

Real Analysis

Muchang Bahng

November 6, 2022

1 The Real Numbers

The entirety of calculus is founded on the definition of the real numbers \mathbb{R} . Therefore, we must properly construct and define it. Before we do, we define order.

1.0.1 Order

Definition 1.1 (Partial, Total/Linear Order). A *partial order* on a set X is a relation \leq such that for any elements $x, y \in X$,

1. Reflexive: $x \leq x$
2. Antisymmetric: $x \leq y, y \leq x \implies x = y$
3. Transitivity: $x \leq y, y \leq z \implies x \leq z$

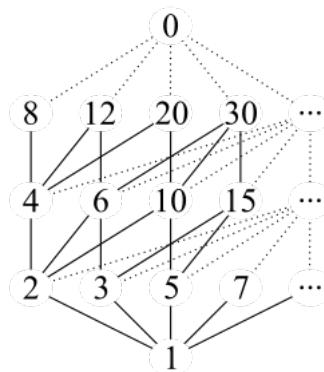
Note that when we say $x \leq y$, this means " x is related to y " (but does not necessarily mean that y is related to x), or " x is less than or equal to y ." A set X with a partial order is called a partially ordered set.

Additionally, given elements x, y of partially ordered set X , if either $x \leq y$ or $y \leq x$, then x and y are *comparable*. Otherwise, they are *incomparable*. A partial order in which every pair of elements is comparable is called a *total order*, or *linear order*. Note that from this \leq relation, we can similarly define

1. \leq : less than or equal to
2. \geq : greater than or equal to
3. $<$: strictly less than ($x < y$ iff $x \leq y, x \neq y$)
4. $>$: strictly greater than ($x > y$ iff $x \geq y, x \neq y$)

Example 1.1 (Partially Ordered Sets). We list some examples of partially ordered sets.

1. The real numbers ordered by the standard "less-than-or-equal" relation \leq (totally ordered set as well).
2. The set of subsets of a given set X ordered by inclusion. That is, the power set 2^X with the partial order \subseteq is partially ordered.
3. The set of natural numbers equipped with the relation of divisibility.



4. The set of subspaces of a vector space ordered by inclusion.
5. For a partially ordered set P , the sequence space containing all sequences of elements from P , where sequence a precedes sequence b if every item in a precedes the corresponding item in b .

Definition 1.2 (Extrema). We list 2 types of extrema of partially ordered sets.

1. *Greatest/Least Element*: An element $g \in P$ is a greatest element if

$$\text{for all } a \in P, a \leq g$$

and a least element if

$$\text{for all } a \in P, g \leq a$$

This means that the relation must exist between g and every other element in P . This also implies that a partially ordered set can only have one greatest and least element.

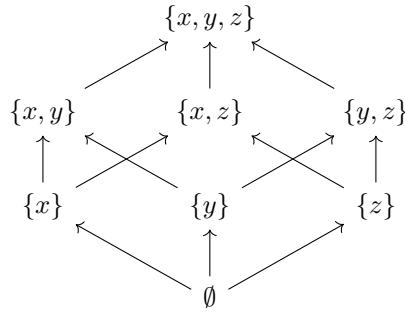
2. *Maximal/Minimal Elements*: An element $g \in P$ is a maximal element if

$$\text{there exists no } a \in P \text{ such that } a > g$$

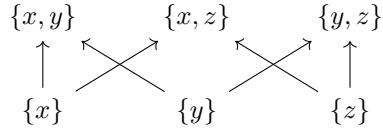
and a least element if

$$\text{there exists no } a \in P \text{ such that } a < g$$

The difference between these two can be seen in the following visual. In here, $A \rightarrow B$ means that $A \leq B$.



In here, the greatest and maximal element of $2^{\{x,y,z\}}$ is $\{x,y,z\}$ and the least and minimal element is \emptyset . However, in the following set



there exists no greatest nor least element. Furthermore, the maximal elements are $\{x,y\}, \{x,z\}, \{y,z\}$ while the minimal elements are $\{x\}, \{y\}, \{z\}$.

Definition 1.3 (Upper, Lower Bounds). We list two definitions for bounds.

1. *Upper/Lower Bounds*: For a subset $A \subset P$, an element $x \in P$ is an upper bound of A if

$$a \leq x \text{ for all } a \in A$$

and is a lower bound of A if

$$a \geq x \text{ for all } a \in A$$

2. *Supremum/Infimum*: For a subset $A \subset P$, an element $x \in P$ is a least upper bound, or supremum, if it is the smallest possible upper bound, and is a greatest lower bound, or infimum, if it is the greatest possible lower bound.

$$\begin{aligned}\sup(X) &\equiv \min \{c \in Z \mid \forall x \in X, x \leq c\} \\ \inf(X) &\equiv \max \{c \in Z \mid \forall x \in X, x \geq c\}\end{aligned}$$

The main difference between the supremum/infimum and greatest/least element is that the supremum/infimum accounts for limit points of the subset A .

1.1 Completeness

Intuitively, completeness implies that there are not any "gaps" or "missing points" in the real number line. This contrasts with the rational numbers, whose corresponding number line has a "gap" at each irrational value (this is formalized with Dedekind cuts). In the decimal number system, completeness is equivalent to the statement that any infinite string of decimal digits is actually a decimal representation of some number.

1.1.1 Least Upper Bound Property

Definition 1.4 (Least Upper Bound Completeness). A totally ordered algebraic field F is complete if every nonempty set of F having an upper bound must have a least upper bound (supremum) in F .

Lemma 1.1 (Least Upper Bound Completeness of \mathbb{R}). \mathbb{R} is least upper bound complete.

On the contrary, the rational number line \mathbb{Q} does not have the least upper bound property. Take the subset of rational numbers

$$S = \{x \in \mathbb{Q} \mid x^2 < 2\}$$

The least upper bound is $\sqrt{2}$, but this does not exist in \mathbb{Q} . We can try constructing smaller and smaller upper bounds of S in \mathbb{Q} , but by denseness of the rationals in \mathbb{R} , there is no end to this; we can always construct a smaller upper bound, infinitely.

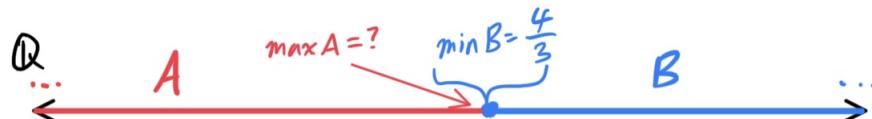


In the example above, we first construct rational upper bound 1.42, then 1.415, then 1.4143, and so on infinitely.

1.1.2 Dedekind Completeness

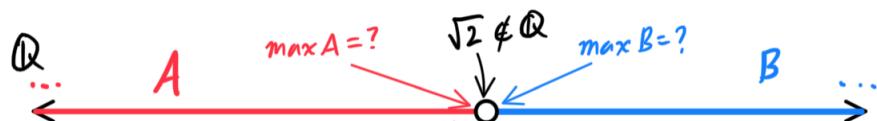
Definition 1.5 (Dedekind Cut). A *Dedekind cut* is a partition of the rational numbers into two sets A and B , such that all elements of A are less than all elements of B , and A contains no greatest element. The set B may or may not have a smallest element among the rationals.

1. If B has a smallest among the rationals, the cut corresponds to that rational, or in other words, the cut is generated by the rational number.



In the visual above, this Dedekind cut of \mathbb{Q} is generated by $4/3$.

2. Otherwise, that cut defines a unique irrational number which, loosely speaking, fills the "gap" between A and B . An irrational cut is equated to an irrational number which is in neither set.



In the visual above, this Dedekind cut of \mathbb{Q} is generated by $\sqrt{2}$, which is not in \mathbb{Q} .

Note that Dedekind cuts can be generalized from the rational numbers to any totally ordered set by defining the partition A and B , where every element in A is less than every element in B , and A contains no greatest element.

Definition 1.6 (Dedekind Completeness). A totally ordered algebraic field F is complete if every Dedekind cut of F is generated by an element of F .

Lemma 1.2 (Dedekind Completeness of \mathbb{R}). \mathbb{R} is Dedekind complete.

1.1.3 Cauchy Completeness

Cauchy completeness actually allows us to generalize completeness for any metric space, which may not need to be ordered. The reader may skip ahead to learn about Cauchy sequences before visiting this definition.

Definition 1.7 (Cauchy Completeness). A metric space (X, d) is complete if every Cauchy sequence in that space converges to an element in X .

Lemma 1.3 (Cauchy Completeness of \mathbb{R}). \mathbb{R} is Cauchy complete.

We can see that \mathbb{Q} is not Cauchy complete since the following rational sequence of successive approximations of π

$$3, \quad 3.1, \quad 3.14, \quad 3.142, \quad 3.1416, \dots$$

does not converge to any rational number since $\pi \notin \mathbb{Q}$.

1.1.4 Nested Intervals Theorem

Definition 1.8 (Nested Interval Completeness). Let F be a totally ordered algebraic field. Let $I_n = [a_n, b_n]$ ($a_n < b_n$) be a sequence of closed intervals, and suppose that these intervals are nested in the sense that

$$I_1 \supset I_2 \supset I_3 \supset \dots$$

where

$$\lim_{n \rightarrow +\infty} b_n - a_n = 0$$

F is complete if the intersection of all of these intervals I_n contains exactly one point. That is,

$$\bigcap_{n=1}^{\infty} I_n \in F$$

Lemma 1.4 (Nested Interval Completeness of \mathbb{R}). \mathbb{R} is Nested Interval complete.

We can also see that \mathbb{Q} is not nested interval complete since the sequence, derived from the digits of π ,

$$[3, 4] \supset [3.1, 3.2] \supset [3.14, 3.15] \supset [3.141, 3.142] \supset \dots$$

is a nested sequence of closed intervals in the rational numbers whose intersection is empty in \mathbb{Q} .

1.2 Construction of the Real Numbers

Definition 1.9 (The Real Numbers). The set of real numbers, denoted \mathbb{R} , is a totally ordered algebraic field equipped with operations $+$ and \cdot , along with relation \leq . Finally, it must satisfy the completeness axiom. Note that

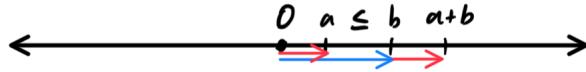
1. Without the axiom of completeness, the set of rational numbers \mathbb{Q} would satisfy the axioms.
2. The fact that the reals are ordered eliminates the complex numbers \mathbb{C} , quaternions \mathbb{H} , and higher-dimensional numbers as candidates.

3. These axioms *uniquely* defines the real numbers up to isomorphism. That is, if two individuals construct sets \mathbb{R}_A and \mathbb{R}_B that satisfy these properties, then

$$\mathbb{R}_A \simeq \mathbb{R}_B$$

For example, let us construct three distinct sets satisfying these axioms:

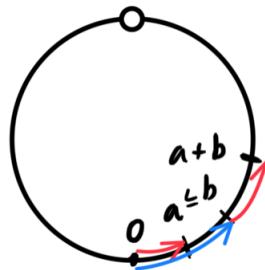
- (a) A line \mathbb{L} with $+$ associated with the translation of \mathbb{L} along itself and \cdot associated with the "stretching/compressing" of the line around the additive origin 0.



- (b) An uncountable list of numbers with possibly infinite decimal points, known as the decimal number system.

$$\dots, -2.583\dots, \dots, 0, \dots, 1.2343\dots, \dots, \sqrt{2}, \dots$$

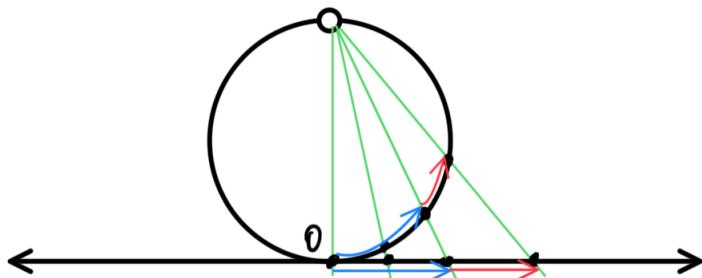
- (c) A circle with a point removed, with addition and multiplication defined similarly as the line.



By the axioms of geometry there exists an isomorphism

$$f : \mathbb{L} \longrightarrow \mathbb{R}$$

between the arbitrary line \mathbb{L} and \mathbb{R} .



4. Additional structures can be put on \mathbb{R} , such as a topology, metric, and norm.

- (a) The basis of the topology of \mathbb{R} consists of open intervals

$$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$$

The δ -neighborhood of a point x is the open interval $(x - \delta, x + \delta)$.

- (b) The metric $d : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}_0^+$ will be defined

$$d(a, b) = |b - a|$$

(c) The norm $\rho : \mathbb{R} \rightarrow \mathbb{R}_0^+$ is defined

$$\rho(x) = |x|$$

Note that a norm induces a metric which induces a topology, so we can define the open interval topology of \mathbb{R} by simply defining ρ .

1.3 Compactness

We now proceed to describe a fundamental topological property of sets that comes up a lot in analysis.

Definition 1.10. A *cover* of a set X is a family of sets X_1, X_2, \dots, X_n such that

$$X = \bigcup_{i=1}^n U_i$$

That is, every point in X must be in at least one of the X_i 's.

Definition 1.11. A subset X of Z is said to be *compact* if each of its open covers has a finite subcover. That is, for every collection C of open subsets of X such that

$$X = \bigcup_{x \in C} x$$

there exists a finite subset $F \subset C$ such that

$$X = \bigcup_{x \in F} x$$

The general notion of compactness for topological spaces is not needed for analysis. Rather, we make use of the following theorem which allows us to focus on the compactness of subsets in Euclidean spaces \mathbb{R}^n .

Theorem 1.5 (Heine-Borel Theorem). *A subset S of Euclidean space \mathbb{R}^n is compact if and only if it is closed and bounded.*

Example 1.2. An open set in \mathbb{R}^2 is not compact. Take the open rectangle $R = (0, 1)^2 \subset \mathbb{R}^2$. There exists an infinite cover of R

$$R = \bigcup_{n=0}^{\infty} (0, 1) \times \left(0, \frac{2^{n+1} - 1}{2^{n+1}}\right)$$

that does not have a finite subcover.

We can view the topological concept of compactness as a generalization of the notion of a Euclidean subset being closed (i.e. containing all its limit points) and bounded (i.e. having all its points lie within a some fixed distance from each other).

According to Terry Tao, a compact set is "small," in the sense that it is easy to deal with. While this may sound counterintuitive at first, since $[0, 1]$ is considered compact while $(0, 1)$, a subset of $[0, 1]$, is considered noncompact. More generally, a set that is compact may be large in area and complicated, but the fact that it is compact means we can interact with it in a finite way using open sets, the building blocks of topology. That finite collection of open sets makes it possible to account for all the points in a set in a finite way. This is easily noticed, since functions defined over compact sets have more controlled behavior than those defined over noncompact sets. Similarly, classifying noncompact spaces are more difficult and less satisfying.

Example 1.3. Any finite space is trivially compact.

We will describe two more equivalent definitions of the completeness of the real numbers.

Definition 1.12. A function $f : \mathbb{N} \rightarrow X$ is called an *infinite sequence*, or a *sequence* of elements of X . It is usually denoted $\{a_n\}$, which is shorthand for

$$a_1, a_2, a_3, \dots$$

A *subsequence* of $\{a_n\}$ is a sequence $\{a_{\gamma_k}\}$, where $\{\gamma_k\}$ is a strictly increasing infinite subset of \mathbb{N} .

Definition 1.13. Let X_1, X_2, X_3, \dots be a sequence of sets. If $X_1 \supset X_2 \supset X_3 \supset \dots$, then the sequence $\{X_n\}$ is *nested*.

Theorem 1.6 (Cantor's Intersection Theorem). *Let $\{I_n\}$ be a nested sequence of intervals $[a_n, b_n]$ in X such that $b_n - a_n \rightarrow 0$ as $n \rightarrow +\infty$. X is complete if there exists exactly one point $c \in X$ such that*

$$c \in \bigcap_{n=1}^{\infty} I_n$$

The set of rational numbers is not complete according to this theorem since the nested sequence leading to the digits of pi

$$[3, 4] \supset [3.1, 3.2] \supset [3.14, 3.15] \supset \dots$$

has an empty intersection in \mathbb{Q} .

Definition 1.14. A point $p \in Z$ is a *limit point* of $X \subset Z$ if every open neighborhood of p has a nontrivial intersection with X . This is equivalent to saying that every open neighborhood of p contains an infinite number of elements of X .

Clearly, the limit point of an open set is its boundary points. Note that a sequence of points can also have a limit point.

Theorem 1.7 (Bolzano-Weierstrass Theorem). *Every bounded infinite sequence in \mathbb{R}^n has an accumulation point. That is, there exists a point $p \in \mathbb{R}^n$ such that every open neighborhood U_p contains an infinite subset of the sequence.*

Proof. The fact that the infinite sequence is bounded means that there exists some closed subset $I \in \mathbb{R}^n$ that contains all point of the sequence. By definition I is compact, so by the Heine-Borel theorem, every cover of I has a finite subcover.

Now, assume that there exists an infinite sequence in I that is not convergent, i.e. has no limit point. Then, each point $x_i \in I$ would have a neighborhood $U(x_i)$ containing at most a finite number of points in the sequence. We can define I such that the union of the neighborhoods is a cover of I . That is,

$$I \subset \bigcup_{i=1}^{\infty} U(x_i)$$

However, since every $U(x_i)$ contains at most a finite number of points, we must have an infinite open neighborhoods to cover $I \implies$ we cannot have a finite subcover. This contradicts the fact that I is compact. ■

1.4 Natural Numbers

Definition 1.15 (Inductive Set, Natural Numbers). A set $X \subset \mathbb{R}$ is inductive if for each number $x \in X$, it also contains $x + 1$. The set of *natural numbers*, denoted \mathbb{N} , is the smallest inductive set containing 1.

