{f(x)}{g(y)} = 0$. Then there is $r < x$ such that for all $y \in (a, r)$, we can get $\left|\dfrac{f(x)}{g(y)}\right| < \dfrac{\varepsilon}{2}$ and $\left|\dfrac{g(x) - g(y)}{g(y)}\right| < 2$. Then

$$\left|\frac{f(y)}{g(y)}\right| \le \left|\frac{f(x)}{g(y)}\right| + \left|\frac{f(y) - f(x)}{g(y)}\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{4} \cdot 2 = \varepsilon.$$

The conclusion follows. The cases $L = -\infty$, $x \to b^-$ and $g(x) \to -\infty$ are similar.

Examples. (1) (Even if $\lim_{x\to a} f(x) = +\infty = \lim_{x\to a} g(x)$ and $\lim_{x\to a} \dfrac{f'(x)}{g'(x)}$ does not exist, $\lim_{x\to a} \dfrac{f(x)}{g(x)}$ may still exist as the following example will show.) Let $f(x) = 2x+\sin x, g(x) = 2x-\sin x, a = +\infty$, then $\lim_{x\to+\infty} f(x) = +\infty = \lim_{x\to+\infty} g(x)$, $\lim_{x\to+\infty} \dfrac{f'(x)}{g'(x)} = \lim_{x\to+\infty} \dfrac{2 + \cos x}{2 - \cos x}$ does not exist. However,

$$\lim_{x\to+\infty} \frac{f(x)}{g(x)} = \lim_{x\to+\infty} \frac{2 + (\sin x)/x}{2 - (\sin x)/x} = 1.$$

---

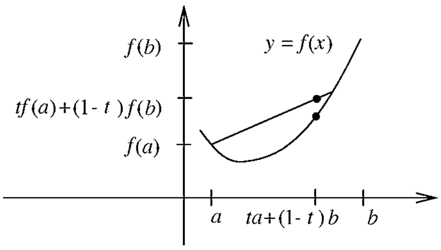
(2) Show that $\lim_{x\to+\infty} \dfrac{x^r}{e^x} = 0$ for every $r \in \mathbb{R}$.

Solution. There is an integer $n \ge |r|$ (by the Archimedian principle). We have $x^r \le x^n$ on $[1, +\infty)$. So $0 \le \dfrac{x^r}{e^x} \le \dfrac{x^n}{e^x}$ on $[1, \infty)$. Since $\dfrac{d^n}{dx^n}x^n = n!$ and $\lim_{x\to+\infty} \dfrac{n!}{e^x} = 0$, applying L'Hôpital's rule n times, we see $\lim_{x\to+\infty} \dfrac{x^n}{e^x} = 0$, which implies $\lim_{x\to+\infty} \dfrac{x^r}{e^x} = 0$ by the sandwich theorem.

(3) Let $f : (a, +\infty) \to \mathbb{R}$ be differentiable. Show that $\lim_{x\to+\infty} f'(x) + f(x) = 0$ if and only if $\lim_{x\to+\infty} f(x) = 0$ and $\lim_{x\to+\infty} f'(x) = 0$. (This type of result is often needed in the studies of differential equations. Here if $\lim_{x\to+\infty} g(x) = 0$, then every solution $y = f(x)$ of the equation $\dfrac{dy}{dx} + y = g(x)$ will have limit 0 as $x \to +\infty$.)

Solution. The if part follows by the computation formulas for limit. For the only if part, consider $f(x) = \dfrac{f(x)e^x}{e^x}$. The denominator on the right side tend to $+\infty$ as x tend to $+\infty$. Since $\lim_{x\to+\infty} \dfrac{\big(f(x)e^x\big)'}{(e^x)'} = \lim_{x\to+\infty} f'(x) + f(x) = 0$, by L'Hôpital's rule, $\lim_{x\to+\infty} f(x) = \lim_{x\to+\infty} \dfrac{f(x)e^x}{e^x} = 0$ and $\lim_{x\to+\infty} f'(x) = \lim_{x\to+\infty}\big(f'(x) + f(x)\big) - f(x) = 0$.

Appendix 1: Convex and Concave Functions.



convex function

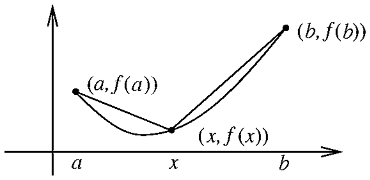
Definitions. Let I be an interval and $f: I \to \mathbb{R}$. We say f is a *convex function on* I iff for every $a, b \in I$, $0 \le t \le 1$,

$$f\big(ta + (1-t)b\big) \le tf(a) + (1-t)f(b).$$

We say f is a *concave function on* I if

$$f\big(ta + (1-t)b\big) \ge tf(a) + (1-t)f(b).$$

Remarks. As t ranges from 0 to 1, the point $\big(ta + (1-t)b, f(ta + (1-t)b)\big)$ traces the graph of $y = f(x)$ for $a \le x \le b$, while the point $\big(ta + (1-t)b, tf(a) + (1-t)f(b)\big)$ traces the chord joining $(a, f(a))$ to $(b, f(b))$. So f is convex on I if and only if the chord joining two points on the graph always lies above or on the graph.



Theorem. *f is convex on I if and only if the slope of the chords always increase, i.e. $a, x, b \in I$, $a < x < b \Rightarrow \dfrac{f(x) - f(a)}{x - a} \le \dfrac{f(b) - f(x)}{b - x}$.*

Proof. $x = ta + (1-t)b$ for some $t \in [0, 1] \iff 0 \le t = \dfrac{b - x}{b - a} \le 1$.

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(x)}{b - x} \iff f(x) \le \frac{b - x}{b - a} f(a) + \frac{x - a}{b - a} f(b) \iff f\big(ta+(1-t)b\big) \le tf(a)+(1-t)f(b).$$

Theorem. *For f differentiable on I, f is convex on $I \iff f'$ is increasing on I.* (For f twice differentiable on I, f is convex on $I \iff f'' \ge 0$ on I.)

Proof. ($\Rightarrow$) If $a, b \in I$, $a < b$, then $f'(a) = \lim\limits_{x\to a^+} \dfrac{f(x) - f(a)}{x - a} \le \lim\limits_{x\to a^+} \dfrac{f(b) - f(x)}{b - x} = \dfrac{f(b) - f(a)}{b - a} = \lim\limits_{x\to b^-} \dfrac{f(x) - f(a)}{x - a} \le \lim\limits_{x\to b^-} \dfrac{f(b) - f(x)}{b - x} = f'(b)$.

($\Leftarrow$) If $a, x, b \in I$, $a < x < b$, then by the mean value theorem, there are r, s such that $a < r < x < s < b$ and $\dfrac{f(x) - f(a)}{x - a} = f'(r) \le f'(s) = \dfrac{f(b) - f(x)}{b - x}$.

Theorem. *If f is convex on (a, b), then f is continuous on (a, b).*

Proof. For $x_0 \in (a, b)$, consider $u, v, w \in (a, b)$ such that $u < x_0 < v < w$. Then $\dfrac{f(x_0) - f(u)}{x_0 - u} \le \dfrac{f(v) - f(x_0)}{v - x_0} \le \dfrac{f(w) - f(v)}{w - v}$. Solving for $f(v)$, we get $\dfrac{f(x_0) - f(u)}{x_0 - u}(v - x_0) + f(x_0) \le f(v) \le \dfrac{f(w) - f(v)}{w - v}(v - x_0) + f(x_0)$. Taking limit as $v \to x_0^+$, we have $f(x_0) \le f(x_0+) \le f(x_0)$, i.e. $f(x_0+) = f(x_0)$. Similarly, we can show $f(x_0-) = f(x_0)$ by taking $u < v < x_0 < w$. Therefore, $\lim\limits_{x\to x_0} f(x) = f(x_0)$ for every $x_0 \in (a, b)$.

Example. The function $f(x) = \begin{cases} 0 & \text{if } 0 \le x < 1 \\ 1 & \text{if } x = 1 \end{cases}$ is convex on $[0, 1]$ by checking the slopes of the chords over $[0, 1]$, but f is not continuous at 1 because $\lim\limits_{x\to 1^-} f(x) = 0 \ne 1 = f(1)$.

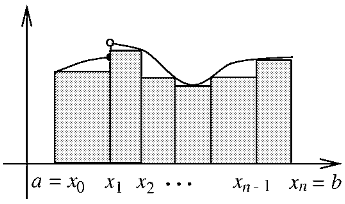
Remark. The proof of the last theorem uses the fact that on an open interval, at any point x_0, there are points on its left and points on its right, which is not true for endpoints of a closed interval.

Example. Prove that if $a, b \ge 0$ and $0 < r < 1$, then $|a^r - b^r| \le |a - b|^r$. (In particular, for $n = 2, 3, 4, \ldots$, we have $|\sqrt[n]{a} - \sqrt[n]{b}| \le \sqrt[n]{|a - b|}$.)

Solution. We may assume $a > b$, otherwise interchange a and b. Define $f : [0, a] \to \mathbb{R}$ by $f(x) = x^r + (a-x)^r$. Since $r - 1 < 0$, so $f''(x) = r(r-1)\big(x^{r-2} + (a-x)^{r-2}\big) \le 0$. So f is concave on $[0, a]$. Since $f(0) = a^r = f(a)$, we get $f(x) = x^r + (a-x)^r \ge a^r$ for all $x \in [0, a]$. Therefore, if $b \in [0, a]$, we have $|a-b|^r = (a-b)^r \ge a^r - b^r = |a^r - b^r|$.

Chapter 9. Riemann Integral

For $a, b \in \mathbb{R}$, let f be a bounded function on $[a, b]$, say $|f(x)| \leq K$ for every $x \in [a, b]$. Let $P = \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$, i.e. $a = x_0 < x_1 < \ldots < x_{n-1} < x_n = b$. The length of $[x_{j-1}, x_j]$ is $\Delta x_j = x_j - x_{j-1}$ and the *mesh* of P is $\|P\| = \max\{\Delta x_1, \ldots, \Delta x_n\}$. Also, on $[x_{j-1}, x_j]$, let $m_j = \inf\{f(x) : x \in [x_{j-1}, x_j]\}$ and $M_j = \sup\{f(x) : x \in [x_{j-1}, x_j]\}$.