We can use this inductive property of natural numbers to prove properties of them.

Lemma 1.8 (Induction Principle). *Given $P(n)$, a property depending on positive integer n ,*

1. if $P(n_0)$ is true for some positive integer n_0 , and
2. if for every $k \geq n_0$, $P(k)$ true implies $P(k + 1)$ true,

then $P(n)$ is true for all $n \geq n_0$.

Lemma 1.9 (Strong Induction Principle). *Given $P(n)$, a property depending on a positive integer n ,*

1. if $P(n_0), P(n_0 + 1), \dots, P(n_0 + m)$ are true for some positive integer n_0 , and nonnegative integer m , and
 2. if for every $k > n_0 + m$, $P(j)$ is true for all $n_0 \leq j \leq k$ implies $P(k)$ is true,
- then $P(n)$ is true for all $n \geq n_0$.*

The idea behind the strong induction principle leads to the proof using infinite descent. Infinite descent combines strong induction with the fact that every subset of the positive integers has a smallest element, i.e. there is no strictly decreasing infinite sequence of positive integers.

Lemma 1.10 (Infinite Descent). *Given $P(n)$, a property depending on positive integer, assume that $P(n)$ is false for a set of integers S . Let the smallest element of S be n_0 . If $P(n_0)$ false implies $P(k)$ false, where $k < n_0$, then by contradiction $P(n)$ is true for all n .*

1.4.1 Countable, Uncountable Sets

Definition 1.16. A set X is *countable* if it is bijective to \mathbb{N} . The cardinality of a countable set with infinite elements is called *countably infinite*. Clearly, all finite sets are countable.

Theorem 1.11 (Induced Countable Sets). *Given a countable sets X, Y .*

1. *A subset $\hat{X} \subset X$ is countable.*
2. *The union $X \cup Y$ is countable.*
3. *The direct product $X \times Y$ is countable.*

Recursively, we can see that any finite union and direct products of countable sets are countable.

Corollary 1.11.1. \mathbb{Z}, \mathbb{Q} are countably infinite.

Proof. It is easy to see that $\mathbb{Z} = \mathbb{N} \cup \{0\} \cup \mathbb{N}$ and $\mathbb{Q} = \mathbb{Z} \times \mathbb{N}$. ■

Theorem 1.12 (Uncountability of \mathbb{R}). *The reals are uncountably infinite.*

$$\text{card } \mathbb{R} > \text{card } \mathbb{N}$$

Therefore, irrational numbers exist.

Proof. Since $(0, 1) \subset \mathbb{R}$, it suffices to prove for $(0, 1)$ uncountable. Assume that $(0, 1)$ is countable. Then, fix a bijection

$$f : \mathbb{N} \longrightarrow (0, 1)$$

Then, for each natural number n we have some decimal sequence that n maps to

$$\begin{aligned} 0 &\mapsto a_{0,0}a_{0,1}a_{0,2}a_{0,3}a_{0,4}\dots \\ 1 &\mapsto a_{1,0}a_{1,1}a_{1,2}a_{1,3}a_{1,4}\dots \\ 2 &\mapsto a_{2,0}a_{2,1}a_{2,2}a_{2,3}a_{2,4}\dots \\ 3 &\mapsto a_{3,0}a_{3,1}a_{3,2}a_{3,3}a_{3,4}\dots \\ 4 &\mapsto a_{4,0}a_{4,1}a_{4,2}a_{4,3}a_{4,4}\dots \\ &\vdots \mapsto \vdots \end{aligned}$$

We will show that this is not a surjection by constructing a number

$$b = 0.b_1b_2b_3b_4\dots$$

that is not in the image of f . We must also realize that decimal numbers do not uniquely represent real numbers since, for example,

$$0.49999\dots = 0.50000\dots$$

For each b_i ($i = 0, 1, 2, \dots$), we let b_i be any number such that

$$b_i \neq a_{i,i}$$

and that $b_i \neq 9$ (this is to ensure that every b_i past a certain point equals 9). This ensures that

1. $b \neq f(0)$ since $b_0 \neq a_{0,0}$
2. $b \neq f(1)$ since $b_1 \neq a_{1,1}$
3. $b \neq f(2)$ since $b_2 \neq a_{2,2}$
4. ...

and therefore $b \notin \text{Im } f$. ■

2 Limits of Sequences

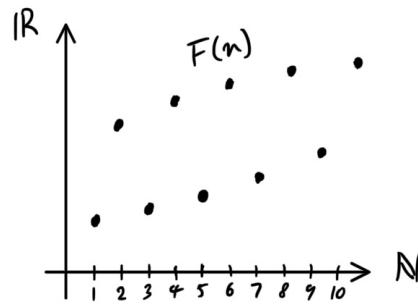
2.1 Sequences, Basic Properties

Definition 2.1 (Sequence). A function $f : \mathbb{N} \rightarrow X$ is a *sequence*, denoted $\{a_n\} = a_1, a_2, a_3, \dots$. Even though we can just let X be ordered, we simplify and assume that $X = \mathbb{R}$ from now on. Let A be some real number.

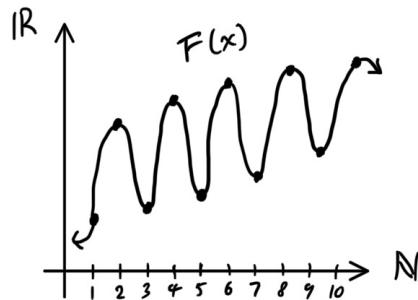
1. $\{x_n\}$ is a *constant sequence* if $a_i = A$ for all i
2. $\{x_n\}$ is an *ultimately constant sequence* if $a_i = A$ for all $i > N$ for some $N \in \mathbb{N}$. If $A = 0$, then $\{x_n\}$ is *finary*.
3. $\{x_n\}$ is *bounded* if there exists M such that $|x_n| < M$ for all $n \in \mathbb{N}$

A *subsequence* of $\{a_n\}$ is a sequence $\{a_{\gamma_k}\}$, where $\{\gamma_k\}$ is a strictly increasing infinite subset of \mathbb{N} .

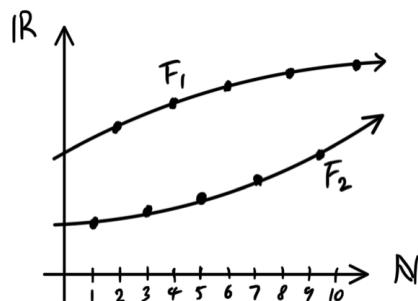
We can visualize a sequence $f : \mathbb{N} \rightarrow X$ as the set of points in $\mathbb{N} \times X$. If, $X = \mathbb{R}$, then the sequence is easy to see in $\mathbb{N} \times \mathbb{R}$ embedded in \mathbb{R}^2 .



Sometimes, it can be useful to visualize the extension of $f : \mathbb{N} \rightarrow \mathbb{R}$, which we will denote as $F : \mathbb{R} \rightarrow \mathbb{R}$.



Likewise, subsequences of f and their extensions can be visualized as F_1 and F_2 .



Definition 2.2 (Sequence Space). The set of all real-valued sequences, denoted $\mathbb{R}^{\mathbb{N}}$, is an infinite dimensional vector space over \mathbb{R} , where addition and scalar multiplication of sequences are defined component-wise.

$$\begin{aligned}\{x_n\} + \{y_n\} &= \{x_n + y_n\} \\ c\{x_n\} &= \{cx_n\}, c \in \mathbb{R}\end{aligned}$$

$X^{\mathbb{N}}$ is also equivalent to the function space of elements $f : \mathbb{N} \rightarrow \mathbb{R}$. We can equip this space with additional operations. The product and quotient of sequences is defined component-wise. That is, given two numerical sequences $\{x_n\}$ and $\{y_n\}$ over \mathbb{R} ,

1. $\{x_n\} \cdot \{y_n\} = \{(x_n \cdot y_n)\}$
2. $\{x_n\}/\{y_n\} = \left\{ \left(\frac{x_n}{y_n} \right) \right\}$ which, of course, is defined only when $y_n \neq 0$ for all $n \in \mathbb{N}$.

Definition 2.3 (Limit of a Sequence). A number $A \in \mathbb{R}$ is called the *limit of the sequence* $\{x_n\}$, written

$$\lim_{n \rightarrow \infty} x_n = A,$$

if for every neighborhood U_A there exists an index N such that

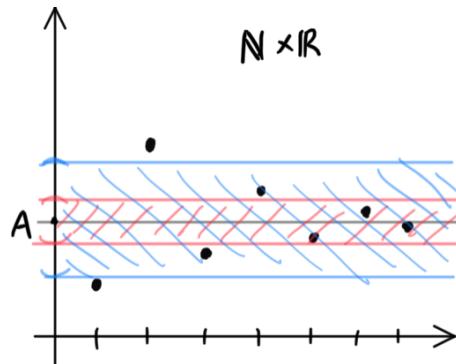
$$x_n \in U_A \text{ for all } n > N$$

Equivalently, A is the limit of $\{x_n\}$ if for every $\epsilon > 0$, there exists an index N such that

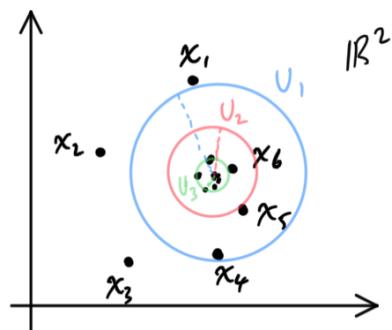
$$|x_n - A| < \epsilon \text{ for all } n > N$$

If A is the limit of $\{x_n\}$, then we say that $\{x_n\}$ *converges* to A . If the limit of $\{x_n\}$ is not well defined or finite, then we say that $\{x_n\}$ is *divergent*.

There are two ways we can visualize the limit of a sequence. The first way is to visualize the sequence on $\mathbb{N} \times \mathbb{R}$ and imagine the various open neighborhoods of the limit A getting smaller and smaller, albeit containing an infinite number of elements of $\{x_n\}$. For example, when $X = \mathbb{R}$, we have



The other way to visualize it is to just imagine the codomain space X and the sequence as a collection of ordered points in X . We can then imagine an open neighborhood of limit A getting smaller and smaller, albeit containing an infinite number of elements of $\{x_n\}$. When $X = \mathbb{R}^2$, we have



Theorem 2.1 (Properties of Limits). *Given that $\{x_n\}, \{y_n\}$ are numerical sequences with $y_n \neq 0$ for all n , and let*

$$\lim_{n \rightarrow \infty} x_n = A, \quad \lim_{n \rightarrow \infty} y_n = B \neq 0$$

then,

$$\begin{aligned}\lim_{n \rightarrow \infty} (x_n + y_n) &= A + B \\ \lim_{n \rightarrow \infty} (cx_n) &= cA \\ \lim_{n \rightarrow \infty} (x_n \cdot y_n) &= A \cdot B \\ \lim_{n \rightarrow \infty} \frac{x_n}{y_n} &= \frac{A}{B}\end{aligned}$$

It immediately follows that the set of all convergent sequences in \mathbb{R}^N is a subspace of \mathbb{R}^N .

Proof. Assume that

$$\lim_{n \rightarrow \infty} x_n = A \text{ and } \lim_{n \rightarrow \infty} y_n = B \neq 0$$

This means that for every $\epsilon > 0$, there exists $N_1, N_2 \in \mathbb{N}$ such that

$$\begin{aligned}|x_n - A| &< \epsilon \text{ for all } n > N_1 \\ |y_n - B| &< \epsilon \text{ for all } n > N_2\end{aligned}$$

Therefore, for a given ϵ , we wish to prove that there exists a N such that for all $n > N$,

1. $|x_n + y_n - (A + B)| < \epsilon$
2. $|cx_n - cA| < \epsilon$
3. $|x_n y_n - AB| < \epsilon$
4. $\left| \frac{x_n}{y_n} - \frac{A}{B} \right| < \epsilon$

1. By the triangle inequality, we can see that

$$|(x_n + y_n) - (A + B)| = |x_n - A| + |y_n - B|$$

Since we can choose the error between x_n and A for $n > N_1$, and y_n and B for $n > N_2$ as small as we want, we set it to $\epsilon/2$. Then, we have

$$|(x_n + y_n) - (A + B)| = |x_n - A| + |y_n - B| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $n > N = \max\{N_1, N_2\}$. Therefore, for a given ϵ , there exists an N such that

$$|(x_n + y_n) - (A + B)| < \epsilon \text{ for all } n > N$$

2. This proof is easy. For a given ϵ , we choose the error to be $\frac{\epsilon}{c}$.

$$|x_n - A| < \frac{\epsilon}{c} \text{ for all } n > N_1$$

Then, there exists natural number N_1 such that

$$|cx_n - cA| < c|x_n - A| = c \frac{\epsilon}{c} = \epsilon \text{ for all } n > N_1$$

3. We first observe that since the limit of $\{y_n\}$ exists, it must be bounded by a value, say B . That is,

$$|y_n| < Y \text{ for all } n \in \mathbb{N}$$

Then, we see that

$$\begin{aligned}|x_n y_n - AB| &= |(x_n y_n - A y_n) + (A y_n - AB)| \\ &< |x_n y_n - A y_n| + |A y_n - AB| \\ &= |y_n| |x_n - A| + |A| |y_n - B|\end{aligned}$$

Suppose $\epsilon > 0$ is given. Then, we can set the error bounds freely; there exists $N_1, N_2 \in \mathbb{N}$ such that

$$\begin{aligned}|x_n - A| &< \frac{\epsilon}{2Y} \text{ for all } n > N_1 \\ |y_n - B| &< \frac{\epsilon}{2|A|} \text{ for all } n > N_2\end{aligned}$$

Then, we can see that

$$|x_n y_n - AB| \leq |y_n||x_n - A| + |A||y_n - B| < Y \cdot \frac{\epsilon}{2Y} + |A| \frac{\epsilon}{2|A|} = \epsilon$$

for all $n > N = \max\{N_1, N_2\}$.

4. We use the estimate

$$\left| \frac{A}{B} - \frac{x_n}{y_n} \right| = \frac{|x_n||y_n - B| + |y_n||x_n - A|}{y_n^2} \cdot \frac{1}{1 - \delta(y_n)}, \quad \delta(y_n) = \frac{|y_n - B|}{|y_n|}$$

For a given $\epsilon > 0$, we find natural numbers N_1, N_2 such that

$$\begin{aligned}|x_n - A| &< \min \left\{ 1, \frac{\epsilon|B|}{8} \right\} \text{ for all } n > N_1 \\ |y_n - B| &< \min \left\{ \frac{|B|}{4}, \frac{\epsilon B^2}{16(|A| + 1)} \right\} \text{ for all } n > N_2\end{aligned}$$

From this we can deduce that

$$|x_n| = |x_n - A + A| < |x_n - A| + |A| < |A| + 1$$

and

$$\begin{aligned}|B| &= |y_n + B - y_n| < |y_n| + |B - y_n| \\ \implies |y_n| &> |B| - |y_n - B| > |B| - \frac{|B|}{4} > \frac{|B|}{2} \\ \implies \frac{1}{|y_n|} &< \frac{2}{|B|} \\ \implies 0 < \delta(y_n) &= \frac{|y_n - B|}{|y_n|} < \frac{|B|/4}{|B|/2} = \frac{1}{2} \\ \implies 1 - \delta(y_n) &> \frac{1}{2} \\ \implies 0 < \frac{1}{1 - \delta(y_n)} &< 2\end{aligned}$$

So, we can substitute

$$\begin{aligned}|x_n| \cdot \frac{1}{y_n^2} \cdot |y_n - B| &< (|A| + 1) \cdot \frac{4}{B^2} \cdot \frac{\epsilon \cdot B^2}{16(|A| + 1)} = \frac{\epsilon}{4} \\ \left| \frac{1}{y_n} \right| \cdot |x_n - A| &< \frac{2}{|B|} \cdot \frac{\epsilon|B|}{8} = \frac{\epsilon}{4}\end{aligned}$$

into the final equation to get

$$\left| \frac{A}{B} - \frac{x_n}{y_n} \right| < \epsilon \text{ for all } n > N = \max\{N_1, N_2\}$$

■

Interestingly, we can interpret the limit as a mapping itself

$$\lim : \mathbb{R}_C^{\mathbb{N}} \longrightarrow \mathbb{R}$$

The properties of the limit imply that

1. \lim is a linear mapping between vector spaces, an element of $\text{Hom}(\mathbb{R}_C^{\mathbb{N}}, \mathbb{R})$
2. \lim is a multiplicative group homomorphism from $\mathbb{R}_C^{\mathbb{N}}$ to \mathbb{R}

Theorem 2.2. Given convergent sequences $\{x_n\}$ and $\{y_n\}$, if

$$\lim_{n \rightarrow \infty} x_n < \lim_{n \rightarrow \infty} y_n$$

then there exists an index $N \in \mathbb{N}$ such that $x_n < y_n$ for all $n > N$.

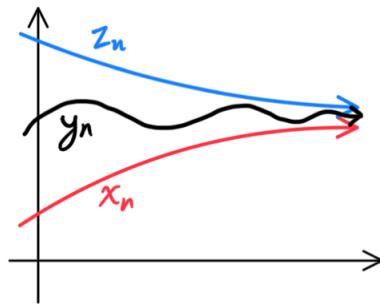
Theorem 2.3 (Squeeze Theorem for Sequences). Given sequences $\{x_n\}, \{y_n\}, \{z_n\}$ such that

$$x_n \leq y_n \leq z_n$$

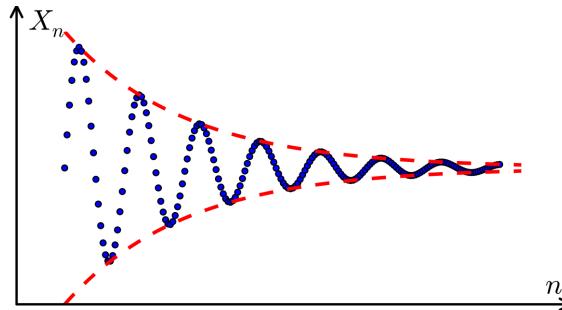
for all $n > N$, if $\{x_n\}$ and $\{z_n\}$ both converge to the same limit, then the sequence $\{y_n\}$ also converges to that limit. That is,

$$\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} z_n = A \implies \lim_{n \rightarrow \infty} y_n = A$$

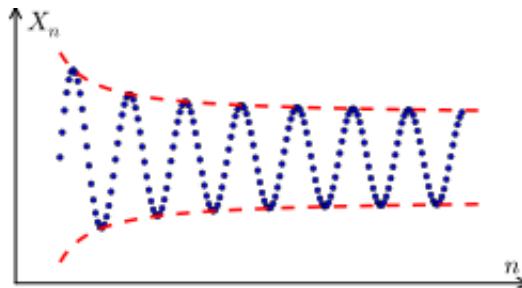
This is quite easy to visualize. For convenience, we use the extension of the sequences $\{x_n\}, \{y_n\}, \{z_n\}$ onto the real numbers.



Definition 2.4 (Cauchy Sequence). A sequence $\{x_n\}$ is called a *Cauchy sequence* if for any $\epsilon > 0$ there exists an index $N \in \mathbb{N}$ such that $|x_m - x_n| < \epsilon$ whenever $n > N$ and $m > N$. That is, for any arbitrarily small interval of length ϵ , there will always be an infinite number of elements of $\{x_n\}$ that all exist within that interval. The plot below shows a Cauchy sequence.



but the sequence below is not Cauchy since the elements of the sequence do not get arbitrarily close to each other as the sequence progresses.



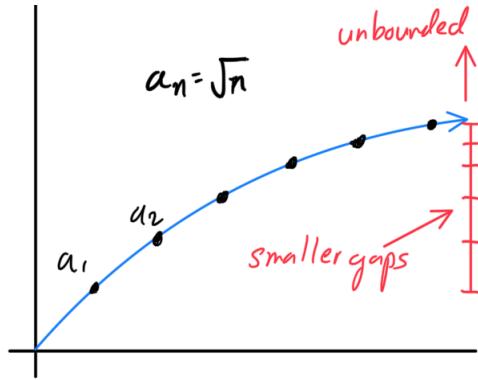
Note that it is not sufficient to say that a sequence is Cauchy by claiming that each term becomes arbitrarily close to the preceding term. That is,

$$\lim_{n \rightarrow \infty} |x_{n+1} - x_n| = 0$$

For example, look at the sequence

$$a_n = \sqrt{n} \implies a_{n+1} - a_n = \frac{1}{\sqrt{n+1} + \sqrt{n}} < \frac{1}{2\sqrt{n}}$$

However, it is clear that a_n gets arbitrarily large, meaning that a finite interval can contain at most a finite number of terms in $\{a_n\}$.



Upon closer observation, the definition of a Cauchy sequence is really just the same as a sequence having a limit.

Theorem 2.4 (Cauchy Convergence Criterion). *A real-valued sequence converges if and only if it is a Cauchy sequence.*

2.1.1 Divergent Sequences

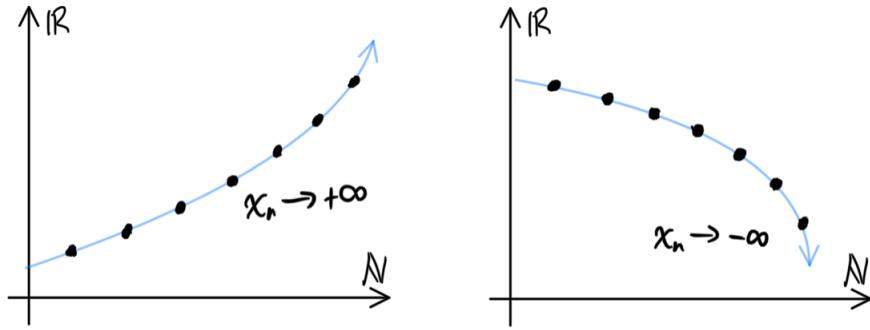
Note that while a convergent sequence can be visualized quite easily by the Cauchy convergence criterion, there are many ways in which a sequence can be divergent.

1. Increasing/decreasing indefinitely
2. Oscillating between two constant values
3. Oscillating between a value tending to $+\infty$ and a value tending to $-\infty$
4. Many other classes of divergence

Definition 2.5 (Sequence Tending to Infinity). The sequence $\{x_n\}$ tends to positive infinity if for each number c there exists $N \in \mathbb{N}$ such that $x_n > c$ for all $n > N$. It is denoted

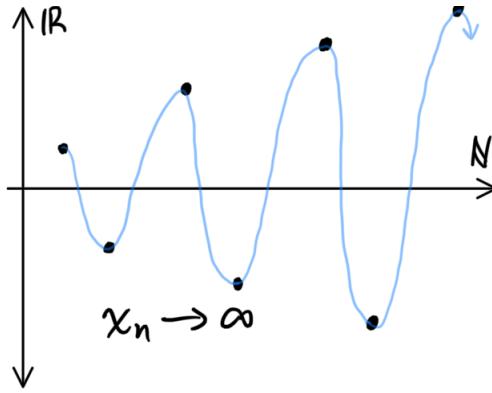
$$x_n \rightarrow +\infty \text{ or } \lim_{n \rightarrow \infty} x_n = +\infty$$

We define sequences that tend to negative infinity similarly.



And $\{x_n\}$ tends to infinity if for each c there exists $N \in \mathbb{N}$ such that $|x_n| > c$ for all $n > N$, which is written

$$x_n \rightarrow \infty$$

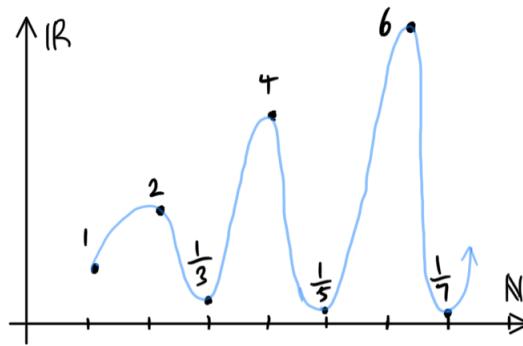


Note that

$$x_n \rightarrow +\infty \text{ or } x_n \rightarrow -\infty \implies x_n \rightarrow \infty$$

but the converse is not necessarily true. The simple example is the sequence $x_n = (-1)^n n$. Also, it is important to know that a sequence may be unbounded and yet not tend to $+\infty$, $-\infty$, or ∞ .

Example 2.1 (Unbounded Sequence that Doesn't tend to ∞). The sequence $x_n = n^{(-1)^n}$ is divergent yet does not tend to positive infinity, negative infinity, nor infinity.



2.1.2 Monotonic Sequences

Definition 2.6 (Monotonic Sequences). A sequence $\{x_n\}$ is *increasing* if $x_{n+1} > x_n$ for all n . Similarly, it is *nondecreasing* if $x_{n+1} \geq x_n$, *decreasing* if $x_{n+1} < x_n$, and *nonincreasing* if $x_{n+1} \leq x_n$. Sequences of these types are called *monotonic*.

Lemma 2.5 (Convergence Criterion for Monotonic Sequences). *In order for a nondecreasing (nonincreasing) sequence to be convergent, it is necessary and sufficient that it is bounded above (or below).*

Theorem 2.6 (Bolzano-Weierstrass Theorem). *Every bounded sequence in \mathbb{R}^n contains a convergent subsequence.*

Proof. It suffices to prove that there exists a monotonic sequence within a bounded sequence $\{x_n\}$. ■

Corollary 2.6.1. *From each sequence of real numbers there exists either a convergent subsequence or a subsequence tending to infinity.*

This allows us to measure the convergence or divergence of subsequences inside a more complicated sequence. We further define additional tools to do this.

Example 2.2. *We claim that*

$$\lim_{n \rightarrow \infty} \frac{n}{q^n} = 0 \text{ if } q > 1$$

Proof. Since $x_n = \frac{n}{q^n} \implies x_{n+1} = \frac{n+1}{q^{n+1}}x_n$ for $n \in \mathbb{N}$. Since

$$\lim_{n \rightarrow \infty} \frac{n+1}{nq} = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) \frac{1}{q} = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right) \cdot \lim_{n \rightarrow \infty} \frac{1}{q} = 1 \cdot \frac{1}{q} = \frac{1}{q} < 1$$

there exists an index N such that $\frac{n+1}{nq} < 1$ for $n > N$. Thus, we have

$$x_n > x_{n+1} = x_n \cdot \frac{n+1}{nq} \text{ for } n > N$$

which means that the sequence will be monotonically decreasing from index N on. The terms of the sequence

$$x_{N+1} > x_{N+2} > x_{N+3} > \dots$$

are positive (bounded below) and are monotonically decreasing, so it must have a limit.

Finding the actual limit is easy. Let $x = \lim_{n \rightarrow \infty} x_n$. It follows from the relation $x_{n+1} = \frac{n+1}{nq}x_n$ that

$$x = \lim_{n \rightarrow \infty} (x_{n+1}) = \lim_{n \rightarrow \infty} \left(\frac{n+1}{nq} x_n \right) = \lim_{n \rightarrow \infty} \frac{n+1}{nq} \cdot \lim_{n \rightarrow \infty} x_n = \frac{1}{q}x$$

which implies that $(1 - \frac{1}{q}) = 0 \implies x = 0$. ■

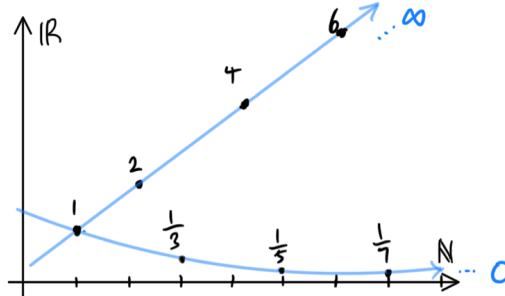
Example 2.3. *We claim that*

$$\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$$

2.1.3 The Number e

2.1.4 Partial, Inferior, Superior Limits

Definition 2.7 (Partial Limits). The *partial limit* of a sequence $\{x_n\}$ is the limit of any of its subsequences.

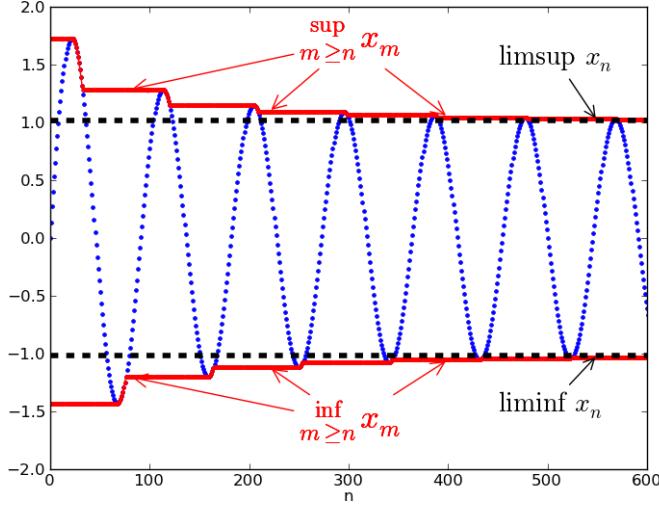


Two (out of the many) partial limits of the sequence above is $+\infty$ and 0.

Definition 2.8 (Inferior, Superior Limits). The *inferior limit* and *superior limit* of a sequence $\{x_k\}$ are defined as follows, and they can be shown to be the smallest and largest partial limits of the sequence. That is,

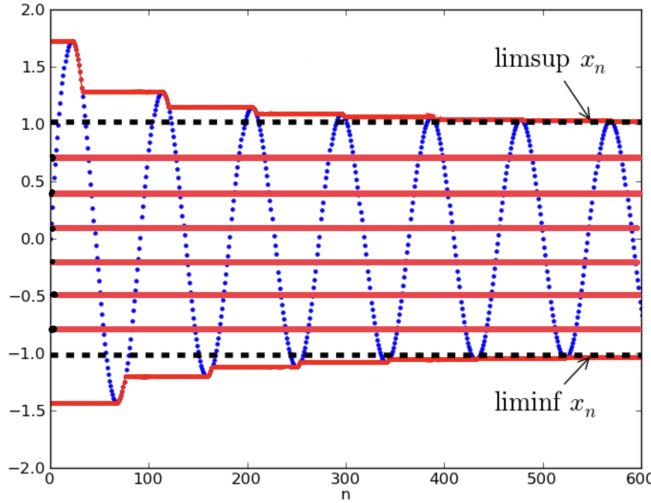
$$\begin{aligned}\liminf_{k \rightarrow \infty} x_k &\equiv \liminf_{n \rightarrow \infty} \inf_{k \geq n} x_k = \min\{\liminf_{r \rightarrow \infty} y_r\} \\ \limsup_{k \rightarrow \infty} x_k &\equiv \limsup_{n \rightarrow \infty} \sup_{k \geq n} x_k = \max\{\limsup_{r \rightarrow \infty} y_r\}\end{aligned}$$

where $\{y_r\}$ is any subsequence of $\{x_n\}$. Despite the definition, it isn't too difficult to visualize this. For example, take a look at the superior and inferior limits of the divergent sequence below.



In order to find the superior limit, we first look the whole sequence in \mathbb{N} and find the supremum. We now "decrease" our domain from \mathbb{N} to $\{2, 3, \dots\}$, then $\{3, 4, \dots\}$, then $\{4, 5, \dots\}$ and so on, continuing to label the supremum of the sequence. The limit of this sequence of supremums is the superior limit. Informally, the superior limit tells us what the supremum of the "end terms" of $\{x_n\}$ will be, and similarly for the inferior limit.

The second property of superior and inferior limits is that they represent the greatest and least possible partial limit of a sequence. For example, the six red lines marked in the middle (along with infinitely many others) are viable partial limits because one can choose a subsequence such that all of its points after a certain n lie in some ϵ -neighborhood of the limit.



Therefore, the superior and inferior limits represent some sort of "bound" on the sequence in the long run. That is, on the long run, the terms of the sequence $\{x_n\}$ cannot be greater than its superior limit

and cannot be less than its inferior limit. With this interpretation, the following theorem should be clear.

Theorem 2.7. *A sequence has a limit or tends to $\pm\infty$ if and only if its inferior and superior limits are the same.*

Corollary 2.7.1. *A sequence converges if and only if every subsequence of it converges.*

2.2 Real Series

Defining limits and convergence for series can be painstaking... unless we construct series as sequences.

Definition 2.9 (Series over \mathbb{R}). Given a sequence of real numbers $\{a_n\}$, the *series* of $\{a_n\}$ is defined

$$s = \sum_{k=1}^{\infty} a_k$$

The series can be interpreted as the sequence of partial sums $\{s_n\}$, where

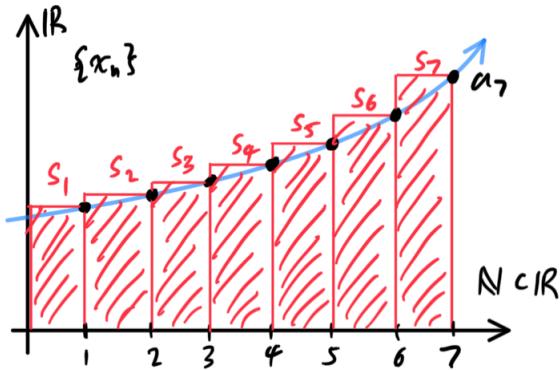
$$s_n = \sum_{k=1}^n a_k$$

is the *n*th partial term of the series. Therefore, we can interpret the sum of the series s as the limit of $\{s_n\}$.

$$\lim_{n \rightarrow \infty} s_n = s$$

If the sequence $\{s_n\}$ converges, the series is *convergent* and *divergent* otherwise.

It can be visualized as the Riemann sums of the smooth extension of the function representing $\{a_n\}$, as shown below, where the a_n 's represent the height of each bar and the s_n 's represent the sums of the area of each bar.



Note that we are really just defining a series as an ordered pair $(\{a_n\}, \{s_n\})$ of sequences connected by the relation

$$s_n = \sum_{k=1}^n a_k \text{ for all } n \in \mathbb{N}$$

Since the convergence of a series is equivalent to convergence of its sequence of partial sums, applying the Cauchy convergence criterion to the sequence $\{s_n\}$ leads to the following theorem.

Theorem 2.8 (Cauchy Convergence Criterion for Series). *The series $a_1 + \dots + a_n + \dots$ converges if and only if for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $m \geq n > N$,*

$$|a_n + \dots + a_m| < \epsilon$$

Definition 2.10 (Cauchy Product of Real Series).

Corollary 2.8.1 (nth Term Test). *A necessary (but not sufficient) condition for convergence of the series $a_1 + \dots + a_n + \dots$ is that the terms tend to 0 as $n \rightarrow \infty$. That is, it is necessary that*

$$\lim_{n \rightarrow \infty} a_n = 0$$

Proof. It suffices to set $m = n$ in the Cauchy convergence criterion. This would mean that for every $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that

$$|a_n| = |a_n - 0| < \epsilon \text{ for all } n > N$$

which, by definition, means that $\{a_n\}$ converges to 0. ■

Example 2.4 (Geometric Series). *The series*

$$1 + q + q^2 + \dots + q^n + \dots$$

is called the geometric series.

Since $|q^n| = |q|^n$, we have $|q^n| \geq 1$ when $|q| \geq 1$. So, if $|q| \geq 1$, the terms q^n does not converge to 0 and the Cauchy convergence criterion is not met.

Now, suppose $|q| < 1$. Then,

$$s_n = 1 + q + \dots + q^{n-1} = \frac{1 - q^n}{1 - q}$$

which implies that

$$\lim_{n \rightarrow \infty} s_n = \frac{1}{1 - q}$$

since $\lim_{n \rightarrow \infty} q^n = 0$ if $|q| < 1$. This, the series converges to if and only if $|q| < 1$, and its sum is $\frac{1}{1-q}$.