Definitions. For the partition P, a *Riemann sum* is a sum of the form $S = \sum_{j=1}^{n} f(t_j)\Delta x_j$, where every t_j is in $[x_{j-1}, x_j]$. The *lower Riemann sum* is $L(f, P) = \sum_{j=1}^{n} m_j \Delta x_j$ and the *upper Riemann sum* is $U(f, P) = \sum_{j=1}^{n} M_j \Delta x_j$. (Note $-K \leq m_j \leq f(t_j) \leq M_j \leq K$ implies $-K(b-a) \leq L(f, P) \leq S \leq U(f, P) \leq K(b-a)$.)

In 1854, George B. Riemann defined the integral of $f(x)$ on $[a, b]$ to be $\lim_{\|P\|\to 0} \sum_{j=1}^{n} f(t_j)\Delta x_j$. There are two immediate technical problems with such a definition.

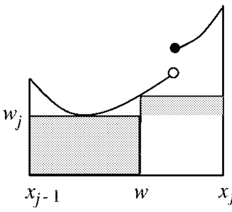
(1) As the choices of the t_j's may vary, it is hard to say how close the Riemann sum is to the actual integral. In particular, it is not clear if the Riemann sum is greater than or less than the integral at any time.

(2) The type of limit involved is not the limit of a sequence nor the limit of a function. In fact, there are many variables in a Riemann sum!

Instead of dealing with these technicalities, we introduce integral following the approach of J. Gaston Darboux in 1875.

Definition. We say P' is a *refinement* of P or P' is *finer than* P iff P and P' are partitions of $[a, b]$ and $P \subseteq P'$.

Refinement Theorem. *If P' is a finer partition of $[a, b]$ than P, then $L(f, P) \leq L(f, P') \leq U(f, P') \leq U(f, P)$.*



Proof. P' can be obtained from P by adding one point at a time. It suffices to consider the case $P' = P \cup \{w\}$ with $w \notin P = \{x_0, x_1, \ldots, x_n\}$. Suppose $x_{j-1} < w < x_j$.

Let $m_j = \inf\{f(x) : x \in [x_{j-1}, x_j]\}$, $m'_j = \inf\{f(x) : x \in [x_{j-1}, w]\}$ and $m''_j = \inf\{f(x) : x \in [w, x_j]\}$. Since $[x_{j-1}, w], [w, x_j] \subseteq [x_{j-1}, x_j]$, we have $m_j \leq m'_j$ and $m_j \leq m''_j$. Then $m_j\Delta x_j = m_j(w - x_{j-1}) + m_j(x_j - w) \leq m'_j(w - x_{j-1}) + m''_j(x_j - w)$. So $L(f, P) \leq L(f, P')$. Similarly, $U(f, P') \leq U(f, P)$.

Observe that $L(f, P)$ "underestimates" the area under the curve; $U(f, P)$ "overestimates" the area under the curve.

Definitions. The *lower integral* of f on $[a, b]$ is

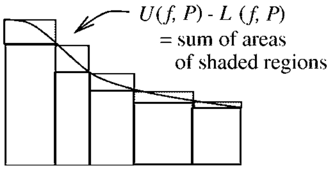
$$(L)\int_a^b f(x)\,dx = \sup\{L(f, P) \text{: } P \text{ is a partition of } [a, b]\} \quad \text{(the largest underestimate)}$$

and the *upper integral* of f on $[a, b]$ is

$$(U)\int_a^b f(x)\,dx = \inf\{U(f, P) \text{: } P \text{ is a partition of } [a, b]\} \quad \text{(the smallest overestimate).}$$

(So for every partition P of $[a, b]$, $L(f, P) \leq (L)\int_a^b f(x)\,dx \leq (U)\int_a^b f(x)\,dx \leq U(f, P)$.)

We say f is (*Riemann*) *integrable* on $[a, b]$ iff $(L)\int_a^b f(x)\,dx = (U)\int_a^b f(x)\,dx$. In that case, we denote the common value by $\int_a^b f(x)\,dx$. (For $b \leq a$, define $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$. In particular, $\int_a^a f(x)\,dx = 0$.)



Theorem (Integral Criterion). *Let $f\colon [a,b] \to \mathbb{R}$ be a bounded function. The function f is (Riemann) integrable on $[a,b]$ if and only if for every $\varepsilon > 0$, there is a partition P of $[a,b]$ such that $U(f,P) - L(f,P) < \varepsilon$.*

Proof. If for every $\varepsilon > 0$, there is a partition P of $[a,b]$ such that $U(f,P) - L(f,P) < \varepsilon$, then $0 \le (U)\int_a^b f(x)\,dx - (L)\int_a^b f(x)\,dx \le U(f,P) - L(f,P) < \varepsilon$. By the infinitesimal principle, we get $(L)\int_a^b f(x)\,dx = (U)\int_a^b f(x)\,dx$, i.e. f is (Riemann) integrable on $[a,b]$.

Conversely, if f is (Riemann) integrable on $[a,b]$, then for every $\varepsilon > 0$, by the supremum property, we have $(L)\int_a^b f(x)\,dx - \varepsilon/2 < L(f,P_1) \le (L)\int_a^b f(x)\,dx$ for some partition P_1 of $[a,b]$. Similarly, $(U)\int_a^b f(x)\,dx \le U(f,P_2) < (U)\int_a^b f(x)\,dx + \varepsilon/2$ for some partition P_2 of $[a,b]$. Let $P = P_1 \cup P_2$, then by the refinement theorem, $L(f,P_1) \le L(f,P) \le U(f,P) \le U(f,P_2)$. Since $U(f,P_2) - L(f,P_1) < [(U)\int_a^b f(x)\,dx + \varepsilon/2] - [(L)\int_a^b f(x)\,dx - \varepsilon/2] = \varepsilon$, so $U(f,P) - L(f,P) \le U(f,P_2) - L(f,P_1) < \varepsilon$.

Question. Are there integrable functions? Are there non-integrable functions?

Below we will show that constant functions and continuous functions are integrable.

Example. Recall the function $f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$ is not continuous anywhere. Partitioning an interval $[a,b]$ into subintervals $[x_{j-1}, x_j]$, we get by the density of rational numbers and irrational numbers that $m_j = 0$ and $M_j = 1$. Then $L(f,P) = 0$ and $U(f,P) = b-a$ for every partition P of $[a,b]$. So $(L)\displaystyle\int_a^b f(x)\,dx = 0$, $(U)\displaystyle\int_a^b f(x)\,dx = b-a$. Therefore, $f(x)$ is not integrable on every interval $[a,b]$ with $a < b$.

Example. If $f(x) = c$ for every $x \in [a,b]$, then $L(f,P) = c(b-a) = U(f,P)$ for every partition P of $[a,b]$. So $(L)\displaystyle\int_a^b f(x)\,dx = c(b-a) = (U)\displaystyle\int_a^b f(x)\,dx$. Therefore, f is integrable on $[a,b]$ and $\displaystyle\int_a^b f(x)\,dx = c(b-a)$.

Uniform Continuity Theorem. *If $f : [a,b] \to \mathbb{R}$ is continuous, then f is uniformly continuous (in the sense that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x, t \in [a,b]$, $|x - t| < \delta$ implies $|f(x) - f(t)| < \varepsilon$.)*

Proof. Suppose f is not uniform continuous. Then there is an $\varepsilon > 0$ such that for every $\delta = \frac{1}{n}, n \in \mathbb{N}$, there are $x_n, t_n \in [a,b]$ such that $|x_n - t_n| < \delta = \frac{1}{n}$ and $|f(x_n) - f(t_n)| \ge \varepsilon$. By the Bolzano-Weierstrass theorem, $\{x_n\}$ has a subsequence $\{x_{n_j}\}$ converging to some $w \in [a,b]$. Since $|t_{n_j} - w| \le |t_{n_j} - x_{n_j}| + |x_{n_j} - w| \le \dfrac{1}{n_j} + |x_{n_j} - w|$, by the sandwich theorem, $\{t_{n_j}\}$ converges to w. Then $0 = |f(w) - f(w)| = \lim\limits_{j\to\infty} |f(x_{n_j}) - f(t_{n_j})| \ge \varepsilon > 0$, which is a contradiction.

Examples. For $\alpha \in (0,1]$, a function $f : S \to \mathbb{R}$ is *Hölder continuous of order α* iff f staisfies the condition that there exists constant $M > 0$ such that for all $x, y \in S$, $|f(x) - f(y)| \le M|x - y|^\alpha$. In case $\alpha = 1$, f is said to be *Lipschitz*. The constant M is called the *Lipschitz constant or Hölder constant* for f. All functions with bounded derivatives (such as $\cos x$ with $M = 1$) are Lipschitz by the mean-value theorem. The function $f(x) = \sqrt{x}$ satisfies $|\sqrt{x} - \sqrt{y}| \le \sqrt{|x - y|}$ on $[0, +\infty)$ is Hölder continuous of order $\frac{1}{2}$. All Hölder functions are uniformly continuous, since for every $\varepsilon > 0$, we can take $\delta = (\varepsilon/m)^{1/\alpha}$.

Theorem. *If f is continuous on $[a,b]$, then f is integrable on $[a,b]$.*

Proof. For $\varepsilon > 0$, since f is uniformly continuous on $[a,b]$, there is $\delta > 0$ such that $x, w \in [a,b]$, $|x - w| < \delta \Rightarrow |f(x) - f(w)| < \varepsilon/(b-a)$. Let $P = \{a = x_0, x_1, \ldots, x_n = b\}$ be a partition of $[a,b]$ with $\max\limits_{1\le j\le n} |x_j - x_{j-1}| < \delta$. By the extreme value theorem, $M_j = f(w_j)$ and $m_j = f(u_j)$ for some $w_j, u_j \in [x_{j-1}, x_j]$. Then

$$U(f,P) - L(f,P) = \sum_{j=1}^{n} \big(f(w_j) - f(u_j)\big)\Delta x_j < \frac{\varepsilon}{b-a} \sum_{j=1}^{n} (x_j - x_{j-1}) = \frac{\varepsilon(x_n - x_0)}{b-a} = \varepsilon.$$

By the integral criterion, f is integrable on $[a,b]$.