Example 2.5 (Harmonic Series). *The series*

$$1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} + \dots$$

is called the harmonic series, since each term from the second on is the harmonic mean of the two terms on either side of it. Clearly,

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

but the sequence of partial sums s_n diverges, and thus the harmonic series diverges.

2.2.1 Convergence Tests

Definition 2.11 (Absolute Convergence). *The series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent if the series*

$$\sum_{n=1}^{\infty} |a_n|$$

converges. Clearly, every absolutely convergent series because

$$\left| \sum_{n=1}^{\infty} a_n \right| \leq \sum_{n=1}^{\infty} |a_n|$$

Theorem 2.9 (Direct Comparison Test). *Let $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ be 2 series with nonnegative terms. If there exists an index $N \in \mathbb{N}$ such that $a_n \leq b_n$ for all $n > N$, then*

$$\begin{aligned} \sum_{n=1}^{\infty} b_n \text{ convergent} &\implies \sum_{n=1}^{\infty} a_n \text{ convergent} \\ \sum_{n=1}^{\infty} a_n \text{ divergent} &\implies \sum_{n=1}^{\infty} b_n \text{ divergent} \end{aligned}$$

Theorem 2.10 (Limit Comparison Test). Suppose the limit

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \alpha$$

exists for the series $\sum_{n=1}^{\infty} a_n$. Then,

$$\begin{aligned}\alpha < 1 &\implies \sum_{n=1}^{\infty} a_n \text{ converges absolutely} \\ \alpha > 1 &\implies \sum_{n=1}^{\infty} a_n \text{ diverges} \\ \alpha = 1 &\implies \text{Inconclusive}\end{aligned}$$

Theorem 2.11 (Root Test). Let $\sum_{n=1}^{\infty}$ be a given series and

$$\alpha = \limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}$$

Then, the following are true

$$\begin{aligned}\alpha < 1 &\implies \sum_{n=1}^{\infty} a_n \text{ converges absolutely} \\ \alpha > 1 &\implies \sum_{n=1}^{\infty} a_n \text{ diverges} \\ \alpha = 1 &\implies \text{Inconclusive}\end{aligned}$$

Theorem 2.12 (Weierstrass M-test for Absolute Convergence). Let $\sum_{n=1}^{\infty}$ and $\sum_{n=1}^{\infty} b_n$ be series. Suppose there exists an index $N \in \mathbb{N}$ such that $|a_n| \leq b_n$ for all $n > N$. Then,

$$\sum_{n=1}^{\infty} b_n \text{ converges} \implies \sum_{n=1}^{\infty} a_n \text{ converges absolutely}$$

The following theorem, while obvious, has interesting consequences.

Theorem 2.13 (Cauchy). If $a_1 \geq a_2 \geq \dots \geq 0$, the series $\sum_{n=1}^{\infty} a_n$ converges if and only if the series

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \dots$$

converges.

Proof. Letting $A_k = a_1 + a_2 + \dots + a_k$ and $S_n = a_1 + 2a_2 + \dots + 2^n a_{2^n}$, it is clear that by adding up the inequalities

$$\begin{aligned}a_2 &\leq a_2 \leq a_1 \\ 2a_4 &\leq a_3 + a_4 \leq 2a_2 \\ 4a_8 &\leq a_5 + a_6 + a_7 + a_8 \leq 4a_4 \\ &\dots \\ 2^n a_{2^{n+1}} &\leq a_{2^n+1} + \dots + a_{2^{n+1}} \leq 2^n a_{2^n},\end{aligned}$$

we get

$$\frac{1}{2}(S_{n+1} - a_1) \leq A_{2^{n+1}} - a_1 \leq S_n$$

Since the sequences $\{A_k\}$ and $\{S_k\}$ are nondecreasing, and hence from the inequalities we can conclude that they are either both bounded above (which means that they are both convergent since it is a bounded, nondecreasing series) or both unbounded above (which means that they are both divergent since they are nondecreasing and unbounded). ■

Corollary 2.13.1 (p-series Test). *The series*

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

converges for $p > 1$ and diverges for $p \leq 1$.

Proof. Suppose $p \geq 0$. By the previous theorem, the series converges or diverges simultaneously with the series

$$\sum_{k=0}^{\infty} 2^k \frac{1}{(2^k)^p} = \sum_{k=0}^{\infty} (2^{1-p})^k$$

which is really just a geometric series. A necessary and sufficient condition for the convergence of this series is that $2^{1-p} < 1$, that is, $p > 1$.

Now suppose $p \leq 0$. The series is then clearly divergent since all of the terms are larger than 1. ■

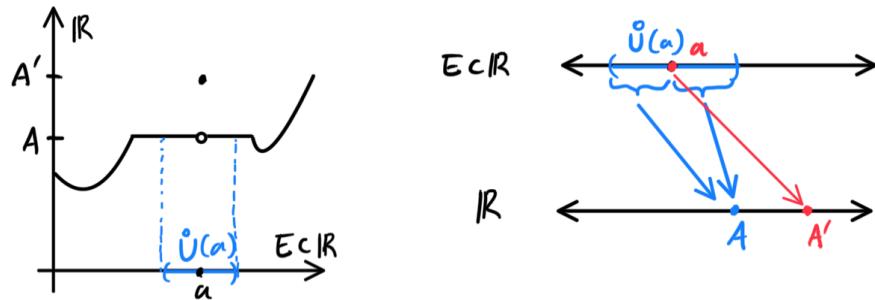
2.2.2 Representation of Euler's Number as a Series

3 Limits of Functions

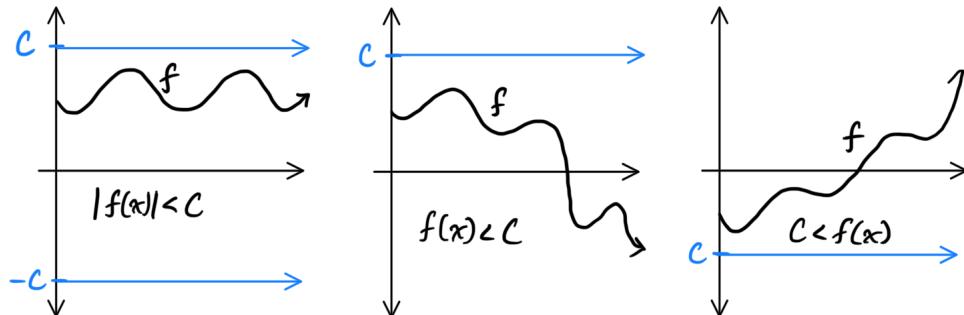
Even though we can generalize the concept of limits to functions mapping between arbitrary topological spaces $f : X \rightarrow Y$, we define this for functions $f : E \subset \mathbb{R} \rightarrow \mathbb{R}$.

Definition 3.1 (Functions). Given a real-valued function $f : E \rightarrow \mathbb{R}$ defined on domain $E \subset \mathbb{R}$,

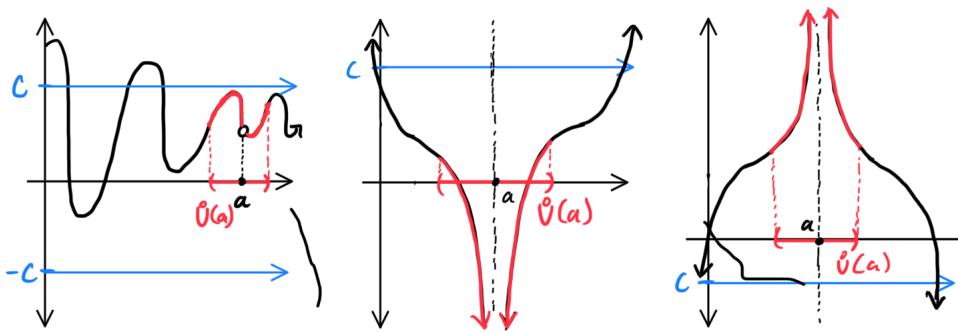
1. f is a *constant function* if $f(x) = A$ for all $x \in E$
2. f is called *ultimately constant* as $x \rightarrow a$ if it is constant in some deleted neighborhood $\dot{U}(a)$, where a is a limit point of E .



3. f is *bounded*, *bounded above*, or *bounded below* respectively if there is a number $C \in \mathbb{R}$ such that $|f(x)| < C$, $f(x) < C$, or $C < f(x)$ for all $x \in E$.



4. f is *ultimately bounded*, *ultimately bounded above*, or *ultimately bounded below* as $x \rightarrow a$ if it is bounded, bounded above, or bounded below in some deleted neighborhood $\dot{U}_E(a)$.



Example 3.1. The function

$$f(x) = \sin \frac{1}{x} + x \cos \frac{1}{x}$$

for $x \neq 0$ is not bounded on the domain of definition, but it is ultimately bounded as $x \rightarrow 0$.

Definition 3.2 ($\epsilon - \delta$ Definition of a Limit). The function $f : E \subset \mathbb{R} \rightarrow \mathbb{R}$ tends to $A \in \mathbb{R}$ as x tends to a , or that

$$\lim_{x \rightarrow a} f(x) = A$$

if for every $\epsilon > 0$ there exists $\delta > 0$ such that

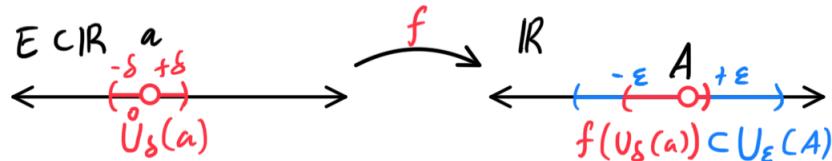
$$0 < |x - a| < \delta \implies |f(x) - A| < \epsilon$$

Note that we set the $0 < |x - a|$ to ensure that $x \neq a$.

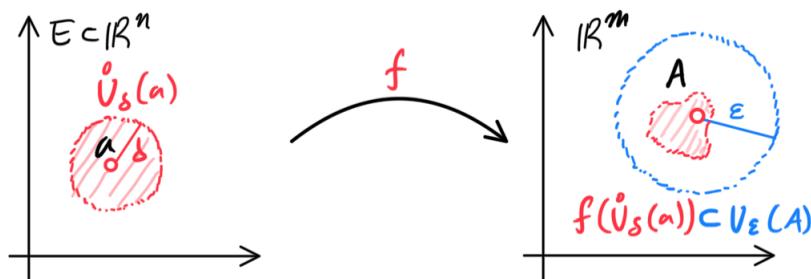
Therefore, in other words, for any arbitrarily small $\epsilon > 0$, if we can find a $\delta > 0$ such that the image of the deleted δ -neighborhood of a (defined to be $\tilde{U}_\delta(a) \equiv U_\delta(a) \setminus a$) is completely within the ϵ -neighborhood $U_\epsilon(A)$, then

$$\lim_{x \rightarrow a} f(x) = A$$

Visually,



In higher dimensional spaces, we have



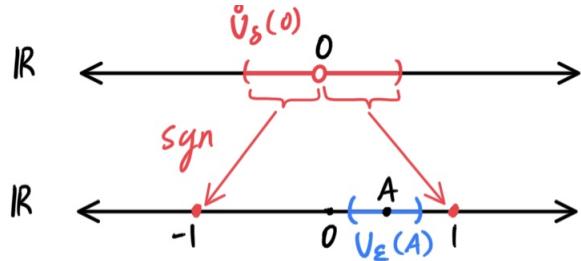
Example 3.2 (Limit of the Signum Function). The function $\text{sgn}: \mathbb{R} \rightarrow \mathbb{R}$ defined

$$\text{sgn } x = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

has no limit as $x \rightarrow 0$.

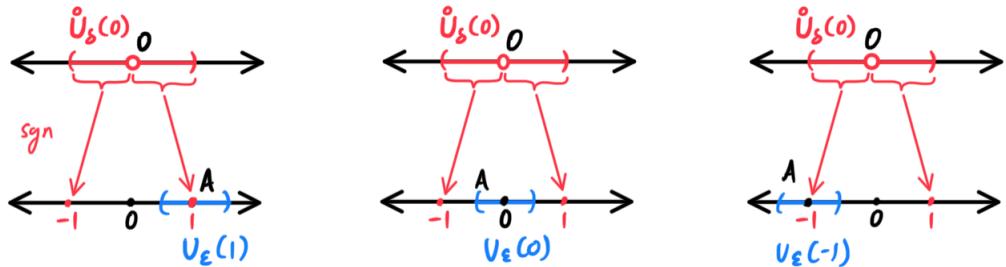
First, it is ludicrous that the limit would be any number that is not $\{-1, 0, 1\}$. If we assume that $A \notin \{-1, 0, 1\}$, then we can choose any arbitrarily small ϵ -neighborhood of A that does not include the three numbers. Clearly, there doesn't exist any $\delta > 0$ such that the deleted δ -neighborhood of 0 maps to a set completely contained in the ϵ -neighborhood of A . That is,

$$\operatorname{sgn}(\dot{U}_\delta(0)) = \{-1, 1\} \not\subset U_\epsilon(A)$$



It doesn't even intersect the ϵ -neighborhood at all.

1. If $A = 1$, we can construct a ϵ -neighborhood V_A for $\epsilon = \frac{1}{2}$. Clearly, there exists no open neighborhood U_0 of 0 that is entirely mapped to V , since U_0 contains both negative numbers and 0 and hence must be mapped to 0, -1.
2. Similarly, given the ($\epsilon = \frac{1}{2}$)-neighborhood of $A = -1$, there exists no open neighborhood U_0 of 0 that is entirely mapped to it, since U_0 contains both positive numbers and 0 and hence must be mapped to 0, 1.
3. Finally, given the ($\epsilon = \frac{1}{2}$)-neighborhood of $A = 0$, there exists no open neighborhood U_0 of 0 that is entirely mapped to it, since U_0 contains both positive and negative numbers and hence must be mapped to ± 1 .

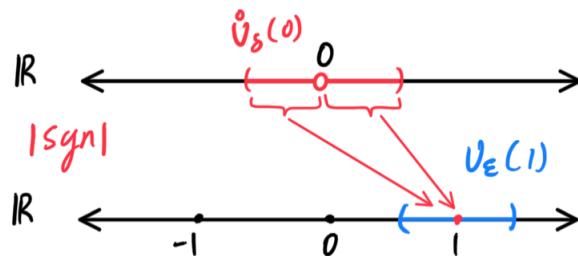


Therefore, the limit does not exist.

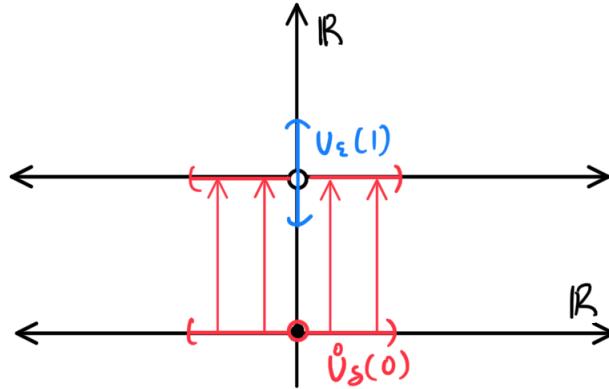
Example 3.3 (Limit of Absolute Value of Signum Function). We will show that

$$\lim_{x \rightarrow 0} |\operatorname{sgn} x| = 1$$

We construct a ϵ -neighborhood $U_\epsilon(1)$ around 1. Given this neighborhood, we can imagine choosing the deleted δ -neighborhood $\dot{U}_\delta(0)$ around 0. Since every element in $\dot{U}_\delta(0)$ maps to 1, it is clearly in U_ϵ .



In fact, for arbitrarily small $\epsilon > 0$, we can choose **any** $\delta > 0$ since everything in $\mathbb{R} \setminus 0$ maps to 1. We can visualize this in \mathbb{R}^2 as

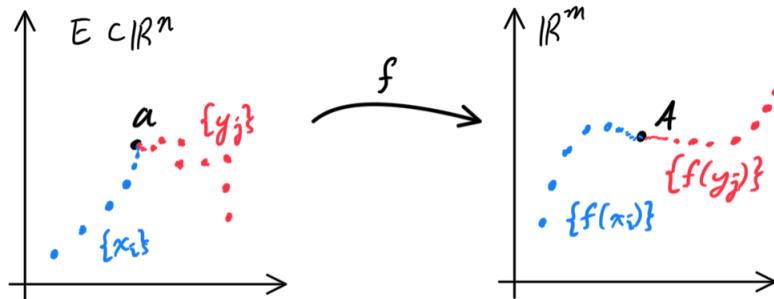


The following lemma nicely interweaves the concepts of limits of sequences and limits of functions. It can be nice for visualization to define the limit of a function using Cauchy sequences rather than the usual $\epsilon - \delta$ definition.

Lemma 3.1 (Cauchy Sequence Criterion of a Limit). *The relation*

$$\lim_{x \rightarrow a} f(x) = A$$

holds if and only if for every sequence $\{x_i\}$ of points $x_n \in E \setminus a$ converging to a , the sequence $\{f(x_n)\}$ converges to A .



Note that we choose the the points x_n to be in the "deleted" neighborhood $E \setminus a$ (neighborhood E with point a removed) to force us to choose a sequence that is not

$$a, a, a, a, a, a, a, \dots$$

That is, it forces us to choose different points for the sequence.

Theorem 3.2 (Properties of Limits). *Given two numerical valued functions $f, g : E \subset \mathbb{R} \rightarrow \mathbb{R}$ with a common domain where $g(x) \neq 0$ for all $x \in E$, let*

$$\lim_{x \rightarrow a} f(x) = A, \quad \lim_{x \rightarrow a} g(x) = B$$

then,

$$\lim_{x \rightarrow a} (f + g)(x) = A + B$$

$$\lim_{x \rightarrow a} (cf)(x) = cA$$

$$\lim_{x \rightarrow a} (f \cdot g)(x) = A \cdot B$$

$$\lim_{x \rightarrow a} \left(\frac{f}{g} \right)(x) = \frac{A}{B}$$

Proof. The the Cauchy sequence criterion for a limit, this theorem is an immediate consequence on the corresponding theorem on limits of sequences. ■

We end this with a theorem connecting the relationship between a limit of a function as $x \rightarrow a$ and its ultimate behavior as $x \rightarrow a$.

Theorem 3.3. *Let $f : E \rightarrow \mathbb{R}$ be a function. Then,*

1. *f is ultimately the constant A as $x \rightarrow a$ implies that $\lim_{x \rightarrow a} f(x) = A$.*
2. *$\lim_{x \rightarrow a} f(x)$ implies that f is ultimately bounded as $x \rightarrow a$.*

3.0.1 Infinitesimal Functions

Definition 3.3 (Infinitesimal Function). A function $f : E \subset \mathbb{R} \rightarrow \mathbb{R}$ is said to be *infinitesimal* as $x \rightarrow a$ if

$$\lim_{x \rightarrow a} f(x) = 0$$

Lemma 3.4 (Sums, Products of Infinitesimals). *It is clear that if α, β are infinitesimal as $x \rightarrow a$, then*

1. $\alpha + \beta$ is infinitesimal as $x \rightarrow a$
2. $\alpha \cdot \beta$ is infinitesimal as $x \rightarrow a$

Furthermore, if α is infinitesimal and β is ultimately bounded as $x \rightarrow a$, then the product $\alpha \cdot \beta$ is infinitesimal as $x \rightarrow a$.

Proof. We prove all three statements.

1. Assume that α and β are infinitesimal as $x \rightarrow a$. Then, let us fix a small $\epsilon > 0$. This means that for every $\frac{\epsilon}{2}$ there exists an open deleted neighborhood $\dot{U}'(a)$ such that its image $\alpha(\dot{U}'(a)) \subset U'_{\epsilon/2}(0) \subset \mathbb{R}$. Additionally, for every $\frac{\epsilon}{2}$ there exists an open deleted neighborhood $\dot{U}''(a)$ such that its image $\beta(\dot{U}''(a)) \subset U'_{\epsilon/2}(0) \subset \mathbb{R}$. Thus, for the deleted neighborhood

$$\dot{U}(a) \subset \dot{U}'(a) \cup \dot{U}''(a)$$

we can see that for all $x \in \dot{U}(a)$,

$$|(\alpha + \beta)(x)| = |\alpha(x) + \beta(x)| \leq |\alpha(x)| + |\beta(x)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and hence $(\alpha + \beta)(\dot{U}(a)) \subset U_\epsilon(0)$.

2. This case is a special case of assertion 3. That is, every function that has a limit is ultimately bounded.
3. Since $\beta(x)$ is ultimately bounded, this means that there exists a constant M and an open deleted neighborhood $\dot{U}'(a) \subset E$ such that for all $x \in \dot{U}'(a)$, its image is bounded: $|\beta(x)| < M$. Let us fix a small $\epsilon > 0$. Then, by definition of the limit, for every $\frac{\epsilon}{M}$ there exists an open deleted neighborhood $\dot{U}''(a)$ such that its image $\beta(\dot{U}''(a)) \subset U_{\epsilon/M}(0) \subset \mathbb{R}$. Therefore, for the deleted neighborhood

$$\dot{U}(a) \subset \dot{U}'(a) \cup \dot{U}''(a)$$

we can see that for all $x \in \dot{U}(a)$,

$$|(\alpha \cdot \beta)(x)| = |\alpha(x)\beta(x)| = |\alpha(x)||\beta(x)| < \frac{\epsilon}{M} \cdot M = \epsilon$$

Therefore, $(\alpha \cdot \beta)(\dot{U}(a)) \subset U_\epsilon(0)$. ■

Note that in proving these properties of the limits, we have used the following fact about open deleted neighborhoods around a .

1. $\mathring{U}(a)$ is not the empty set.
2. Given open deleted neighborhoods $\mathring{U}'(a)$ and $\mathring{U}''(a)$, there exists an open deleted neighborhood in the intersections of these neighborhoods.

$$\mathring{U}(a) \subset \mathring{U}'(a) \cup \mathring{U}''(a)$$

These facts can be used to generalize the concept of limits as limits over a certain *filter base*.

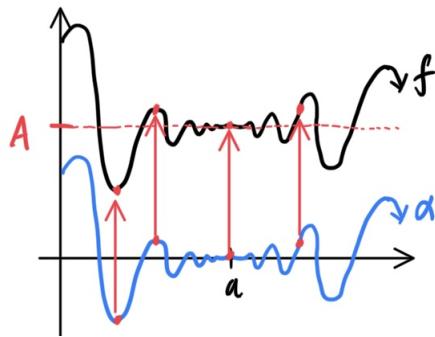
Theorem 3.5 (Representation of a Convergent Function as a Shift of its Infinitesimal). *Given a function $f : E \subset \mathbb{R} \rightarrow \mathbb{R}$, its limit exists and*

$$\lim_{x \rightarrow a} f(x) = A$$

if and only if f can be represented as

$$f(x) = A + \alpha(x)$$

where α is infinitesimal as $x \rightarrow a$. We can visualize this theorem by thinking of a function f that results from a "shift" of an infinitesimal.



Finally, we reiterate some limit theorems already stated for sequences, but now corresponding to functions. Interpreting the function limit as the Cauchy sequence definition of limits renders the proofs of these theorems trivial.

Theorem 3.6 (Behavior of Functions with Different Limits). *If the functions $f, g : E \rightarrow \mathbb{R}$ are such that*

$$\lim_{x \rightarrow a} f(x) = A < B = \lim_{x \rightarrow a} g(x)$$

then there exists a deleted neighborhood $U_\delta(a)$ in E at each point of which $f(x) < g(x)$.

Theorem 3.7 (Squeeze Theorem for Limits of Functions). *Given the functions $f, g, h : E \subset \mathbb{R} \rightarrow \mathbb{R}$ such that*

$$f(x) \leq g(x) \leq h(x) \text{ for all } x \in E$$

then,

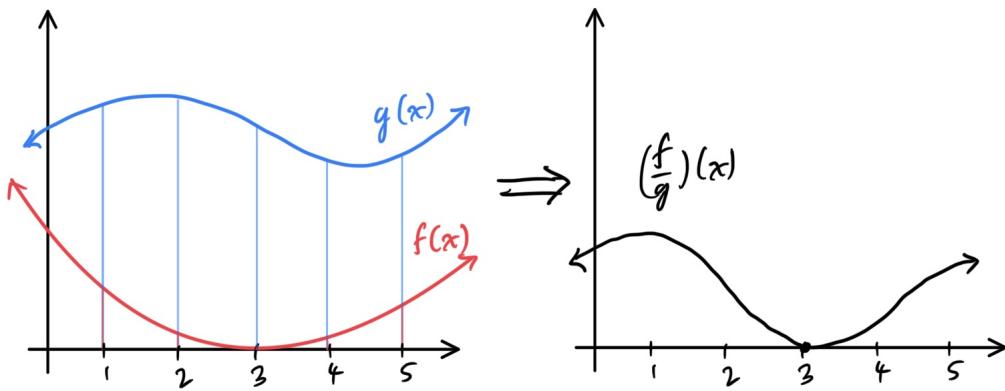
$$\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} h(x) = C \implies \lim_{x \rightarrow a} g(x) = C$$

3.1 Asymptotic Behavior of Functions

Definition 3.4 (Little-O Notation). The function $f : E \rightarrow \mathbb{R}$ is said to be *infinitesimal compared with the function $g : E \rightarrow \mathbb{R}$ as $x \rightarrow a$* , written (by abuse of notation) $f = o(g)$ as $x \rightarrow a$, if

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 1$$

or in other words, if f/g is an infinitesimal function as $x \rightarrow a$. Therefore, $f = o(1)$ as $x \rightarrow a$ means that f is infinitesimal as $x \rightarrow a$.

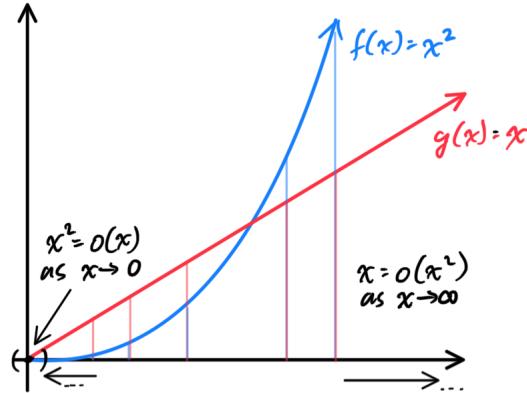


Note that writing $f = o(g)$ is again, an abuse of notation. $f = o(g)$ is really a shorthand way of writing that f is in the class of functions that is infinitesimal compared with the function g .

Intuitively, $f = o(g)$ means that the ratio between $f(x)$ and $g(x)$ will tend to infinity as $x \rightarrow a$ (this does not mean that f will be infinitely greater than g , however!). For example, looking at the two functions $f(x) = x^2$ and $g(x) = x$, we have

1. $x^2 = o(x)$ as $x \rightarrow 0$ (since $\frac{x^2}{x} = x$ is infinitesimal as $x \rightarrow 0$)
2. $x = o(x^2)$ as $x \rightarrow \infty$ (since $\frac{x}{x^2} = \frac{1}{x}$ is infinitesimal as $x \rightarrow \infty$)

We can visualize $g/f(x)$ tending to infinity within a neighborhood of 0 and $f/g(x)$ tending to infinity within a neighborhood of ∞ .



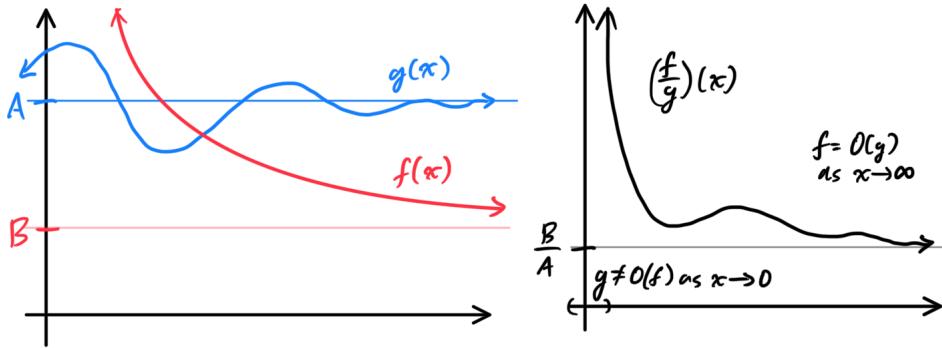
Definition 3.5 (Orders of Infinitesimals, Infinities). If $f = o(g)$ and g is infinitesimal as $x \rightarrow a$, then f is an *infinitesimal of higher order than g as $x \rightarrow a$* . Furthermore, if f and g are infinite functions as $x \rightarrow a$ and $f = o(g)$ as $x \rightarrow a$, then g is a *higher order infinity than f as $x \rightarrow a$* .

Definition 3.6 (Big-O Notation). By abuse of notation, $f = O(g)$ as $x \rightarrow a$ means that

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \infty$$

or in other words, f/g is ultimately bounded as $x \rightarrow a$. In particular, $f = O(1)$ as $x \rightarrow a$ means that f is bounded within a certain neighborhood $U(a)$ of a .

In the visual below, we can see that $f = O(g)$ as $x \rightarrow +\infty$ since the limit converges to constant $\frac{B}{A}$ which is bounded. In fact, at any other positive real number x , $(f/g)(x)$ is finite and is therefore bounded. However, at every neighborhood of $x = 0$, $(f/g)(x)$ is unbounded, meaning that $g \neq O(f)$ as $x \rightarrow 0$.

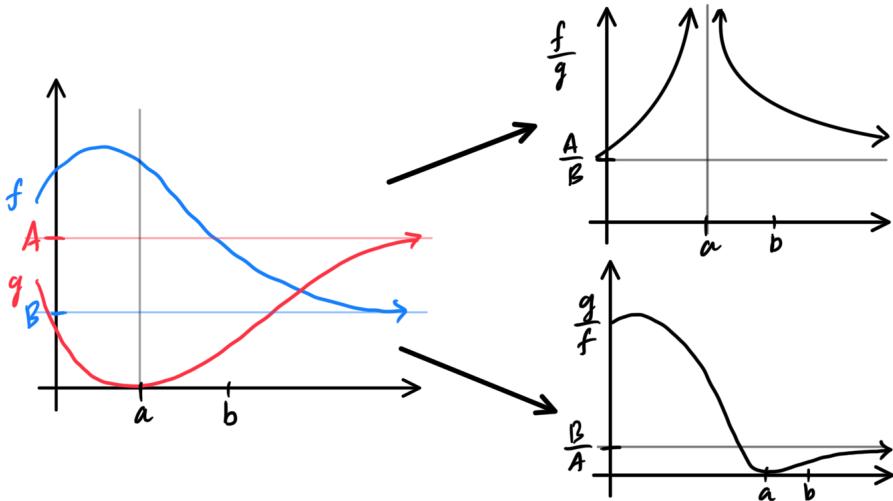


Definition 3.7 (Functions of Same Order). The functions f and g are of the same order as $x \rightarrow a$, written

$$f \asymp g \text{ as } x \rightarrow a$$

if $f = O(g)$ and $g = O(f)$ as $x \rightarrow a$. Intuitively, this means that the ratio between f and g within some deleted neighborhood of a is finite.

In the visual below, we can see that as long as $k \neq a$, $f = O(g)$ as $x \rightarrow k$ and as $x \rightarrow \infty$. In other words, the function f/g becomes ultimately bounded at every other point other than a , and f/g is unbounded within every neighborhood of a . When looking at g/f , we can see that this function is bounded for all $x \in \mathbb{R}$ and therefore $g = O(f)$ as $x \rightarrow k$ for all k .



Therefore, we can see that as long as $k \neq a$, $f \asymp g$ as $x \rightarrow k$.

Note that the condition that f and g be of the same order as $x \rightarrow a$ is (by definition of ultimately bounded functions) equivalent to the condition that there exist $c_1, c_2 > 0$ and an open neighborhood $U(a)$ such that the relations

$$c_1|g(x)| \leq |f(x)| \leq c_2|g(x)|$$

is true for $x \in U(a)$.

Definition 3.8 (Asymptotic Equivalence of Functions). For functions f and g , if

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 1$$

we say that f behaves asymptotically like g as $x \rightarrow a$, or that f is equivalent to g as $x \rightarrow a$, written

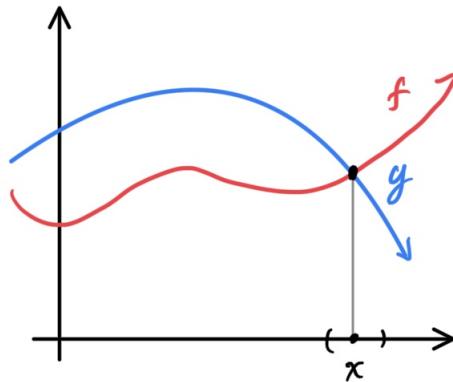
$$f \sim g \text{ as } x \rightarrow a$$

Moreover, \sim is an equivalence relation, which means that

1. $f \sim f$ as $x \rightarrow a$
2. $f \sim g$ as $x \rightarrow a \implies g \sim f$ as $x \rightarrow a$
3. $f \sim g$ and $g \sim h$ as $x \rightarrow a \implies f \sim h$ as $x \rightarrow a$

We list a few examples in order to develop some sort of visual intuition for when two functions are asymptotically equivalent.

1. If $f(a) = g(a) \neq 0$, then $f \sim g$ trivially since the ratio of f and g converges to 1 within a neighborhood of a .

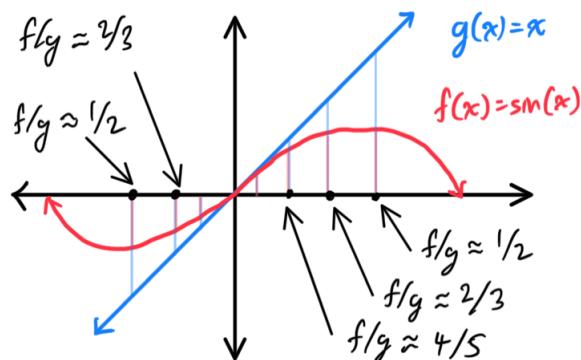


2. When $f(a) = g(a) = 0$, it may be f may be equivalent to g or one function may be infinitesimally smaller than the other.

- (a) When $f(x) = \sin x$ and $g(x) = x$, then $f \sim g$ since we see that

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$$

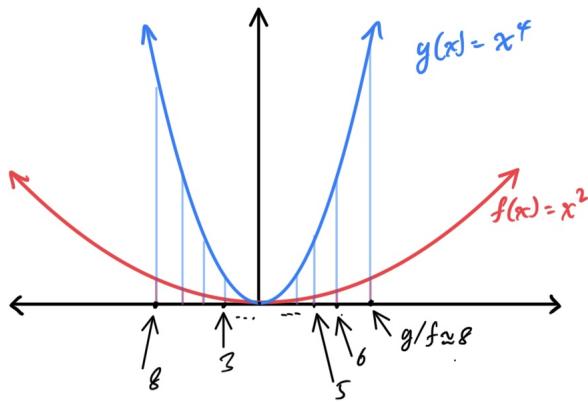
and so $\sin x \sim x$ as $x \rightarrow 1$



- (b) When $f(x) = x^2$ and $g(x) = x^4$, then

$$\lim_{x \rightarrow 0} \frac{x^4}{x^2} = 0$$

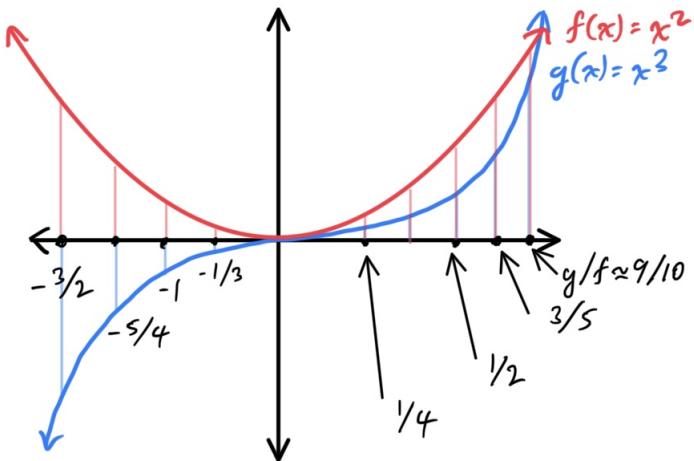
and so $x^4 \not\sim x^2$. In fact, $x^4 = o(x^2)$. Therefore, since x^4 decreases to 0 infinitely faster than x^2 , they are not equivalent.