Exercises. (1) Let $f : [a,b] \to \mathbb{R}$ be a function and $c \in [a,b]$. Show that f is integrable on $[a,b]$ if and only if f is integrable on $[a,c]$ and $[c,b]$.

(2) If $f : [a,b] \to \mathbb{R}$ is bounded and discontinuous only at $x_1, \ldots, x_n \in [a,b]$, show that f is integrable on $[a,b]$.

(*Hint:* This can be done directly or by using (1) to reduce the problem to intervals having only one discontinuity.)

Questions. How bad can an integrable function be discontinuous? Which functions are integrable?

Definitions.(i) A set $S \subseteq \mathbb{R}$ is *of measure* 0 (or has *zero-length*) iff for every $\varepsilon > 0$, there are intervals (a_1, b_1), (a_2, b_2), $(a_3, b_3), \ldots$ such that $S \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n)$ and $\sum_{n=1}^{\infty} |a_n - b_n| < \varepsilon$.

(ii) We say a property holds *almost everywhere* iff the property holds except on a set of measure 0. (It is common to abbreviate almost everywhere by *a.e.* in advanced courses. In probability, *almost surely* is used instead of almost everywhere.)

Remarks. For example, it is known that monotone functions on intervals are differentiable almost everywhere (see H.L. Royden's book, *Real Analysis*, 3rd ed., p. 100). Also, H. Rademacher proved that Lipschitz functions on intervals are differentiable almost everwhere (see L.C. Evans' book, *Partial Differential Equations*, p. 281).

Lebesgue's Theorem(1902). *For a bounded function* $f : [a,b] \to \mathbb{R}$, f *is integrable on* $[a,b]$ *if and only if the set* $S_f = \{x \in [a,b] : f$ is discontinuous at $x\}$ *is of measure* 0 *(i.e.* f *is continuous almost everywhere).*

For a proof of Lebesgue's theorem, please go to appendix 1 of this chapter.

Examples. (1) The empty set $\emptyset$ is of measure 0 because $\emptyset \subseteq \bigcup_{n=1}^{\infty} (0,0)$ and $\sum_{n=1}^{\infty} |0-0| = 0 < \varepsilon$. So every continuous function $f : [a,b] \to \mathbb{R}$ is integrable on $[a,b]$, since the set of discontinuities S is $\emptyset$.

(2) A countable set $\{x_1, x_2, \ldots\}$ is of measure 0 because

$$\{x_1, x_2, x_3, \ldots\} \subseteq (x_1 - \tfrac{\varepsilon}{4}, x_1 + \tfrac{\varepsilon}{4}) \cup (x_2 - \tfrac{\varepsilon}{4^2}, x_2 + \tfrac{\varepsilon}{4^2}) \cup (x_3 - \tfrac{\varepsilon}{4^3}, x_3 + \tfrac{\varepsilon}{4^3}) \cup \cdots \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{2\varepsilon}{4^n} = \frac{2\varepsilon}{3} < \varepsilon.$$

So every monotone function on $[a,b]$ is integrable on $[a,b]$, since the set of discontinuities S is countable by the monotone function theorem, hence of measure 0.

(3) There exist uncountable sets, which are of measure 0 (*eg.* the Cantor set).

(4) A countable union of sets of measure 0 is of measure 0. (If $S_1, S_2, S_3, \ldots$ are of measure 0 and S is their union, then for every $\varepsilon > 0$, since S_k is of measure 0, there are intervals $(a_{k,n}, b_{k,n})$ such that $S_k \subseteq \bigcup_{n=1}^{\infty} (a_{k,n}, b_{k,n})$ and $\sum_{n=1}^{\infty} |a_{k,n} - b_{k,n}| < \frac{\varepsilon}{4^k}$. Then $S \subseteq \bigcup_{k=1}^{\infty} \bigcup_{n=1}^{\infty} (a_{k,n}, b_{k,n})$ and $\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} |a_{k,n} - b_{k,n}| < \sum_{k=1}^{\infty} \frac{\varepsilon}{4^k} = \frac{\varepsilon}{3} < \varepsilon$.)

(5) If a set S is of measure 0 and $S' \subseteq S$, then S' is also of measure 0. (For S', use the same intervals (a_n, b_n) for S.)

(6) Arrange $\mathbb{Q} \cap [0,1]$ in a sequence $r_1, r_2, r_3, \ldots$ and define $f_n(x) = \begin{cases} 1 & \text{if } x = r_1 \text{ or } r_2 \text{ or } \ldots \text{ or } r_n \\ 0 & \text{otherwise} \end{cases}$. Then $f_n(x)$ is integrable on $[0,1]$, since the set of discontinuities $S_{f_n} = \{r_1, r_2, \ldots, r_n\}$ is finite, hence of measure 0. Now

$\lim_{n\to\infty} f_n(x) = f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$ is not integrable on $[0, 1]$ since $S_f = \{x \in [0, 1] : f \text{ is discontinuous at } x\} = [0, 1]$, which is not of measure 0. So *the limit of a sequence of Riemann integrable functions may not be Riemann integrable.*

Remarks. In B. R. Gelbaum and J. M. H. Olmsted's book *Counterexamples in Analysis*, p. 106, there is an example of the limit of a sequence of continuous functions on $[0, 1]$, which is not Riemann integrable.

Theorem. *For $c \in [a, b]$, f is integrable on $[a, b]$ if and only if f is integrable on $[a, c]$ and $[c, b]$.*

Proof. Let S, S_1, S_2 be the set of discontinuous points of f on $[a, b], [a, c], [c, b]$, respectively. If f is integrable on $[a, b]$, then by Lebesgue's theorem, S is of measure 0. Since $S_1, S_2 \subseteq S$, so both S_1, S_2 are of measure 0. By Lebesgue's theorem, f is integrable on $[a, c]$ and $[c, b]$.

Conversely, if f is integrable on $[a, c]$ and $[c, b]$, then S_1, S_2 are of measure 0. Since $S \subseteq S_1 \cup S_2 \cup \{c\}$, by examples (4) and (5), S is of measure 0. By Lebesgue's theorem, f is integrable on $[a, b]$.

Theorem. *If $f, g : [a, b] \to \mathbb{R}$ are integrable on $[a, b]$, then $f + g$, $f - g$ and fg are integrable on $[a, b]$.*

Proof. If f and g are integrable on $[a, b]$, then f and g are bounded on $[a, b]$. So $f + g$, $f - g$ and fg are bounded on $[a, b]$.

Observe that if f and g are continuous at x, then $f + g$ is continuous at x. Taking contrapositive, we see that if $f + g$ is discontinuous at x, then f or g is discontinuous at x. So if $x \in S_{f+g}$, then $x \in S_f$ or $x \in S_g$, which implies $S_{f+g} \subseteq S_f \cup S_g$. Since f, g are integrable on $[a, b]$, by Lebesgue's theorem, S_f, S_g are of measure 0. By example (4), $S_f \cup S_g$ is of measure 0. By example (5), S_{f+g} is also of measure 0. Therefore, by Lebesgue's theorem $f + g$ is integrable on $[a, b]$. Using a similar argument, we can see that $f - g$, fg are integrable on $[a, b]$.

Remarks. By a similar argument, we can show that if $f : [a, b] \to \mathbb{R}$ is integrable on $[a, b]$ and g is bounded and continuous on $f([a, b])$, then $S_{g\circ f} \subseteq S_f$. (This is because if f is continuous at x, then $g \circ f$ is continuous at x. Taking contrapositive, we get if $x \in S_{g\circ f}$, then $x \in S_f$. So $S_{g\circ f} \subseteq S_f$.) Since S_f is of measure 0, by example (5), $S_{g\circ f}$ is of measure 0. So $g \circ f : [a, b] \to \mathbb{R}$ is integrable on $[a, b]$. In particular, if f is integrable on $[a, b]$, then taking $g(x) = |x|, x^2, e^x, \cos x, \ldots$, respectively, we get $|f|, f^2, e^f, \cos f, \ldots$ are integrable on $[a, b]$.

However, even if $f : [a, b] \to [c, d]$ is integrable on $[a, b]$ and $g : [c, d] \to \mathbb{R}$ is integrable on $[c, d]$, $g \circ f$ may not be integrable on $[a, b]$. (For example, define $f : [0, 1] \to [0, 1]$ by $f(0) = 1$, $f(m/n) = 1/n$, where m, n are positive integers with no common prime factors, and $f(x) = 0$ for every $x \in [0, 1] \setminus \mathbb{Q}$. Next, define $g : [0, 1] \to [0, 1]$ by $g(0) = 0$ and $g(x) = 1$ for every $x \in (0, 1]$. As an exercise, it can be showed that $S_f = [0, 1] \cap \mathbb{Q}$ and $S_g = \{0\}$. So f and g are integrable on $[0, 1]$. However, $g \circ f$ is the nonintegrable function that is 1 on $[0, 1] \cap \mathbb{Q}$ and 0 on $[0, 1] \setminus \mathbb{Q}$.)

There is even an example of a continuous function $f : [0, 1] \to [0, 1]$ and an integrable function $g : [0, 1] \to \mathbb{R}$ such that $g \circ f$ is not integrable (see B. R. Gelbaum and J. M. H. Olmstad's book *Counterexample in Analysis*, pp. 106-107).

Up to now, we have been trying to determine which functions are integrable. Below we will look at how the integrals of functions can be computed.

Theorem (Simple Properties of Riemann Integrals). Let f and g be integrable on $[a, b]$.

(1) $\int_a^b \big(f(x) \pm g(x)\big)\, dx = \int_a^b f(x)dx \pm \int_a^b g(x)dx$, $\int_a^b cf(x)\, dx = c \int_a^b f(x)\, dx$ for every $c \in \mathbb{R}$.

(2) If $f(x) \le g(x)$ for all $x \in [a, b]$, then $\int_a^b f(x)\, dx \le \int_a^b g(x)\, dx$. Also, $\left|\int_a^b f(x)\, dx\right| \le \int_a^b |f(x)|\, dx$.

(3) $\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx$ for $c \in [a, b]$.

Proof. To get (1), by the supremum and infimum properties, for every $\varepsilon > 0$, there is a partition P of $[a, b]$ such that

$$\int_a^b f(x)\, dx + \int_a^b g(x)\, dx - \varepsilon < L(f, P) + L(g, P) \le L(f + g, P) \le \int_a^b \big(f(x) + g(x)\big)\, dx$$

$$\leq U(f+g,P) \leq U(f,P)+U(g,P) < \int_a^b f(x)\,dx + \int_a^b g(x)\,dx + \varepsilon.$$

Letting $\varepsilon \to 0$, we get $\int_a^b \big(f(x)+g(x)\big)\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$. Next, from $U(-f,P) = -L(f,P)$ and $\inf(-S) = -\sup S$, we get $\int_a^b -f(x)\,dx = -\int_a^b f(x)\,dx$. Then

$$\int_a^b \big(f(x)-g(x)\big)\,dx = \int_a^b \big(f(x)+(-g(x))\big)dx = \int_a^b f(x)\,dx + \int_a^b -g(x)\,dx = \int_a^b f(x)\,dx - \int_a^b g(x)\,dx.$$

For the second statement, the case $c = 0$ is clear since $\int_a^b 0f(x)\,dx = 0 = 0\int_a^b f(x)\,dx$. If $c > 0$, then

$$\begin{aligned}\int_a^b cf(x)dx &= \sup\{cL(f,P)\colon P \text{ is a partition of } [a,b]\}\\ &= c\sup\{L(f,P)\colon P \text{ is a partition of } [a,b]\} = c\int_a^b f(x)dx.\end{aligned}$$

If $c < 0$, then

$$\int_a^b cf(x)\,dx = \int_a^b -(-c)f(x)\,dx = -\int_a^b (-c)f(x)\,dx = -(-c)\int_a^b f(x)\,dx = c\int_a^b f(x)\,dx.$$

For (2), observe that $g - f \geq 0$ on $[a,b]$ implies $L(g-f,P) \geq 0$ for every partition P. So

$$\int_a^b \big(g(x)-f(x)\big)\,dx = \sup\{L(g-f,P)\colon P \text{ is a partition of } [a,b]\} \geq 0 \quad \text{implies} \quad \int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx.$$

For the second statement, since $-|f| \leq f \leq |f|$ on $[a,b]$, we get $-\int_a^b |f(x)|\,dx \leq \int_a^b f(x)\,dx \leq \int_a^b |f(x)|\,dx$.

For (3), let P be a partition of $[a,b]$, $P' = P \cup \{c\}$, $P_1 = P' \cap [a,c]$ and $P_2 = P' \cap [c,b]$. Then P' is a finer partition of $[a,b]$ than P, P_1 is a partition of $[a,c]$, P_2 is a partition of $[c,b]$ and $L(f,P') = L(f,P_1) + L(f,P_2)$. Let

$$A = \{L(f,P)\colon P \text{ is a partition of } [a,b]\} \quad \text{and} \quad B = \{L(f,P')\colon P \text{ is a partition of } [a,b] \text{ and } P' = P \cup \{c\}\}.$$

Since P' is also a partition of $[a,b]$, we see $B \subseteq A$ and so $\sup B \leq \sup A$. By the refinement theorem, $L(f,P) \leq L(f,P') \leq \sup B$. Hence $\sup A \leq \sup B$. Therefore, $\sup A = \sup B$ and

$$\begin{aligned}\int_a^b f(x)\,dx &= \sup A = \sup B\\ &= \sup\{L(f,P_1)+L(f,P_2)\colon P_1, P_2 \text{ are partitions of } [a,c], [c,b], \text{ respectively}\}\\ &= \sup\{L(f,P_1)\colon P_1 \text{ is a partition of } [a,c]\} + \sup\{L(f,P_2)\colon P_2 \text{ is a partition of } [c,b]\}\\ &= \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.\end{aligned}$$

The most important tool for computing an integral is to find an *antiderivative* or *primitive function* of the integrable function, which is a function whose derivative is the integrable function. What can we say about such a function?

Example. For $x \in [-1,1]$, define $f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$. Since $f(x)$ has only one point of discontinuity at 0, $f(x)$ is integrable on $[-1,1]$. Now the function

$$F(x) = \int_0^x f(t)\,dt = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0 \end{cases} = |x|$$

is continuous on $[-1, 1]$, hence uniformly continuous there by the uniform continuity theorem, but $F(x)$ is not differentiable at 0, the same point where $f(x)$ is discontinuous!

Theorem. *If f is integrable on $[a, b]$ and $c \in [a, b]$, then $F(x) = \int_c^x f(t)\,dt$ is uniformly continuous on $[a, b]$.*

Proof. Recall there is $K > 0$ such that $|f(t)| \le K$ for every $t \in [a, b]$. For every $\varepsilon > 0$, let $\delta = \frac{\varepsilon}{K}$. Then $|x - x_0| < \delta$ implies $|F(x) - F(x_0)| = \left|\int_{x_0}^x f(t)\,dt\right| \le K|x - x_0| < \varepsilon$.

The next theorem is the most important theorem in calculus. It not only tells us how to compute an integral, but also the deep fact that differentiation and integration are inverse operations on functions. *Roughly,* it may be summarized by the formulas $\frac{d}{dx}\int_c^x h(t)\,dt = h(x)$ and $\int_c^x \frac{d}{dt}h(t)\,dt = h(t)\Big|_c^x$.

Fundamental Theorem of Calculus. *Let $c, x_0 \in [a, b]$.*

(1) *If f is integrable on $[a, b]$, continuous at $x_0 \in [a, b]$ and $F(x) = \int_c^x f(t)\,dt$, then $F'(x_0) = f(x_0)$.*

(2) *If G is differentiable on $[a, b]$ with $G' = g$ integrable on $[a, b]$, then $\int_a^b g(x)\,dx = G(b) - G(a)$.*

(Note G' need not be continuous by an example in the chapter on differentiation.)

Proof. (1) We will check $\lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} = f(x_0)$ using the definition of limit. For every $\varepsilon > 0$, since f is continuous at x_0, there exists a $\delta > 0$ such that for every $x \in [a, b]$, $|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon/2$. With the same δ, we get $0 < |x - x_0| < \delta$ implies

$$\left|\frac{F(x) - F(x_0)}{x - x_0} - f(x_0)\right| = \left|\frac{\int_c^x f(t)dt - \int_c^{x_0} f(t)dt}{x - x_0} - \frac{\int_{x_0}^x f(x_0)dt}{x - x_0}\right| = \left|\frac{\int_{x_0}^x (f(t) - f(x_0))dt}{x - x_0}\right| \le \left|\frac{\int_{x_0}^x \varepsilon/2\,dt}{x - x_0}\right| < \varepsilon.$$

(2) The conclusion will follow from the infinitesimal principle if we can show $\left|\int_a^b g(x)\,dx - \big(G(b) - G(a)\big)\right| < \varepsilon$ for every $\varepsilon > 0$. Now for every $\varepsilon > 0$, since g is integrable on $[a, b]$, by the integral criterion, there is a partition $P = \{a = x_0, x_1, \ldots, x_n = b\}$ such that $U(g, P) - L(g, P) < \varepsilon$. For $j = 1, 2, \ldots, n$, by the mean value theorem, there exists $t_j \in (x_{j-1}, x_j)$ such that $G(x_j) - G(x_{j-1}) = G'(t_j)(x_j - x_{j-1}) = g(t_j)\Delta x_j$. Then

$$L(g, P) \le \sum_{j=1}^n g(t_i)\Delta x_j = \sum_{j=1}^n (G(x_j) - G(x_{j-1})) = G(b) - G(a) \le U(g, P).$$

Also, $L(g, P) \le \int_a^b g(x)\,dx \le U(g, P)$. Then $\left|\int_a^b g(x)\,dx - \big(G(b) - G(a)\big)\right| \le U(g, P) - L(g, P) < \varepsilon$.

Remarks. In (2) above, the condition G' is integrable is important. The function $G(x) = x^2 \sin \frac{1}{x^2}$ for $x \ne 0$ and $G(0) = 0$ is differentiable on $\mathbb{R}$, but $G'(x)$ is unbounded on $[0, 1]$, hence not integrable there. In Y. Katznelson and K. Stromberg's paper *Everywhere Differentiable, Nowhere Monotone Functions*, *American Mathematical Monthly*, vol. 81, pp. 349-354, there is even an example of a differentiable function on $\mathbb{R}$ such that its derivative is bounded, but not Riemann integrable on any interval $[a, b]$ with $a < b$.

Theorem (Integration by Parts). *If f, g are differentiable on $[a, b]$ with f', g' integrable on $[a, b]$, then*

$$\int_a^b f(x)g'(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,dx.$$

Proof. By part (2) of the fundamental theorem of calculus, $\int_a^b (fg)'(x)\,dx = f(b)g(b) - f(a)g(a)$. Since $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$, we can subtract the integral of $f'(x)g(x)$ on both sides to get the integration by parts formula.

Theorem (Change of Variable Formula). *If $\phi : [a,b] \to \mathbb{R}$ is differentiable, ϕ' is integrable on $[a,b]$ and f is continuous on $\phi([a,b])$, then* $\int_{\phi(a)}^{\phi(b)} f(t)\,dt = \int_a^b f(\phi(x))\phi'(x)\,dx$.

Proof. Let $g(x) = \int_{\phi(a)}^{\phi(x)} f(t)\,dt$. By part (1) of the fundamental theorem of calculus and chain rule, $g'(x) = f(\phi(x))\phi'(x)$. So $\int_a^b f(\phi(x))\phi'(x)\,dx = \int_a^b g'(x)\,dx = g(b) - g(a) = g(b) = \int_{\phi(a)}^{\phi(b)} f(t)\,dt$.

Appendix 1: Proof of Lebesgue's Theorem

We first modify the uniform continuity theorem to obtain a lemma.