(c) When $f(x) = x^2$ and $g(x) = x^3$, then

$$\lim_{x \rightarrow 0} \frac{x^3}{x^2} = 0$$

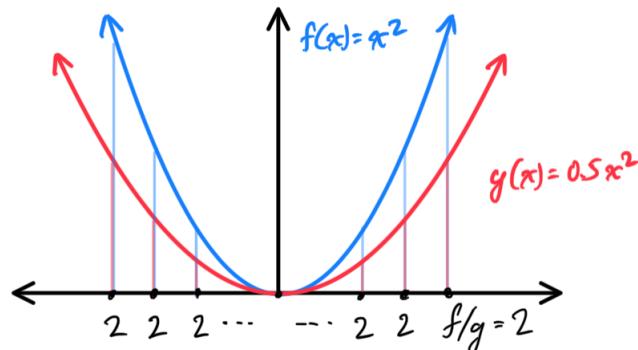
and so $x^3 \not\sim x^2$. In fact, $x^3 = o(x^2)$. Therefore, since x^4 decreases to 0 infinitely faster than x^2 , they are not equivalent.



(d) When $f(x) = x^2$ and $g(x) = 0.5x^2$, then

$$\lim_{x \rightarrow 0} \frac{0.5x^2}{x^2} = \frac{1}{2}$$

and so $0.5x^2 \not\sim x^2$. Therefore, since $0.5x^2$ is always as small as x^2 , they are not equivalent.

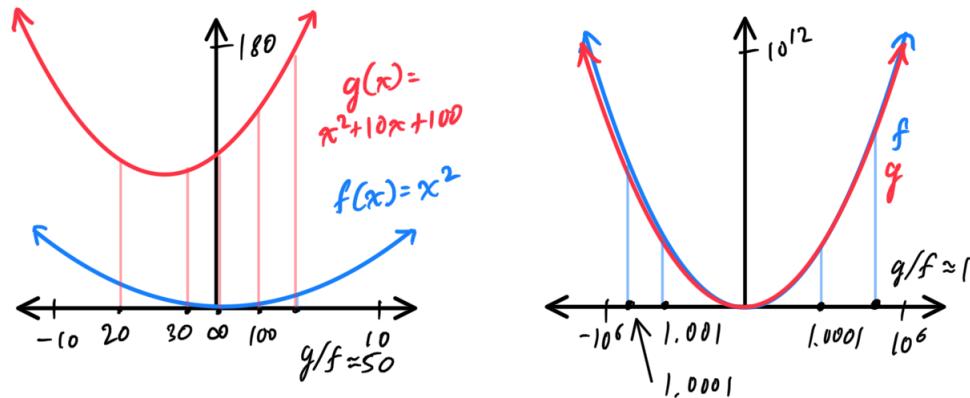


3. When analyzing the behavior of functions as $x \rightarrow \infty$, we can picture the two graphs of f and g on the plane and "zoom out" to see if the ratio of the values converge to 1. This would mean that as

$x \rightarrow \infty$, we should see the graphs overlapping more and more. For example, taking $f(x) = x^2$ and $g(x) = x^2 + 10x + 100$, we can see that the discrepancy is high around a neighborhood of $x = 0$. But as $x \rightarrow +\infty$, we get

$$\lim_{x \rightarrow +\infty} \frac{x^2 + 10x + 100}{x^2} = 1$$

and so the graphs look like they are overlapping.



Notice that even though the absolute difference $|(x^2 + 10x + 100) - x^2| = |10x + 100|$ tends to infinity, this difference increases infinitesimally compared to f and g .

From this, we can see that if $f \sim g$ as $x \rightarrow a$, then their difference

$$f - g = o(g) = o(f)$$

That is, $(f - g)(x)$ is infinitesimal compared to g or f (doesn't matter which one we compare it to). This leads to our next section, where we formalize this concept with absolute and relative errors.

3.1.1 Approximations of Functions

It is useful to note that since the relation $\lim_{x \rightarrow a} \gamma(x) = 1$ is equivalent to

$$\gamma(x) = 1 + \alpha(x), \text{ where } \lim_{x \rightarrow a} \alpha(x) = 0$$

the relation $f \sim g$ as $x \rightarrow a$ is equivalent to saying that

$$\frac{f(x)}{g(x)} = \gamma(x), \text{ where } \lim_{x \rightarrow a} \gamma(x) = 1$$

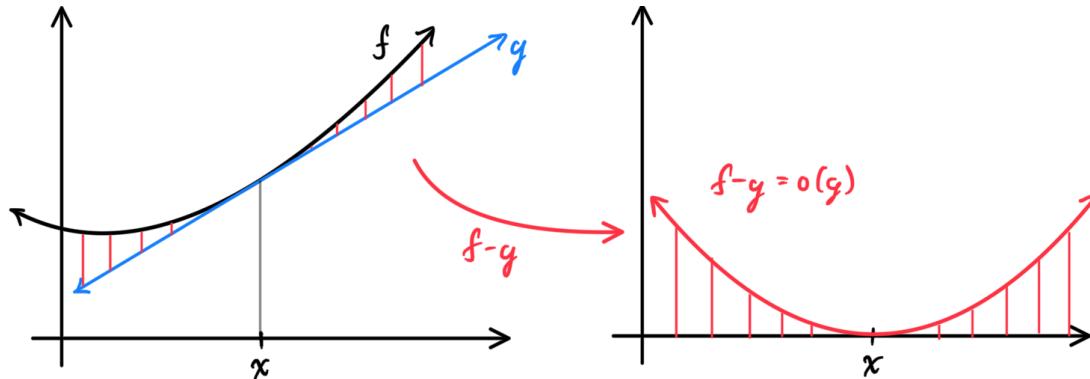
which implies

$$f(x) = g(x) + \alpha(x)g(x) = g(x) + o(g(x)) \text{ as } x \rightarrow a$$

or, symmetrically,

$$g(x) = f(x) + \alpha(x)f(x) = f(x) + o(f(x)) \text{ as } x \rightarrow a$$

This means that f can be exactly represented by another function g , plus another (error) function $o(g(x))$ that is infinitesimal compared to g .



Note that it is not a sufficient condition that the error function be infinitesimal! The error function $f - g$ must be infinitesimal **compared to g** ! This tells us that not only does the error function decrease infinitesimally, but also is infinitesimal compared to the approximation function we already have, which is in general a much stronger claim. This representation of certain types functions will provide the foundation for differential calculus when we talk about "good" approximations for a function.

Definition 3.9 (Relative Error). Since $f \sim g$ as $x \rightarrow a$ means that

$$f(x) = g(x) + o(x)g(x) = g(x) + o(g(x))$$

we can define the *relative error* of g as an approximation of f to be

$$|\alpha(x)| = \left| \frac{f(x) - g(x)}{g(x)} \right|$$

Clearly, since $f \sim g$, the relative error must be infinitesimal as $x \rightarrow a$.

We use the following lemma to check whether two functions are asymptotically equivalent.

Lemma 3.8. $f \sim g$ as $x \rightarrow a$ if and only if the relative error of g is infinitesimal as $x \rightarrow a$.

Example 3.4. We claim that

$$x^2 + x = \left(1 + \frac{1}{x}\right)x^2 \sim x^2 \text{ as } x \rightarrow \infty$$

We see that the absolute error of this approximation

$$|(x^2 + x) - x^2| = |x|$$

tends to infinity, but the relative error

$$\frac{|x|}{x^2} = \frac{1}{|x|}$$

tends to 0 as $x \rightarrow \infty$.

Theorem 3.9 (Prime Number Theorem). Let $\pi(x)$ be the number of prime numbers strictly less than x . Then $\pi \sim \frac{x}{\ln x}$ as $x \rightarrow +\infty$, or more precisely,

$$\pi(x) = \frac{x}{\ln x} + o\left(\frac{x}{\ln x}\right) \text{ as } x \rightarrow +\infty$$

Example 3.5. It is a fact that $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$, so we have $\sin x \sim x$ as $x \rightarrow 0$. So,

$$\sin x = x + o(x) \text{ as } x \rightarrow 0$$

The following theorem proves useful when computing limits.

Theorem 3.10. If $f \sim \tilde{f}$ as $x \rightarrow a$, then

$$\lim_{x \rightarrow a} f(x)g(x) = \lim_{x \rightarrow a} \tilde{f}(x)g(x)$$

provided one of these limits exist.

Theorem 3.11 (Properties of $o(g)$ and $O(g)$ Functions). For $x \rightarrow a$,

1. $o(f) + o(f) = o(f)$
2. $o(f)$ is also $O(f)$
3. $o(f) + O(f) = O(f)$
4. $O(f) + O(f) = O(f)$
5. If $g(x) \neq 0$, then

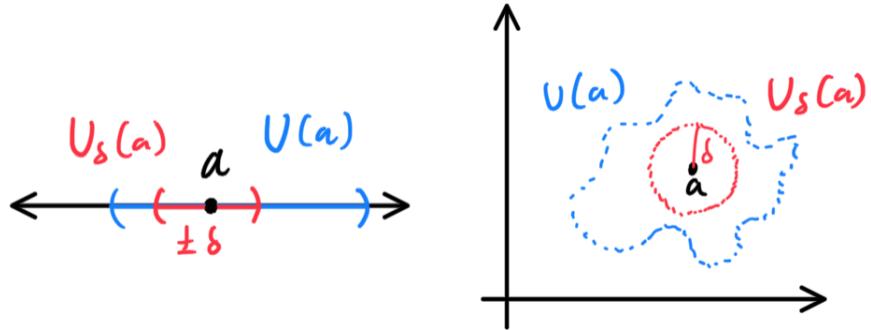
$$\frac{o(f(x))}{g(x)} = o\left(\frac{f(x)}{g(x)}\right), \text{ and } \frac{O(f(x))}{g(x)} = O\left(\frac{f(x)}{g(x)}\right)$$

4 Continuous Functions

Definition 4.1 (Continuity of a Function). A function f is *continuous at point a* if for any neighborhood $V(f(a))$ of $f(a)$, there is a neighborhood $U(a)$ of a whose image under the mapping f is contained in $V(f(a))$.

Generalizing this, we say that a function is *(globally) continuous* if the preimage of every neighborhood in its codomain is an open set in its domain.

The equivalent of these statements for functions $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ follows from the fact that any neighborhood of a point contains a symmetric neighborhood of the point.



Lemma 4.1 (Existence of Limits of Continuous Functions). $f : E \rightarrow \mathbb{R}$ is continuous at $a \in E$, where a is a limit point of E if and only if

$$\lim_{x \rightarrow a} f(x) = f(a)$$

Proof. The limit equalling $f(a)$ means that, by definition, for any arbitrarily small deleted neighborhood of $f(a)$, denoted $U_{f(a)} \setminus f(a)$, its preimage will be an open neighborhood of a , which itself will contain an open set. ■

This also means that we can use the Cauchy limit definition to define continuity of a function at a point. That is, for any sequence $\{a_n\}$ of points in codomain E which converges to point a , the function f is continuous at a if the corresponding sequence $\{f(a_n)\}$ converges to $f(a)$.

Theorem 4.2. This means that the continuous functions commute with the operation of passing to the limit at a point.

$$\lim_{x \rightarrow a} f(x) = f\left(\lim_{x \rightarrow a} x\right)$$

Lemma 4.3 (Properties of Continuous Functions). Let $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^m \rightarrow \mathbb{R}^p$ with $c \in \mathbb{R}$.

1. f continuous at $x_0 \implies cf$ continuous at x_0 .
2. f, g continuous at $x_0 \implies f + g$ continuous at x_0 .
3. Let $m = 1$. f, g continuous at $x_0 \implies fg$ continuous at x_0 .
4. f continuous at x_0 and $f(x) \neq 0 \forall x \in \mathbb{R}^n \implies 1/f$ continuous at x_0 .
5. If $f(x) = (f_1(x), f_2(x), \dots, f_n(x))$ coordinate-wise, then

$$f \text{ continuous at } x_0 \iff f_1, f_2, \dots, f_m \text{ continuous at } x_0$$

6. f continuous at x_0 and g continuous at $y_0 = f(x_0) \implies g \circ f$ continuous at x_0 .

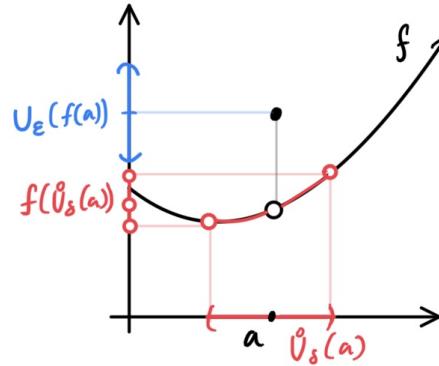
Proof. This is an immediate result of the equivalence of a function being continuous at point a and its limit at point a existing. ■

4.1 Points of Discontinuity

Definition 4.2 (Discontinuity). If the function $f : E \rightarrow \mathbb{R}$ is not continuous at a point of E , then this point is called a *point of discontinuity*, or simply a *discontinuity* of f .

That is, a is a point of discontinuity of f if for some neighborhood $V(f(a))$ of $f(a)$, there exists no neighborhood of a whose image under the mapping f is contained in $V(f(a))$. There are three types of discontinuities:

1. A *removable discontinuity* is characterized by the fact that the limit $\lim_{x \rightarrow a} f(x) = A$ exists, but $A \neq f(a)$.



This means that we can modify f and define a new function $\tilde{f} : E \rightarrow \mathbb{R}$ as

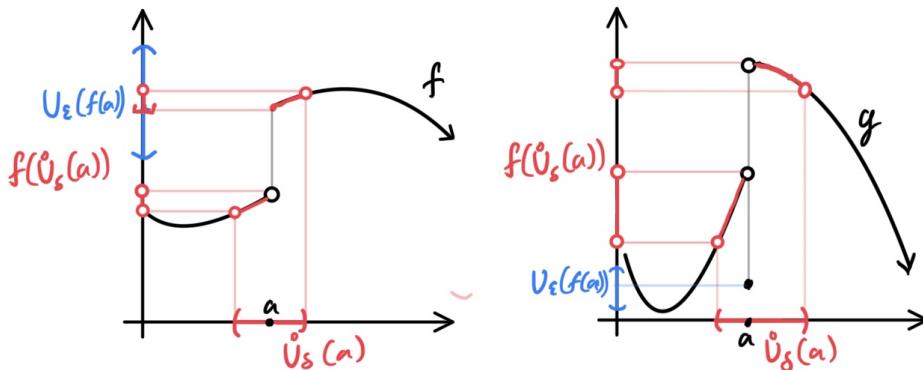
$$\tilde{f}(x) = \begin{cases} f(x), & x \in E \setminus a \\ A, & x = a \end{cases}$$

which would be continuous on E .

2. A *discontinuity of first kind*, also known as a jump/step discontinuity, is characterized by both the left and right-hand limits

$$\lim_{x \rightarrow a-0} f(x) \text{ and } \lim_{x \rightarrow a+0} f(x)$$

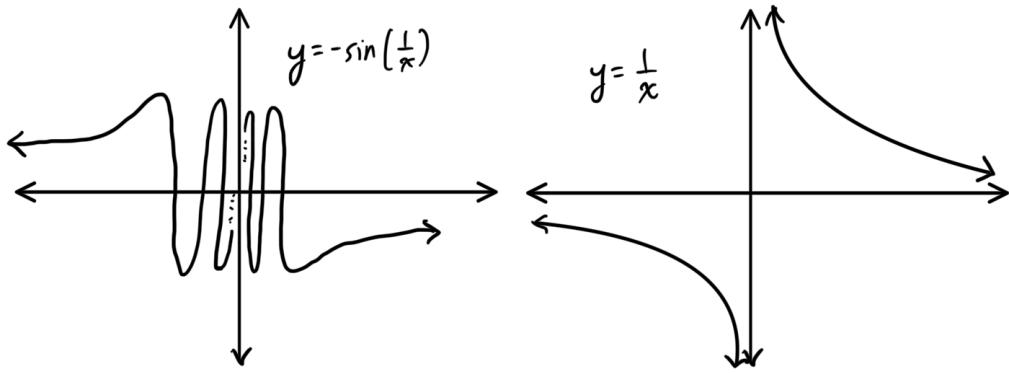
existing, but at least one of them is not equal to the value $f(a)$ that the function assumes at a .



3. A *discontinuity of second kind*, also known as an essential discontinuity, is characterized by at least one of the two limits

$$\lim_{x \rightarrow a-0} f(x) \text{ and } \lim_{x \rightarrow a+0} f(x)$$

not existing.



Note that strictly speaking, a removable discontinuity is really a discontinuity of first kind, but in this context we distinguish them.