Lemma. *Let $f : [a,b] \to \mathbb{R}$ be a function continuous on a subset $K = [a,b] \setminus \bigcup_{j=1}^{\infty} (\alpha_j, \beta_j)$. Then for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x \in K$, $t \in [a,b]$, $|x - t| < \delta$ implies $|f(x) - f(t)| < \varepsilon$.*

Proof. Suppose the lemma is false. Then there is an $\varepsilon > 0$ such that for every $\delta = \frac{1}{n}$, $n \in \mathbb{N}$, there are $x_n \in K$, $t_n \in [a,b]$ such that $|x_n - t_n| < \delta = \frac{1}{n}$ and $|f(x_n) - f(t_n)| \geq \varepsilon$. By the Bolzano-Weierstrass theorem, $\{x_n\}$ has a subsequence $\{x_{n_j}\}$ converging to some $w \in [a,b]$. Since

$$|t_{n_j} - w| \leq |t_{n_j} - x_{n_j}| + |x_{n_j} - w| \leq \frac{1}{n_j} + |x_{n_j} - w|,$$

by the sandwich theorem, $\{t_{n_j}\}$ converges to w.

Next we will show $w \in K$. Suppose $w \notin K$, then $w \in (\alpha_i, \beta_i)$ for some i. Since $\lim_{j\to\infty} x_{n_j} = w$ and $w \in (\alpha_i, \beta_i)$, by the definition of limit, there will be some $x_{n_p} \in (\alpha_i, \beta_i)$, contradicting $x_{n_p} \in K = [a,b] \setminus \bigcup_{j=1}^{\infty} (\alpha_j, \beta_j)$. So $w \in K$.

Since f is continuous at w, by the sequential continuity theorem,

$$0 = |f(w) - f(w)| = \lim_{j\to\infty} |f(x_{n_j}) - f(t_{n_j})| \geq \varepsilon > 0,$$

which is a contradiction.

Proof of Lebesgue's Theorem. First suppose $f : [a,b] \to \mathbb{R}$ is integrable. Note that if f is discontinuous at x, then there is an $\varepsilon > 0$ such that for every $\delta > 0$, there is $z \in (x-\delta, x+\delta) \cap [a,b]$ such that $|f(x) - f(z)| \geq \varepsilon$. Let D_k be the set of all $x \in [a,b]$ such that for every open interval I with $x \in I$, there is $z \in I \cap [a,b]$ such that $|f(x) - f(z)| > \frac{1}{k}$. Since every positive $\varepsilon > \frac{1}{k}$ for some positive integer k by the Archimedean principle and every open interval I with $x \in I$ contains an interval $(x-\delta, x+\delta)$ for some $\delta > 0$, it follows that $S_f = \{x \in [a,b] : f \text{ is discontinuous at } x\} = \bigcup_{k=1}^{\infty} D_k$. We will show each D_k is of measure 0, which will imply S_f is of measure 0.

For every $\varepsilon > 0$, by the integral criterion, there is a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a,b]$ such that $U(f,P) - L(f,P) < \frac{\varepsilon}{2k}$. If there is $x \in D_k \cap (x_{j-1}, x_j)$, then $M_j - m_j \geq |f(x) - f(z)| > \frac{1}{k}$ for some $z \in (x_{j-1}, x_j)$. Let J be the set of j such that $D_k \cap (x_{j-1}, x_j) \neq \emptyset$. Then $D_k \setminus \{x_0, x_1, \ldots, x_n\} \subseteq \bigcup_{j\in J} (x_{j-1}, x_j)$ and $\sum_{j\in J} |x_{j-1} - x_j| \leq \sum_{j\in J} k \underbrace{(M_j - m_j)}_{>1/k} \Delta x_j \leq k\big(U(f,P) - L(f,P)\big) < \frac{\varepsilon}{2}$. Next around each x_j, we take an open interval

I_j containing x_j with length $\dfrac{\varepsilon}{2(n+1)}$. Then the intervals (x_{j-1}, x_j) with $j \in J$ and I_j's together contain D_k and the sum of their lengths is less than ε. So each D_k (and hence S_f) is of measure 0.

Conversely, suppose S_f is of measure 0. For every $\varepsilon > 0$, let $\varepsilon_0 = \dfrac{\varepsilon}{3(N+b-a)}$, where N is an upper bound for $|f(x)|$ on $[a,b]$. By the definition of measure 0, there are intervals (α_i, β_i) such that $S_f \subseteq \bigcup_{i=1}^{\infty} (\alpha_i, \beta_i)$ and $\sum_{i=1}^{\infty} |\alpha_i - \beta_i| < \varepsilon_0$. Then f is continuous on $K = [a,b] \setminus \bigcup_{i=1}^{\infty} (\alpha_i, \beta_i)$. Now take the δ in the lemma for ε_0 and a partition $P = \{x_0, x_1, \ldots, x_n\}$ such that $|x_{j-1} - x_j| < \delta$ for $j = 1, 2, \ldots, n$. If $K \cap [x_{j-1}, x_j] = \emptyset$, then $[x_{j-1}, x_j] \subseteq \bigcup_{i=1}^{\infty} (\alpha_i, \beta_i)$. If there is $x \in K \cap [x_{j-1}, x_j]$, then $M_j - m_j = \sup\{|f(t_0) - f(t_1)| : t_0, t_1 \in [x_{j-1}, x_j]\} \le \sup\{|f(t_0) - f(x)| + |f(x) - f(t_1)| : t_0, t_1 \in [x_{j-1}, x_j]\} \le 2\varepsilon_0$. So

$$U(f,P) - L(f,P) = \sum_{K \cap [x_{j-1}, x_j] = \emptyset} \underbrace{(M_j - m_j)}_{\le 2N} \Delta x_j + \sum_{K \cap [x_{j-1}, x_j] \ne \emptyset} \underbrace{(M_j - m_j)}_{\le 2\varepsilon_0} \Delta x_j \le 2N\varepsilon_0 + 2\varepsilon_0 (b-a) < \varepsilon.$$

Therefore, by the integral criterion, f is integrable on $[a,b]$.

Appendix 2: Riemann's Definition of the Integral

Recall the definition of $\lim_{x \to 0} f(x) = L$ is that for every $\varepsilon > 0$, there is $\delta > 0$ such that for every x with $|x| < \delta$, we have $|f(x) - L| < \varepsilon$. Based on this, Riemann's approach of integral can be defined as follow.

Definition. Let f be a bounded function on $[a,b]$. We write $\lim_{\|P\| \to 0} \sum_{j=1}^{n} f(t_j) \Delta x_j = I$ iff for every $\varepsilon > 0$, there is a $\delta > 0$ such that for every partition P of $[a,b]$ with $\|P\| < \delta$ and every choice $t_j \in [x_{j-1}, x_j]$, $j = 1, 2, \ldots, n$, we have $\left| \sum_{j=1}^{n} f(t_j) \Delta x_j - I \right| < \varepsilon$. (Such I is unique as in the proof of uniqueness of limit.)

Below we will show Darboux's approach is the same as Riemann's by establishing the following theorem.

Darboux's Theorem. *Let f be a bounded function on $[a,b]$. Then the following are equivalent:*

(a) $\int_a^b f(x)\,dx = I$;

(b) $\lim_{\|P\| \to 0} \sum_{j=1}^{n} f(t_j) \Delta x_j = I$;

(c) for every $\varepsilon > 0$, there is a partition P of $[a,b]$ such that for every refinement $Q = \{x_0, x_1, \ldots, x_n\}$ of P, every choice $t_j \in [x_{j-1}, x_j]$, we have $\left| \sum_{j=1}^{n} f(t_j) \Delta x_j - I \right| < \varepsilon$. (Such a number I is unique as in the proof of uniqueness of limit.)

Proof. $(a) \Rightarrow (b)$ Suppose $\int_a^b f(x)\,dx = I$. For every $\varepsilon > 0$, by the proof of the integral criterion, there is a partition $P_\varepsilon = \{w_0, w_1, \ldots, w_q\}$ of $[a,b]$ such that $I - \frac{\varepsilon}{2} < L(f, P_\varepsilon) \le U(f, P_\varepsilon) < I + \frac{\varepsilon}{2}$. Let $\delta = \dfrac{\varepsilon}{4qK}$, where $K = \sup\{|f(x)| : x \in [a,b]\}$. If $P = \{x_0, x_1, \ldots, x_n\}$ is a partition of $[a,b]$ with $\|P\| < \delta$, then $U(f,P) = S_1 + S_2$, where S_1 is the sum of the terms from intervals not containing points in P_ε and S_2 the sum of the remaining terms. Note the rectangles for the terms of S_1 are inside the rectangles for $U(f, P_\varepsilon)$. So we have $S_1 < U(f, P_\varepsilon) < I + \frac{\varepsilon}{2}$. Since $P_\varepsilon = \{w_0, w_1, \ldots, w_q\}$, S_2 has at most $2q$ terms (one for w_0, w_q each and at most two for $w_1, \ldots, w_{q-1}$

each). Then $S_2 \le 2q(K\|P\|) \le 2qK\delta = \frac{\varepsilon}{2}$. Thus $U(f,P) = S_1 + S_2 < I + \varepsilon$. Similarly, $L(f,P) > I - \varepsilon$. So $I - \varepsilon < L(f,P) \le \sum_{j=1}^{n} f(t_j)\Delta x_j \le U(f,P) < I + \varepsilon$. Then $\left|\sum_{j=1}^{n} f(t_j)\Delta x_j - I\right| < \varepsilon$.

$(b) \Rightarrow (c)$ For every $\varepsilon > 0$, take δ as the definition above. Let P be any partition of $[a,b]$ with $\|P\| < \delta$. Then for every refinement Q of P, we have $\|Q\| \le \|P\| < \delta$ and so (c) follows from the conclusion of (b).

$(c) \Rightarrow (a)$ For $\varepsilon > 0$, let P be the partition in (c) for $\frac{\varepsilon}{3}$. For each $j = 1, 2, \ldots, n$, by the supremum limit theorem, there are sequences $t_{j,k}$ such that $\lim_{k\to\infty} f(t_{j,k}) = M_j = \sup\{f(x) : x \in [x_{j-1}, x_j]\}$. Since $\left|\sum_{j=1}^{n} f(t_{j,k})\Delta x_j - I\right| < \frac{\varepsilon}{3}$, so taking limit, we get $|U(f,P) - I| \le \frac{\varepsilon}{3}$. Similarly, we can also get $|L(f,P) - I| \le \frac{\varepsilon}{3}$. Then $|U(f,P) - L(f,P)| \le \frac{2\varepsilon}{3} < \varepsilon$. By the integral criterion, it follows that $\int_a^b f(x)\,dx$ exists. By $(a) \Rightarrow (b)$ and $(b) \Rightarrow (c)$, it must equal I by uniqueness.

Improper Riemann Integral – Integration of Unbounded Functions or on Unbounded Intervals

We would like to integrate unbounded functions on unbounded intervals by taking limit of integrals on bounded intervals. Since the functions may or may not be continuous, we have to make sure the functions are integrable on bounded intervals first.

Definition. Let I be an interval. We say $f : I \to \mathbb{R}$ is *locally integrable* on I iff f is integrable on every closed and bounded subintervals of I. We denote this by $f \in L_{\text{loc}}(I)$.

Roughly, there are three cases of improper integrals.

Case 1: (Unbounded near one endpoint) Let f be locally integrable on $[a,b)$, where a is real and b is real or $+\infty$. We define the *improper (Riemann) integral* of f on $[a,b)$ to be $\int_a^b f(x)\,dx = \lim_{d\to b^-} \int_a^d f(x)\,dx$, provided the limit exists in $\mathbb{R}$, and we say f is *improper integrable* on $[a,b)$ in that case. For f locally integrable on an interval of the form $(a,b]$, where a is real or $-\infty$ and b is real, the definitions of the improper (Riemann) integral of f on $(a,b]$ and f is improper (Riemann) integrable on $(a,b]$ are similar.

Case 2: (Unbounded near both endpoints) Let f be locally integrable on (a,b), with a, b real or infinity, and $x_0 \in (a,b)$. We define the *improper (Riemann) integral of f on (a,b)* to be $\int_a^b f(x)\,dx = \lim_{c\to a^+} \int_c^{x_0} f(x)\,dx + \lim_{d\to b^-} \int_{x_0}^d f(x)\,dx$ provide both limits exist in $\mathbb{R}$, and say f is *improper integrable* on (a,b) in that case. (*Note the integral is the same regardless of the choices of $x_0 \in (a,b)$.*)

Case 3: (Unbounded inside interval) Let f be improper integrable on intervals $[a,c)$ and $(c,b]$. The *improper integral* of f on $[a,b]$ is the sum of the improper integrals on the two intervals. The cases where a or b is excluded are similarly defined.

In each case, if the improper integral is a number, then we say the improper integral *converges* to the number, otherwise we say it *diverges.*

Examples.(1) Consider the unbounded function $f(x) = \ln x$ on $(0,1]$. Observe that f is continuous, hence integrable, on every closed subinterval of $(0,1]$. So $f \in L_{\text{loc}}(0,1]$. Now $\lim_{c\to 0^+} \int_c^1 \ln x\,dx = \lim_{c\to 0^+}(-1 - c\ln c + c) = -1$, so $\int_0^1 \ln x\,dx = -1$ and $\ln x$ is improper integrable on $(0,1]$.

(2) Consider $f(x) = \frac{1}{x^2}$ on the unbounded interval $[2, +\infty)$. Observe that $f \in L_{\text{loc}}[2, +\infty)$ because of continuity. Now $\lim_{d\to+\infty} \int_2^d \frac{1}{x^2}\,dx = \lim_{d\to+\infty} \left(-\frac{1}{d} + \frac{1}{2}\right) = \frac{1}{2}$, so $\int_2^\infty \frac{1}{x^2}\,dx = \frac{1}{2}$ and $\frac{1}{x^2}$ is improper integrable on $[2, +\infty)$.

(3) Consider $f(x) = e^x$ on $\mathbb{R} = (-\infty, +\infty)$. Since f is continuous everywhere, $f \in L_{\text{loc}}(-\infty, +\infty)$. Now take a $x_0 \in (-\infty, +\infty)$, say $x_0 = 0$. Then $\lim_{c\to-\infty} \int_c^0 e^x dx = \lim_{c\to-\infty} (1 - e^c) = 1$, but $\lim_{d\to+\infty} \int_0^d e^x dx = \lim_{d\to+\infty} (e^d - 1)$ does not exist in $\mathbb{R}$. So e^x is not improper integrable on $(-\infty, +\infty)$.

Remarks. By taking limits, many properties of Riemann integrals extend to improper Riemann integrals as well. There are also some helpful tests to determine if a function is improper integrable.

***p*-test.** *For $0 < a < \infty$, we have $\int_a^\infty \frac{1}{x^p}\,dx < \infty$ if and only if $p > 1$. Also, $\int_0^a \frac{1}{x^p}\,dx < \infty$ if and only if $p < 1$.*

Proof. In the proof of p-test for series, we saw $\int_1^\infty \frac{1}{x^p}\,dx < \infty$ if and only if $p > 1$. Adding $\int_a^1 \frac{1}{x^p}\,dx < \infty$, we get the first statement. Next, by the change of variable $y = \frac{1}{x}$, we have $\int_c^1 \frac{1}{x^p}\,dx = \int_1^{1/c} \frac{1}{y^{2-p}}\,dy$. Taking limit as $c \to 0+$, we see that $\int_0^1 \frac{1}{x^p}\,dx < \infty$ if and only if $\int_1^\infty \frac{1}{y^{2-p}}\,dy < \infty$. By the first statement, this is the same as $2 - p > 1$, i.e. $p < 1$. Again adding $\int_1^a \frac{1}{x^p}\,dx < \infty$, we get the second statement.

Theorem (Comparison Test). *Suppose $0 \le f(x) \le g(x)$ on I and $f, g \in L_{\text{loc}}(I)$. If g is improper integrable on I, then f is improper integrable on I. If f is not improper integrable on I, then g is not improper integrable on I.*

Proof. For the case $I = [a, b)$, if $0 \le f \le g$ on I and g is improper integrable on I, then $\int_a^d f(x)\,dx$ is an increasing function of d and is bounded above by $\int_a^b g(x)\,dx$. So $\int_a^b f(x)\,dx = \lim_{d\to b^-} \int_a^d f(x)\,dx$ exists in $\mathbb{R}$ by the monotone function theorem. The other cases $I = (a, b]$ and (a, b) are similar.

Theorem (Limit Comparison Test). *Suppose $f(x), g(x) > 0$ on $(a, b]$ and $f, g \in L_{\text{loc}}((a, b])$. If $\lim_{x\to a+} \frac{g(x)}{f(x)}$ is a positive number L, then either (both $\int_a^b f(x)\,dx$ and $\int_a^b g(x)\,dx$ converge) or (both diverge). If $\lim_{x\to a+} \frac{g(x)}{f(x)} = 0$, then $\int_a^b f(x)\,dx$ converges implies $\int_a^b g(x)\,dx$ converges. If $\lim_{x\to a+} \frac{g(x)}{f(x)} = \infty$, then $\int_a^b f(x)\,dx$ diverges implies $\int_a^b g(x)\,dx$ diverges. In the case $[a, b)$, we take $\lim_{x\to b-} \frac{g(x)}{f(x)}$ and the results are similar.*

Proof. If $\lim_{x\to a+} \frac{g(x)}{f(x)}$ is a positive number L, then for $\varepsilon = \frac{L}{2} > 0$, there is $\delta > 0$ such that for every $x \in (a, a+\delta)$, we have $\frac{L}{2} = L - \varepsilon < \frac{g(x)}{f(x)} < L + \varepsilon = \frac{3L}{2}$. Then $\frac{L}{2} \int_a^{a+\delta} f(x)\,dx \le \int_a^{a+\delta} g(x)\,dx \le \frac{3L}{2} \int_a^{a+\delta} f(x)\,dx$. So either (both $\int_a^b f(x)\,dx$ and $\int_a^b g(x)\,dx$ converge) or (both diverge).

If $\lim_{x\to a+} \frac{g(x)}{f(x)} = 0$, then there exists $\delta' > 0$ such that for every $x \in (a, a + \delta')$, we have $0 < \frac{g(x)}{f(x)} < 1$, which implies $0 < g(x) < f(x)$. Then $0 \le \int_a^{a+\delta'} g(x)\,dx \le \int_a^{a+\delta'} f(x)\,dx$. So $\int_a^b f(x)\,dx < \infty \Rightarrow \int_a^b g(x)\,dx < \infty$.

If $\lim_{x\to a+} \frac{g(x)}{f(x)} = \infty$, then there is $\delta'' > 0$ such that for every $x \in (a, a+\delta'')$, we have $\frac{g(x)}{f(x)} > 1$, which implies $g(x) > f(x)$. Hence $\int_a^{a+\delta''} g(x)\,dx \geq \int_a^{a+\delta''} f(x)\,dx$. So $\int_a^b f(x)\,dx = \infty \Rightarrow \int_a^b g(x)\,dx = \infty$.

Theorem (Absolute Convergence Test). *If $f \in L_{\mathrm{loc}}(I)$ and $|f|$ is improper integrable on I, then f is improper integrable on I.*

Proof. We have $-|f| \leq f \leq |f|$ on I, which implies $0 \leq f + |f| \leq 2|f|$ on I. By the comparison test, $f + |f|$ is improper integrable on I. Therefore, $f = (f + |f|) - |f|$ is improper integrable on I.

Examples. (4) Is $\frac{\ln x}{1+x^2}$ improper integrable on $(0, 1]$? Observe first that the function is locally integrable on $(0, 1]$ by continuity. Also, $\left|\frac{\ln x}{1+x^2}\right| \leq |\ln x|$ on $(0, 1]$. Since $|\ln x| = -\ln x$ is improper integrable on $(0, 1]$ (cf. example (1)), so $\left|\frac{\ln x}{1+x^2}\right|$ is improper integrable on $(0, 1]$ by the comparison test. Then $\frac{\ln x}{1+x^2}$ is improper integrable on $(0, 1]$ by the absolute convergence test.