Example 4.1 (Dirichlet Function). *The Dirichlet function, defined*

$$\mathcal{D}(x) = \begin{cases} 1, & \text{if } x \in \mathbb{Q} \\ 0, & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

is discontinuous at every point, and obviously all of its discontinuities are of second kind, since in every interval there are both rational and irrational numbers and therefore there exists no limit at any point $a \in \mathbb{R}$.

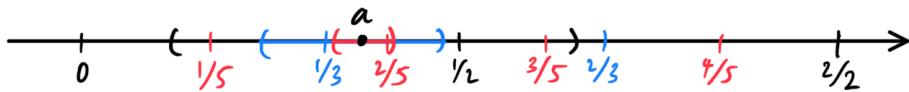
More specifically, given any point $a \in \mathbb{R}$, assume that a is rational. We can set $\epsilon = 0.1$ -neighborhood around the value 1, but no matter how small we let δ , the interval $(a - \delta, a + \delta)$ will contain both rationals and irrationals, meaning that it will map to $\{0, 1\}$ always, which is not fully contained in $(0.9, 1.1)$.

Here is a slightly more interesting example.

Example 4.2 (Riemann Function). *Let the Riemann function \mathcal{R} be defined*

$$\mathcal{R}(x) = \begin{cases} \frac{1}{n}, & \text{if } x = \frac{m}{n} \in \mathbb{Q}, \text{ where } \gcd(m, n) = 1 \\ 0, & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

We first note that for any point $a \in \mathbb{R}$, any bounded neighborhood $U(a)$ of it, and any number $N \in \mathbb{N}$, the neighborhood $U(a)$ contains only a finite number of rational numbers $\frac{m}{n}$, where $n < N$. By shrinking the neighborhood, we can assume that the denominators of all rational numbers in the neighborhood are larger than N . We can visualize why this is by seeing that rational numbers with larger denominators have smaller "gaps" between them.



Thus, at any point $x \in U(a) \setminus a$, we have

$$|\mathcal{R}(x)| < \frac{1}{N}$$

and therefore

$$\lim_{x \rightarrow a} \mathcal{R}(x) = 0$$

at any point $a \in \mathbb{R} \setminus \mathbb{Q}$. Hence, the Riemann function is continuous at any irrational number.

4.2 Properties of Continuous Functions

Theorem 4.4 (Local Properties of Continuous Functions). *Let $f : E \rightarrow \mathbb{R}$ be a function that is continuous at the point $a \in E$. Then,*

1. *f is bounded in some neighborhood $U(a)$.*
2. *If $f(a) \neq 0$, then in some neighborhood $U(a)$ all the values of the function have the same sign as $f(a)$.*
3. *If the function $g : U(a) \subset E \rightarrow \mathbb{R}$ is defined in some neighborhood of a and is continuous at a , then the following functions*

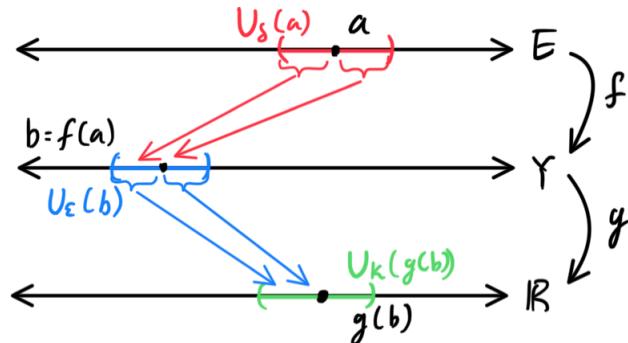
$$\begin{aligned} & (f + g)(x) \\ & (f \cdot g)(x) \\ & \left(\frac{f}{g}\right)(x) \text{ where } g(a) \neq 0 \end{aligned}$$

are also defined in $U(a)$ and continuous at a .

4. *If the function $g : Y \rightarrow \mathbb{R}$ is continuous at a point $b \in Y$ and f is such that $f : E \rightarrow Y$, $f(a) = b$, and f is continuous at a , then the composite function*

$$g \circ f : E \rightarrow \mathbb{R}$$

is defined on E and continuous at a .



This is easy to see because given the open neighborhood of $g(b)$, we know for a fact that $U_\delta(a)$ maps completely into $U_\epsilon(b)$, and that $U_\epsilon(b)$ maps completely into $U_\kappa(g(b))$ and so the composition of these mappings must mean that $U_\delta(a)$ maps completely into $U_\kappa(g(b))$.

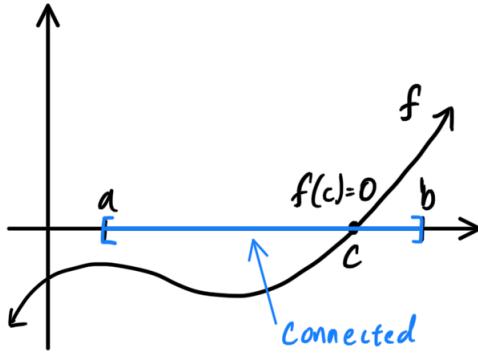
Example 4.3. An algebraic polynomial

$$P(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \dots + a_{n-1} x + a_n$$

is a continuous function on \mathbb{R} . Since $f(x) = x$ and $f(x) = c$ are continuous functions, by induction on x , we can multiply them together to find that $f(x) = x^n$ is continuous, which implies that ax^n is continuous, which implies that the sums of these functions are also continuous.

Unlike local properties, the global property of a function is a property involving the entire domain of definition of the function.

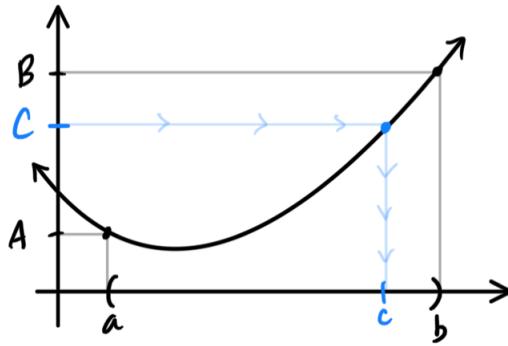
Theorem 4.5 (Intermediate Value Theorem). *If a function that is continuous on a closed interval assumes values with different signs at the endpoints of the interval, then there is a point in the interval where it assumes the value 0.*



Proof. ■

This following proof provides a very simple algorithm for finding the zero of the equation $f(x) = 0$ on an interval whose endpoints has values with opposite signs. Note that the colloquial description of the intermediate value theorem, that is it is impossible to pass continuously from positive to negative values without assuming the value 0 along the way), assumes more than they state. That is, this theorem is actually dependent on the domain of definition: that is it is a closed interval, or more generally, that it is *connected*.

Corollary 4.5.1. *If a function f is continuous on an open interval and assumes values $f(a) = A, f(b) = B$, then for any number $C \in (A, B)$, there is a point c between a and b such that $f(c) = C$.*



Theorem 4.6 (Weierstrass Maximum-Value Theorem). *A function that is continuous on a closed interval is bounded in that interval, with a maximum and minimum.*

4.2.1 Uniform Continuity

Roughly speaking, a function f is uniformly continuous if it is possible to guarantee that $f(x)$ and $f(y)$ be as close to each other as we please by requiring only that x and y be sufficiently close to each other.

Definition 4.3 (Uniform Continuity). A function $f : E \rightarrow \mathbb{R}$ is *uniformly continuous* on a set $E \subset \mathbb{R}$ if for every $\epsilon > 0$, there exists $\delta > 0$ such that

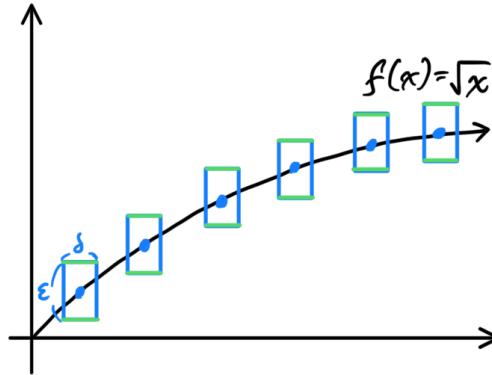
$$|f(x_1) - f(x_2)| < \epsilon$$

for all points $x_1, x_2 \in E$ such that $|x_1 - x_2| < \delta$.

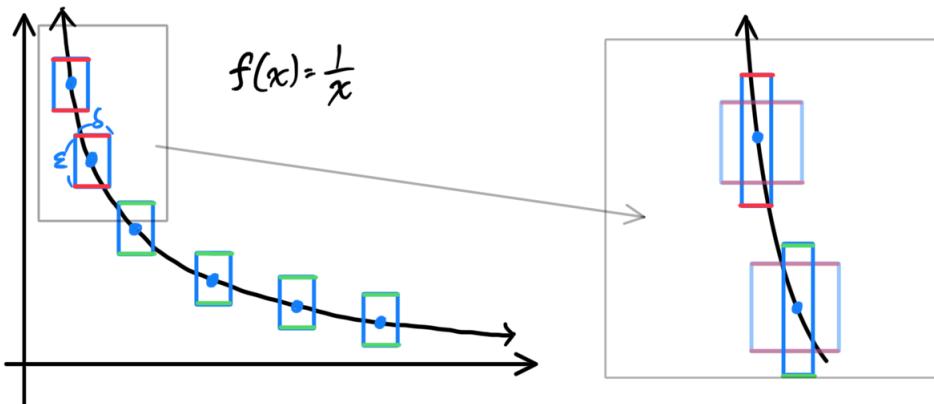
Intuitively, uniform continuity says that given any two points x, y in the domain where their distance is arbitrarily small (δ apart), we can guarantee that the distance between $f(x), f(y)$ is at maximum some arbitrarily small ϵ .

The following visual shows the radical function $f(x) = \sqrt{x}$ defined on \mathbb{R}^+ . We can see that it satisfies uniform continuity because the graph does not escape the top and/or bottom of the $\epsilon \times \delta$ window, no

matter where the box is located on the graph. More strictly speaking, no matter what we set the ϵ (how long the box is), uniform continuity says that we can choose a sufficient δ (width of the box) such that the graph does not escape the top/bottom of the window no matter where the window is.

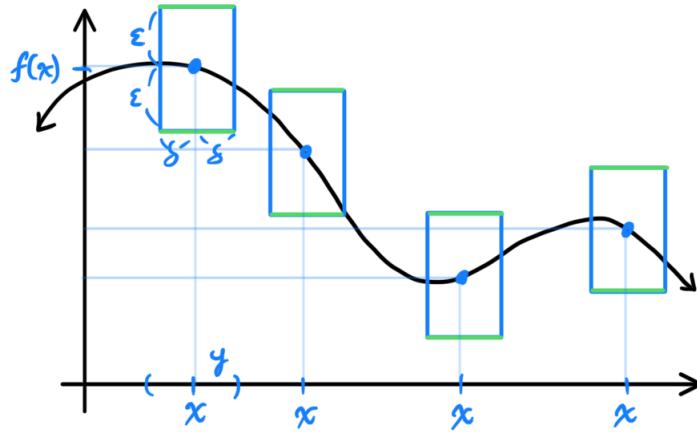


We can clearly see that the function $f(x) = 1/x$ is not uniformly continuous, since the graph escapes the $\epsilon \times \delta$ window at some point (marked in red). More strictly speaking, given any length ϵ of the window, we cannot create a thin-enough δ box that will contain the graph, since as $x \rightarrow 1$, the function becomes unbounded.



That is, arbitrarily thin boxes don't help when the slope is arbitrarily steep.

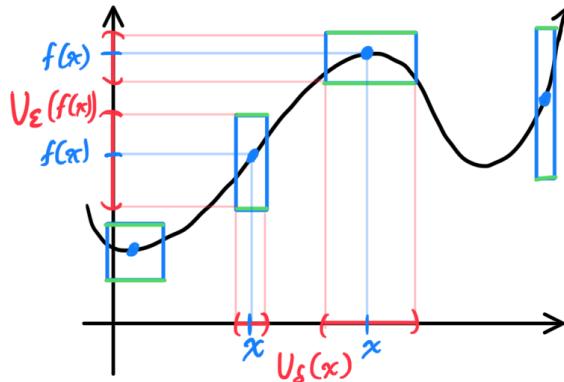
To compare uniform continuity with regular continuity, we can adapt this alternate (yet equivalent interpretation): Let there exist function $f : E \rightarrow \mathbb{R}$. Given any $\epsilon > 0$, we can choose a $\delta > 0$ such that given any point $x \in E$ and $f(x)$, as long as a second point y is δ away from x , then $f(y)$ is ϵ away from $f(x)$. This visualization would lead to there being a $2\epsilon \times 2\delta$ window around point x .



Uniform continuity means that the box above does not change dimensions no matter where the point is (hence, the name uniform). Therefore, given a certain $\epsilon > 0$, the way we choose δ is only dependent on ϵ , and so it must be a function of ϵ :

$$\delta = \delta(\epsilon)$$

However, in continuity, there just has to exist **some** δ -neighborhood of x such that its image is contained in the ϵ -neighborhood of $f(x)$. There are no restrictions on the dimensions of this box; it just has to exist.



Lemma 4.7. If f is uniformly continuous on the set E , it is continuous at each point of that set. However, the converse is not generally true.

Theorem 4.8 (Cantor's Theorem on Uniform Continuity). A function that is continuous on a closed interval is uniformly continuous on that interval.

Example 4.4. Let $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = 3x + 7$. Then f is uniformly continuous. Choose $\epsilon > 0$. Let $\delta = \epsilon/3$. Choose $x, y \in \mathbb{R}$ and assume $|x - y| < \delta$. Then,

$$|f(x) - f(y)| = |3x + 7 - 3y - 7| = 3|x - y| < 3\delta = \epsilon$$

Example 4.5. Let $f : (0, 4) \subset \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^2$. Then f is uniformly continuous on $(0, 4)$. Choose $\epsilon > 0$. Let $\delta = \epsilon/8$. Choose $x, y \in (0, 4)$ and assume $|x - y| < \delta$. Then,

$$|f(x) - f(y)| = |x^2 - y^2| = (x + y)|x - y| < (4 + 4)|x - y| = 8\delta = \epsilon$$

In both examples, the function satisfied an inequality of form

$$|f(x_1) - f(x_2)| \leq M|x_1 - x_2|$$

this is called the Lipschitz inequality.

4.2.2 Lipschitz Continuity

Lipschitz continuity is a strong form of uniform continuity for functions. Intuitively, a Lipschitz continuous function is limited in how fast it can change (by the Lipschitz constant).

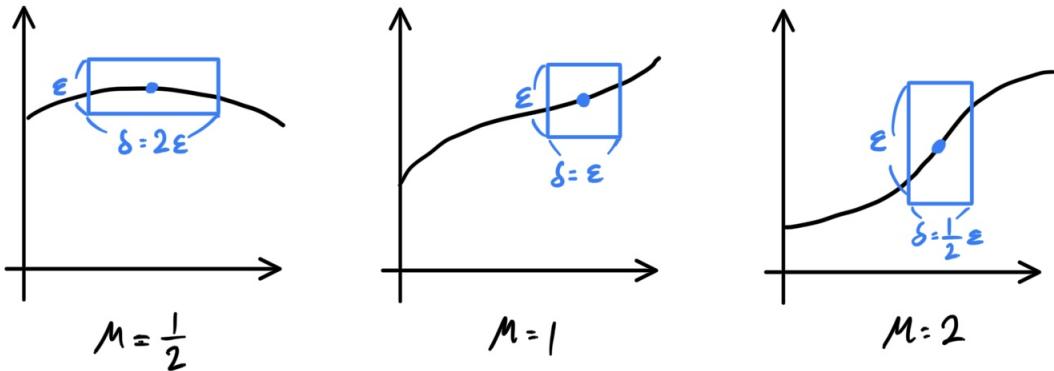
Definition 4.4 (Lipschitz Continuous Function). Given $f : E \subset \mathbb{R} \rightarrow \mathbb{R}$, f is *Lipschitz continuous* if there exists a positive real constant M such that for all real $x, y \in E$,

$$|f(x) - f(y)| \leq M|x - y|$$

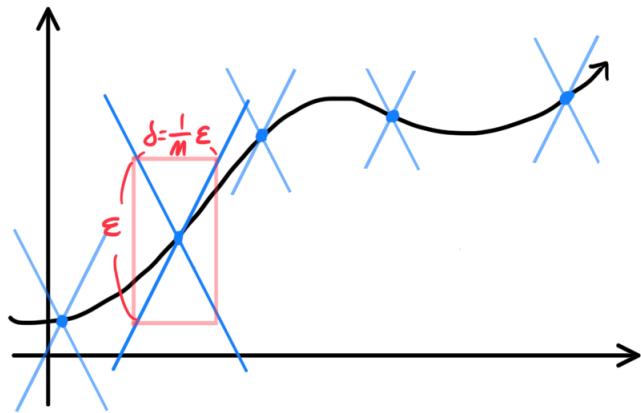
The corresponding M is called the *Lipschitz constant*, and the smallest constant M satisfying this inequality is called the *best Lipschitz constant*.

Note that Lipschitz continuity pops up as a very natural extension of uniform continuity. The inequality above just means that given an ϵ , we can choose a δ such that a linear multiple of δ is always greater than ϵ . This means that Lipschitz continuity is just uniform continuity such that the δ function is linear:

$$\delta = \delta(\epsilon) = \frac{1}{M}\epsilon$$



Another way to interpret uniform continuity is by seeing that the derivative of f is bounded by the slope M .

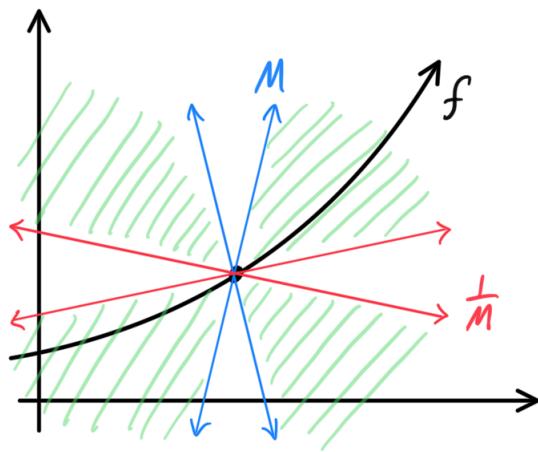


This slope bound implies that for every pair of points on the graph of this function, the absolute value of the slope of the line connecting them is not greater than M' . The smallest M' is the best Lipschitz constant.