(5) Does $\int_2^{+\infty} \frac{dx}{\sqrt{x^2-1}}$ converge? Observe that on $[2, +\infty)$, $0 < \frac{1}{x} \leq \frac{1}{\sqrt{x^2-1}}$. Both $\frac{1}{x}$ and $\frac{1}{\sqrt{x^2-1}}$ are continuous, hence locally integrable, on $[2, +\infty)$. Now $\int_2^{+\infty} \frac{1}{x} dx = \infty$ by p-test. By the comparison test, $\int_2^{+\infty} \frac{dx}{\sqrt{x^2-1}}$ diverges.

(6) Does $\int_1^{+\infty} \frac{\sin x}{x}\,dx$ converge? Since $\frac{\sin x}{x}$ is continuous on $[1, +\infty)$, it is locally integrable there. Integrating by parts,

$$\int_1^c \frac{\sin x}{x}\,dx = \frac{-\cos c}{c} + \cos 1 - \int_1^c \frac{\cos x}{x^2}\,dx.$$

Since $|\cos c| \leq 1$, $\lim_{c\to+\infty} \frac{\cos c}{c} = 0$. Since $\left|\frac{\cos x}{x^2}\right| \leq \frac{1}{x^2}$ on $[1, +\infty)$ and $\int_1^{+\infty} \frac{1}{x^2}\,dx < \infty$ by p-test, we have $\int_1^{+\infty} \frac{\cos x}{x^2}\,dx$ converges by the comparison test and the absolute convergence test. Therefore, $\int_1^{+\infty} \frac{\sin x}{x}\,dx$ converges.

(7) For the improper integral $\int_0^1 \frac{dx}{1-x^3}$, the integrand is continuous (hence locally integrable) on $[0, 1)$. Observe that $\frac{1}{1-x^3} = \frac{1}{1-x}\left(\frac{1}{1+x+x^2}\right)$ and the second factor on the right has a positive limit as $x \to 1-$. More precisely, since $\lim_{x\to 1-} \frac{1/(1-x)}{1/(1-x^3)} = \lim_{x\to 1-} (1+x+x^2) = 3$ and $\int_0^1 \frac{dx}{1-x} = \lim_{d\to 1-} \int_0^d \frac{dx}{1-x} = \lim_{d\to 1-} -\ln(1-d) = \infty$, by the limit comparison test, $\int_0^1 \frac{dx}{1-x^3}$ diverges.

(8) For the improper integral $\int_0^5 \frac{dx}{\sqrt[3]{7x+2x^4}}$, the integrand is continuous (hence locally integrable) on $(0, 5]$. Observe that $\frac{1}{\sqrt[3]{7x+2x^4}} = \frac{1}{\sqrt[3]{x}}\left(\frac{1}{\sqrt[3]{7+2x^3}}\right)$ and the second factor on the right has a positive limit as $x \to 0+$. More precisely, since $\lim_{x\to 0+} \frac{1/\sqrt[3]{x}}{1/\sqrt[3]{7x+2x^4}} = \sqrt[3]{7}$ and $\int_0^5 \frac{dx}{\sqrt[3]{x}} < \infty$ by p-test as $p = \frac{1}{3} < 1$, by the limit comparison test, $\int_0^5 \frac{dx}{\sqrt[3]{7x+2x^4}}$ converges.

Principal Value – Symmetric Integration about Singularities.

Definition. Let $f \in L_{\text{loc}}(\mathbb{R})$, then the *principal value* of $\int_{-\infty}^{\infty} f(x)\,dx$ is P.V. $\int_{-\infty}^{\infty} f(x)\,dx = \lim_{c\to+\infty} \int_{-c}^{c} f(x)\,dx$.

Examples. (1) Because of continuity, $f(x) = \dfrac{1}{1+x^2} \in L_{\text{loc}}(\mathbb{R})$. Now P.V. $\int_{-\infty}^{\infty} \dfrac{1}{1+x^2}\,dx = \lim_{c\to+\infty} \int_{-c}^{c} \dfrac{1}{1+x^2}\,dx$ $= \lim_{c\to+\infty} (2\tan^{-1} c) = \pi$ exists and the improper integral

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = \lim_{c\to-\infty} \int_{c}^{0} \frac{1}{1+x^2}\,dx + \lim_{d\to+\infty} \int_{0}^{d} \frac{1}{1+x^2}\,dx = \lim_{c\to-\infty} (-\tan^{-1} c) + \lim_{d\to+\infty} (\tan^{-1} d) = \pi.$$

(2) Because of continuity, $f(x) = x \in L_{\text{loc}}(\mathbb{R})$. Now P.V. $\int_{-\infty}^{\infty} x\,dx = \lim_{c\to+\infty} \int_{-c}^{c} x\,dx = \lim_{c\to+\infty} 0 = 0$ exists, however the improper integral $\int_{-\infty}^{\infty} x\,dx = \lim_{c\to-\infty} \int_{c}^{0} x\,dx + \lim_{d\to+\infty} \int_{0}^{d} x\,dx = \lim_{c\to-\infty} (-\frac{c^2}{2}) + \lim_{d\to+\infty} (\frac{d^2}{2})$ does not exist.

Theorem. *If the improper integral* $\int_{-\infty}^{\infty} f(x)\,dx$ *exists, then* P.V. $\int_{-\infty}^{\infty} f(x)\,dx$ *exists and equals the improper integral* $\int_{-\infty}^{\infty} f(x)\,dx$. *(Note the converse is false by example (2).)*

Proof. If the improper integral $\int_{-\infty}^{\infty} f(x)\,dx$ exists, then it implies that both $\lim_{d\to-\infty} \int_{d}^{0} f(x)\,dx$ and $\lim_{c\to+\infty} \int_{0}^{c} f(x)\,dx$ exist. So

$$\begin{aligned} \text{P.V.} \int_{-\infty}^{\infty} f(x)\,dx &= \lim_{c\to+\infty} \int_{-c}^{c} f(x)\,dx = \lim_{c\to+\infty} \Big(\int_{-c}^{0} f(x)\,dx + \int_{0}^{c} f(x)\,dx \Big) \\ &= \lim_{d\to-\infty} \int_{d}^{0} f(x)\,dx + \lim_{c\to+\infty} \int_{0}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} f(x)\,dx. \end{aligned}$$

Definition. For $a, b \in \mathbb{R}$ or ∞, let $f\colon [a, c) \cup (c, b] \to \mathbb{R}$ be locally integrable on $[a, c)$ and on $(c, b]$. Define

$$\text{P.V.} \int_{a}^{b} f(x)\,dx = \lim_{\varepsilon\to0^+} \left(\int_{a}^{c-\varepsilon} f(x)\,dx + \int_{c+\varepsilon}^{b} f(x)\,dx \right).$$

Example. Consider the improper sense and the principal value sense of $\int_{-1}^{1} \frac{1}{x}\,dx$.

Because of continuity, $\frac{1}{x} \in L_{\text{loc}}[-1, 0)$ and $\frac{1}{x} \in L_{\text{loc}}(0, 1]$. In the improper sense, $\int_{0}^{1} \frac{1}{x}\,dx = \lim_{c\to0^+} \int_{c}^{1} \frac{1}{x}\,dx$ $= \lim_{c\to0^+} (-\ln c)$ does not exist. So the improper integral on $[-1, 0) \cup (0, 1]$ does not exist. However,

$$\text{P.V.} \int_{-1}^{1} \frac{1}{x}\,dx = \lim_{\varepsilon\to0^+} \left(\int_{-1}^{-\varepsilon} \frac{1}{x}\,dx + \int_{\varepsilon}^{1} \frac{1}{x}\,dx \right) = \lim_{\varepsilon\to0^+} (\ln|-\varepsilon| - \ln|\varepsilon|) = 0.$$

Remarks. (1) It is incorrect to say $\int_{-1}^{1} \frac{1}{x}\,dx = \ln|-1| - \ln|1| = 0$, which attempts to use part (2) of the fundamental theorem of calculus, but here $f(x) = \ln|x|$ is not differentiable on $[-1, 1]$ failing the required condition.

(2) There is a similar theorem for this type of principal value integrals as the theorem above.

Taylor Series of Common Functions

Question. How do calculators compute $\sin x$, $\cos x$, e^x, x^y, $\ln x, \ldots$?

Recall Taylor's theorem with the Lagrange form of the remainder is the following.

Taylor's Theorem. *Let $f\colon (a,b) \to \mathbb{R}$ be n times differentiable on (a,b). For every $x, c \in (a,b)$, there is x_0 between x and c such that*

$$f(x) = f(c) + \frac{f'(c)}{1!}(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \ldots + \frac{f^{(n-1)}(c)}{(n-1)!}(x-c)^{n-1} + \frac{f^{(n)}(x_0)}{n!}(x-c)^n.$$

(This is called the n-th Taylor expansion of f about c. The term $R_n(x) = \frac{f^{(n)}(x_0)}{n!}(x-c)^n$ is called the Lagrange form of the remainder.)

There are two other forms of the remainder.

Theorem (Taylor's Formula with Integral Remainder). *Let f be n times differentiable on (a,b). For every $x, c \in (a,b)$, if $f^{(n)}$ is integrable on the closed interval with endpoints x and c, then*

$$f(x) = f(c) + \frac{f'(c)}{1!}(x-c) + \cdots + \frac{f^{(n-1)}(c)}{(n-1)!}(x-c)^{n-1} + R_n(x), \quad \textit{where } R_n(x) = \frac{1}{(n-1)!}\int_c^x (x-t)^{n-1} f^{(n)}(t)\,dt.$$

Proof. Note $\frac{d}{dt}\big(-(x-t)\big) = 1$. Applying integration by parts $n-1$ times, we get

$$\begin{aligned} f(x) - f(c) &= \int_c^x f'(t)\cdot 1\,dt = \int_c^x f'(t)\big(-(x-t)\big)'\,dt \\ &= -f'(t)(x-t)\Big|_c^x + \int_c^x (x-t) f''(t)\,dt \\ &= f'(c)(x-c) + \Big(-f''(t)\frac{(x-t)^2}{2!}\Big|_c^x + \frac{1}{2!}\int_c^x (x-t)^2 f'''(t)\,dt\Big) = \cdots \\ &= f'(c)(x-c) + \cdots + \frac{f^{(n-1)}(c)}{(n-1)!}(x-c)^{n-1} + \frac{1}{(n-1)!}\int_c^x (x-t)^{n-1} f^{(n)}(t)\,dt. \end{aligned}$$

Mean Value Theorem for Integrals. *If f is continuous on $[a,b]$, then $\int_a^b f(x)\,dx = f(x_1)(b-a)$ for some $x_1 \in [a,b]$, (i.e. the mean value or average of f on $[a,b]$ is $\frac{1}{b-a}\int_a^b f(x)\,dx = f(x_0)$). More generally, if g is integrable and $g \ge 0$ on $[a,b]$, then $\int_a^b f(x)g(x)\,dx = f(x_1)\int_a^b g(x)\,dx$ for some $x_1 \in [a,b]$.*

Proof. The first statement is the special case of the second statement when $g(x) = 1$. So we only need to prove the second statement. Let M and m be the maximum and the minimum of f on $[a,b]$, respectively. Since $m \le f(x) \le M$ on $[a,b]$, we have $m\int_a^b g(x)\,dx \le \int_a^b f(x)g(x)\,dx \le M\int_a^b g(x)\,dx$. If $\int_a^b g(x)\,dx = 0$, then the last sentence implies $\int_a^b f(x)g(x)\,dx = 0$ and so we may take x_1 to be any element of $[a,b]$. If $\int_a^b g(x)\,dx > 0$, then dividing by $\int_a^b g(x)\,dx$, we see that $\int_a^b f(x)g(x)\,dx \Big/ \int_a^b g(x)\,dx$ is between m and M. By the intermediate value theorem, this expression equals $f(x_1)$ for some $x_1 \in [a,b]$, which gives the second statement.

Theorem (Taylor's Formula with Cauchy Form Remainder). *Let f be n times differentiable on (a,b). For $x, c \in (a,b)$, if $f^{(n)}$ is continuous on the closed interval with endpoints x and c, then there is x_1 between x and c such that*

$$f(x) = f(c) + \frac{f'(c)}{1!}(x-c) + \cdots + \frac{f^{(n-1)}(c)}{(n-1)!}(x-1)^{n-1} + R_n(x), \quad \textit{where } R_n(x) = \frac{(x-c)(x-x_1)^{n-1} f^{(n)}(x_1)}{(n-1)!}.$$

Proof. This follows by applying the mean value theorem for integrals to the integral remainder of Taylor's formula above.

Remarks. For every $x \in \mathbb{R}$, the series $\sum_{k=0}^{\infty} \frac{x^k}{k!}$ converges. This follows from the ratio test because $\lim_{k\to\infty} \left| \frac{x^{k+1}}{(k+1)!} \frac{k!}{x^k} \right| = \lim_{k\to\infty} \frac{|x|}{k+1} = 0 < 1$. By term test, we have $\lim_{k\to\infty} \frac{x^k}{k!} = 0$ for every $x \in \mathbb{R}$.

Examples.

(1) For $f(x) = \sin x$, $f^{(n)}(x) = \begin{cases} (-1)^k \cos x & \text{if } n = 2k+1 \\ (-1)^k \sin x & \text{if } n = 2k \end{cases}$. So $|f^{(n)}(x)| \le 1$ for every $x \in \mathbb{R}$.
Taking $c = 0$, we have $f^{(2k+1)}(0) = (-1)^k$ and $f^{(2k)}(0) = 0$ for every $k \in \mathbb{N}$. By Taylor's theorem, $\sin x = \sum_{k=0}^{n-1} \frac{(-1)^k x^{2k+1}}{(2k+1)!} + R_{2n}(x)$ for every $x \in \mathbb{R}$. Now $|R_{2n}(x)| \le \frac{1}{(2n)!}|x|^{2n} \to 0$ as $n \to \infty$ by the remarks above. Therefore,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} \qquad \text{for } -\infty < x < +\infty.$$

Remarks. For $0 \le x \le \frac{\pi}{2}$, $|R_{18}(x)| \le \frac{|x|^{18}}{18!} \le \frac{(\pi/2)^{18}}{18!} < 6 \times 10^{-13}$. So $x - \frac{x^3}{3!} + \cdots + \frac{x^{17}}{17!}$ can be used to compute $\sin x$ to 10 decimal places.

---

(2) For $f(x) = \cos x$, $f^{(n)}(x) = \begin{cases} (-1)^k \sin x & \text{if } n = 2k-1 \\ (-1)^k \cos x & \text{if } n = 2k \end{cases}$. So $|f^{(n)}(x)| \le 1$ for every $x \in \mathbb{R}$.
Taking $c = 0$, we have $f^{(2k-1)}(0) = 0$ and $f^{(2k)}(0) = (-1)^k$ for every $k \in \mathbb{N}$. By Taylor's theorem, $\sin x = \sum_{k=0}^{n-1} \frac{(-1)^k x^{2k}}{(2k)!} + R_{2n-1}(x)$ for every $x \in \mathbb{R}$. Now $|R_{2n-1}(x)| \le \frac{1}{(2n-1)!}|x|^{2n-1} \to 0$ as $n \to \infty$ by the remarks above. Therefore,

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{(2k)!} \qquad \text{for } -\infty < x < +\infty.$$

---

(3) For $f(x) = e^x$, $f^{(n)}(x) = e^x$. Taking $c = 0$, we have $f^{(n)}(0) = 1$ for every $n \in \mathbb{N}$. By Taylor's theorem, $e^x = \sum_{k=0}^{n-1} \frac{x^k}{k!} + R_n(x)$ for every $x \in \mathbb{R}$. Now $|R_n(x)| = \frac{e^{x_0}}{n!}|x|^n \le \frac{\max(e^0, e^x)}{n!}|x|^n \to 0$ as $n \to \infty$ by the remarks above. So

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{x^k}{k!} \qquad \text{for } -\infty < x < +\infty.$$

Remarks. The Taylor series for $\sin x$, $\cos x$ and e^x also converge if x is a complex number! In fact, this is how the sine, cosine, and exponential functions are defined for complex numbers. For every $x \in \mathbb{R}$, we have

$$e^{ix} = \sum_{k=0}^{\infty} \frac{(ix)^k}{k!} = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n+1}}{(2n+1)!} = \cos x + i \sin x.$$

In particular, we have $e^{i\pi} = \cos\pi + i\sin\pi = -1$ so that $e^{i\pi} + 1 = 0$. This is known as *Euler's formula.* It is the most beautiful formula in mathematics because it connects the five most important constants $1, 0, \pi, i, e$ of mathematics in one single equation!

---

(4) (*Binomial Theorem*) For $a \in \mathbb{R}$, if $-1 < x < 1$, then

$$(1+x)^a = 1 + ax + \frac{a(a-1)}{2!}x^2 + \frac{a(a-1)(a-2)}{3!}x^3 + \cdots = 1 + \sum_{k=1}^{\infty} \frac{a(a-1)\cdots(a-k+1)}{k!}x^k.$$

To see this, let $f(x) = (1+x)^a$, then $f^{(n)}(x) = a(a-1)\cdots(a-n+1)(1+x)^{a-n}$. By Taylor's formula with integral formula,

$$R_n(x) = \frac{1}{(n-1)!}\int_0^x (x-t)^{n-1} f^{(n)}(t)\,dt = \frac{a(a-1)\cdots(a-n+1)}{(n-1)!}\int_0^x \left(\frac{x-t}{1+t}\right)^{n-1}(1+t)^{a-1}\,dt.$$

For $x \in [0,1)$, the function $g(t) = \dfrac{x-t}{1+t}$ has derivative $g'(t) = \dfrac{-1-x}{(1+t)^2} < 0$. On $[0,x]$, $g(t) \le g(0) = x$. Let $k = \displaystyle\int_0^x (1+t)^{a-1}\,dt$, then $|R_n(x)| \le \underbrace{\dfrac{|a(a-1)\cdots(a-n+1)kx^{n-1}|}{(n-1)!}}_{\text{call this } b_n}$. Since $\displaystyle\lim_{n\to\infty}\frac{b_{n+1}}{b_n} = \lim_{n\to\infty}\frac{|a-n||x|}{n} = |x| < 1$, by ratio test, $\displaystyle\sum_{n=1}^{\infty} b_n$ converges. By term test, $\displaystyle\lim_{n\to\infty} b_n = 0$. Then $\displaystyle\lim_{n\to\infty} R_n(x) = 0$ by the sandwich theorem. So the binomial formula is true for $x \in [0,1)$. The case $x \in (-1,0]$ is similar.

Here are a few more common Taylor series. (Note the series only equal the functions on a small interval.) They are obtained from the cases $a = -1$, $a = -1$ (with x replaced by x^2) and $a = -\frac{1}{2}$ (with x replaced by $-x^2$) of the binomial theorem by term-by-term integration.

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots = \sum_{k=1}^{\infty}\frac{(-1)^{k-1}x^k}{k} \qquad \text{for } -1 < x \le 1$$

$$\tan^{-1} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots = \sum_{k=0}^{\infty}\frac{(-1)^k x^{2k+1}}{2k+1} \qquad \text{for } -1 \le x \le 1$$

$$\sin^{-1} x = x + \frac{1}{2}\frac{x^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{x^5}{5} + \cdots = x + \sum_{k=1}^{\infty}\frac{1\cdot 3\cdot 5\cdots(2k-1)}{2\cdot 4\cdot 6\cdots(2k)}\frac{x^{2k+1}}{2k+1} \qquad \text{for } -1 \le x \le 1$$

Remarks. Unfortunately, the Taylor series of a function does not always equal to the function away from the center. For example, the function $f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \ne 0 \\ 0 & \text{if } x = 0 \end{cases}$ can be shown to be infinitely differentiable and $f^{(n)}(0) = 0$ for every $n \in \mathbb{N}$. So the Taylor series of $f(x)$ about $c = 0$ is $\displaystyle\sum_{k=0}^{\infty} 0x^k = 0$, i.e. the Taylor series is the zero function. Therefore, the Taylor series of $f(x)$ equals $f(x)$ only at the center $c = 0$.