Definition 4.5 (Bi-Lipschitz Continuity). A function $f : E \subset \mathbb{R} \rightarrow \mathbb{R}$ is *Bi-Lipschitz continuous* if there exists constant $M \geq 1$ such that for all real $x, y \in E$,

$$\frac{1}{M}|x - y| \leq |f(x) - f(y)| \leq M|x - y|$$

A visual of this map is shown, where the function f must always land in the shaded green area.



It immediately follows that for $x \neq y$, $|f(x) - f(y)|$ cannot equal 0, which means that a bilipschitz map is injective. A bilipschitz map is really just Lipschitz map with its inverse also being Lipschitz.

Proposition 4.9. *A bilipschitz map f is a homeomorphism onto its image.*

4.2.3 Inverse Function Theorem

We begin by introducing this intuitive lemma.

Lemma 4.10. *A continuous mapping $f : E \rightarrow \mathbb{R}$ of a closed interval $E = [a, b]$ into \mathbb{R} is injective if and only if the function f is strictly monotonic on $[a, b]$.*

Furthermore, every strictly monotonic function $f : X \subset \mathbb{R} \rightarrow \mathbb{R}$ (for arbitrary X) has an inverse

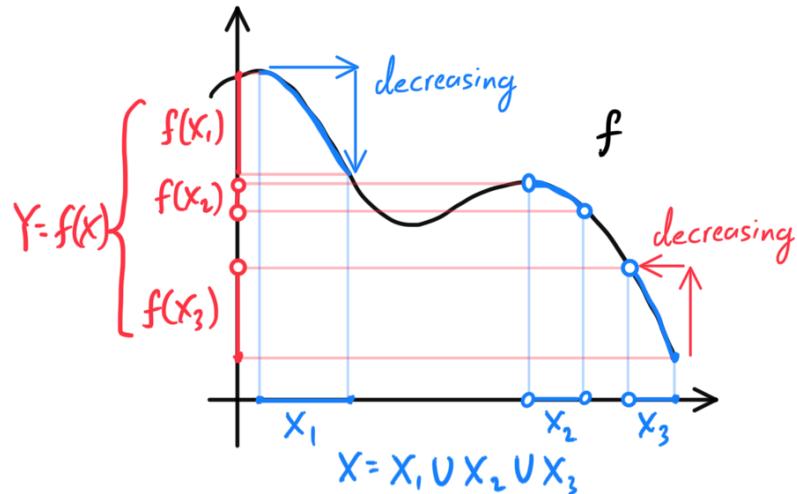
$$f^{-1} : f(X) \subset \mathbb{R} \rightarrow \mathbb{R}$$

with the same kind of monotonicity on $f(X)$ that f has on X .

Lemma 4.11 (Criterion for Continuity of a Monotonic Function). *A monotonic function $f : E \rightarrow \mathbb{R}$ defined on a closed interval $E = [a, b]$ is continuous if and only if its set of values $f(E)$ is the closed interval with endpoints $f(a)$ and $f(b)$.*

Note that both conditions imply that there are no points of discontinuities in the graph of f .

Theorem 4.12 (Inverse Function Theorem). *A function $f : X \rightarrow \mathbb{R}$ that is strictly monotonic on a set $X \subset \mathbb{R}$ has an inverse $f^{-1} : Y \rightarrow \mathbb{R}$ defined on the set $Y = f(X)$ of values of f . The function $f^{-1} : Y \rightarrow \mathbb{R}$ is monotonic and has the same type of monotonicity on Y that f has on X .*

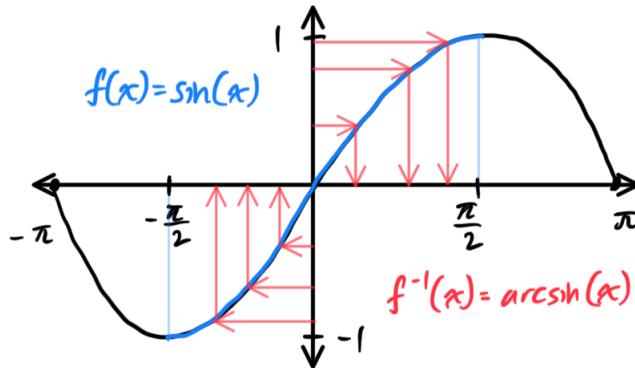


If in addition, X is a closed interval $[a, b]$ and f is continuous on X , then the set $Y = f(X)$ is the closed interval with endpoints $f(a)$ and $f(b)$ and the function $f^{-1} : Y \rightarrow \mathbb{R}$ is continuous on it.

Example 4.6. The function $f(x) = \sin x$ is increasing and continuous on the closed interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Hence, the restriction to the closed interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ has an inverse $x = f^{-1}(y)$, which is denoted by

$$x = \arcsin y$$

This function is defined on the closed interval $[-\sin(-\frac{\pi}{2}), \sin(\frac{\pi}{2})] = [-1, 1]$ and increases continuously from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$.



5 Differential Calculus

5.1 Functions Differentiable at a Point

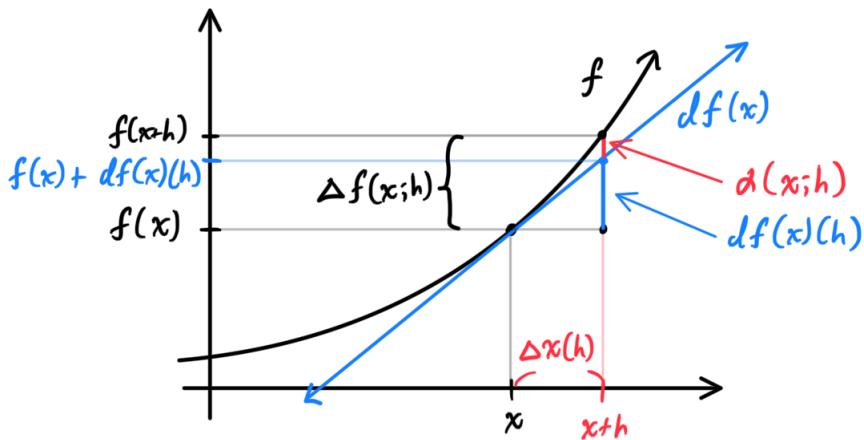
Definition 5.1 (Differentiable Function). A function $f : E \subset \mathbb{R} \rightarrow \mathbb{R}$ is *differentiable* at a given point x (that is a limit point of E) if there exists a linear function $h \mapsto df(x)h$ (called the *differential* of f) and an infinitesimal $\alpha(x; h) = o(h)$ as $h \rightarrow 0$, such that

$$f(x+h) - f(x) = df(x)(h) + \alpha(x; h)$$

Note that x is fixed; what we are really interested here is the h value. Furthermore,

1. $\Delta x(h) \equiv (x+h) - x = h$ is called the *increment of the argument*
2. $\Delta f(x; h) \equiv f(x+h) - f(x)$ is called the *increment of the function*

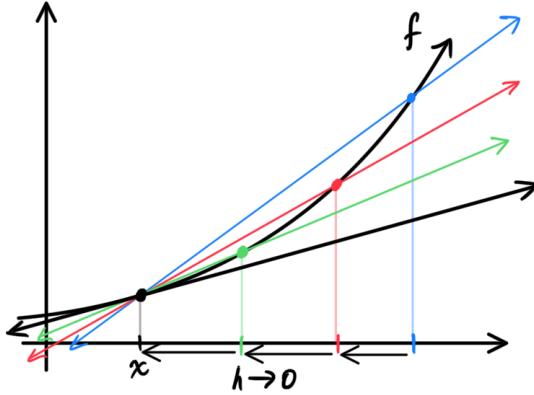
They are often denoted (inappropriately) by the symbols Δx and $\Delta f(x)$ representing functions of h . The differential and the infinitesimal can be visualized below.



Definition 5.2 (Derivative). Given function $f : E \subset \mathbb{R} \rightarrow \mathbb{R}$, the number

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

is called the *derivative* of the function f at x . It can be visualized as the sequence of the slopes of the secant lines converging onto the slope of the black tangent line as shown.



This equality can also be written in the equivalent form:

$$\frac{f(x+h) - f(x)}{h} = f'(x) + o(h)$$

where α is infinitesimal as $h \rightarrow 0$. This is also equivalent to:

$$f(x+h) - f(x) = f'(x)h + o(h)$$

where the error term $o(h) \rightarrow 0$ as $h \rightarrow 0$.

Note that we have defined the differentiability of a function at a point and the existence of its derivative at a point completely separately. But it turns out that the existence of this arbitrary number $f'(x)$ we call the "derivative," defined

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

actually has an equivalent form of

$$f(x+h) - f(x) = f'(x)h + o(h)$$

But since $f'(x)$ is in \mathbb{R} , the function $h \mapsto f'(x)h$ is linear and $o(h)$ is infinitesimal, so it is in the form

$$f(x+h) - f(x) = df(x)(h) + \alpha(x; h)$$

which, by definition, means that it is differentiable! Therefore, we have determined the equivalence between the differentiability of a function at a point and the existence of its derivative at the same point. Furthermore, this function $h \mapsto f'(x)h$ is precisely the differential of f , meaning that

$$df(x)(h) = f'(x)h$$

Furthermore,

$$\Delta f(x; h) - df(x)(h) = \alpha(x; h)$$

and $\alpha(x; h) = o(h)$ as $h \rightarrow 0$, or in other words, the difference between the increment of the function and the value of the function $df(x)$ in h is an infinitesimal of higher order than the first in h . For this reason, we say that the differential is the *principal linear part of the increment of the function*.

In particular, if $f(x) \equiv x$, then we have $f'(x) \equiv 1$ and

$$dx(h) = 1 \cdot h = h$$

Substituting this equality into $df(x)(h) = f'(x)h$, we get

$$df(x)(h) = f'(x) dx(h)$$

or without the input parameter h ,

$$df(x) = f'(x) dx$$

Note that this is an equality between two functions of h . From this, we obtain the familiar *Leibniz notation* of the derivative:

$$\frac{df(x)(h)}{dx(h)} = f'(x) \iff \frac{df(x)}{dx} = f'(x)$$

That is, the function $\frac{df(x)}{dx}$, which is the ratio of the functions $df(x)$ and dx , is constant and equals $f'(x)$.

5.2 Tangent Line: Geometric Meaning of the Derivative, Differential

Let us try to construct successive approximations to an arbitrary function $f : E \rightarrow \mathbb{R}$ at a given limit point x_0 . That is, we find a function g such that

$$f = g + o(g)$$

Depending on what g is, we can construct better approximations of f .

5.2.1 Constant Approximation

The 0th order approximation is when g is a constant. That is, $g \equiv c_0$ for some $c_0 \in \mathbb{R}$. This means

$$f(x) = c_0 + o(c_0) = c_0 + o(1) \text{ as } x \rightarrow x_0$$

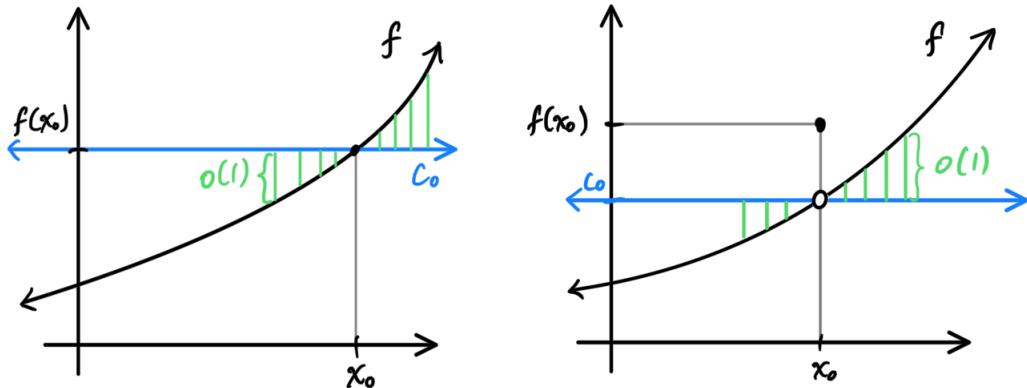
More precisely, we want this difference $f(x) - c_0$ to be $o(1)$ as $x \rightarrow x_0$, which means that it is simply infinitesimal. Visualizing this, we can see that given a constant approximation (labeled in blue) to a function at x_0 , its error term (labeled in green) is in fact, infinitesimal. All this boils down to the fact that

$$\lim_{x \rightarrow x_0} f(x) = c_0$$

If the function is continuous at x_0 , then

$$\lim_{x \rightarrow x_0} f(x) = f(x_0)$$

and naturally $c_0 = f(x_0)$. Both the continuous (left) and noncontinuous case (right) is shown, but in most cases, we will assume continuity.



5.2.2 Linear Approximation

The 1st order approximation is a linear function that approximates f as

$$f(x) = c_0 + c_1(x - x_0) + o(x - x_0) \text{ as } x \rightarrow x_0$$

Following the previous logic, assuming f continuous means that $c_0 = f(x_0)$. Furthermore, as $x \rightarrow x_0$

$$\begin{aligned} f(x) = c_0 + c_1(x - x_0) + o(x - x_0) &\implies c_1 = \frac{f(x) - c_0 - o(x - x_0)}{x - x_0} \\ &\implies c_1 = \frac{f(x) - c_0}{x - x_0} - \frac{o(x - x_0)}{x - x_0} \\ &\implies c_1 = \frac{f(x) - c_0}{x - x_0} - o(1) \\ &\implies c_1 = \lim_{x \rightarrow x_0} \frac{f(x) - c_0}{x - x_0} = f'(x_0) \end{aligned}$$

But this just means that $f'(x_0) = c_1$. Note that before, we have proved the equivalence of the existence of a derivative at x_0 with differentiability at x_0 (which itself means that there exists a linear approximation $df(x)(h)$ that is a function of h). Here, we have created a linear approximation with respect to $x = x_0 + h$, rather than h (shifted the function).

Therefore, the function

$$\alpha(x) = f(x_0) + f'(x_0)(x - x_0)$$

provides the best linear approximation to the function f in a neighborhood of x_0 in the sense that for any other function $\beta(x)$ of the form

$$\beta(x) = c_0 + c_1(x - x_0)$$

we have $f(x) - \beta(x) \neq o(x - x_0)$ as $x \rightarrow x_0$. The graph of the function α is the straight line

$$y - f(x_0) = f'(x_0)(x - x_0)$$

This leads to the definition of our familiar tangent line.

Definition 5.3 (Tangent Line). If a function $f : E \rightarrow \mathbb{R}$ is differentiable at a point $x_0 \in E$, the line defined by

$$y - f(x_0) = f'(x_0)(x - x_0)$$

is called the *tangent* to the graph of f at the point $(x_0, f(x_0))$.

5.2.3 Higher Order Approximations

We can continue this pattern to get a quadratic approximation of f in the form

$$f(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)^2 + o((x - x_0)^2) \text{ as } x \rightarrow x_0$$

As we have done in the previous subsection, we can derive (assuming continuity of f) $c_0 = f(x_0)$, $c_1 = f'(x_0)$. To derive what c_2 should be, we see that the equation above implies

$$c_2 = \frac{f(x) - c_0 - c_1(x - x_0) - o((x - x_0)^2)}{(x - x_0)^2} = \frac{f(x) - c_0 - c_1(x - x_0)}{(x - x_0)^2} - o(1)$$

which means

$$c_2 = \lim_{x \rightarrow x_0} \frac{f(x) - c_0 - c_1(x - x_0)}{(x - x_0)^2}$$

Extending this, if we are seeking a polynomial $P_n(x_0; x) = c_0 + c_1(x - x_0) + \dots + c_n(x - x_0)^n$ such that

$$f(x) = c_0 + c_1(x - x_0) + \dots + c_n(x - x_0)^n + o((x - x_0)^n) \text{ as } x \rightarrow x_0$$

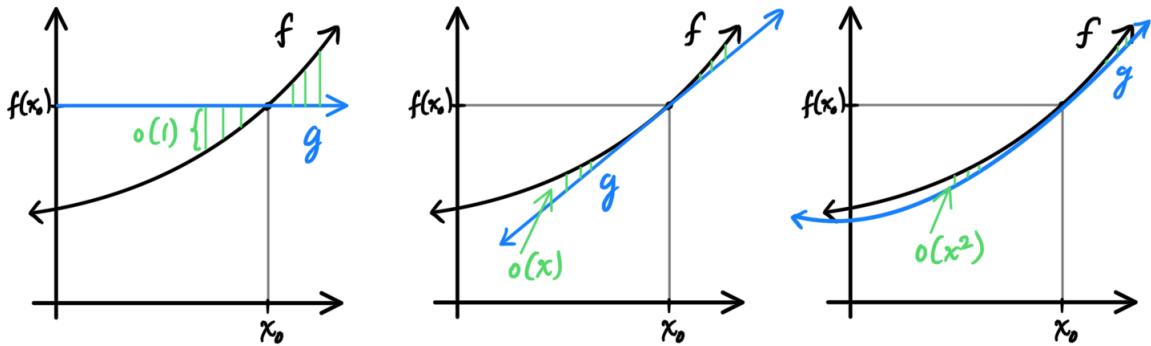
we would find

$$\begin{aligned} c_0 &= \lim_{x \rightarrow x_0} f(x) \\ c_1 &= \lim_{x \rightarrow x_0} \frac{f(x) - c_0}{x - x_0} \\ c_2 &= \lim_{x \rightarrow x_0} \frac{f(x) - c_0 - c_1(x - x_0)}{(x - x_0)^2} \\ \dots &= \dots \\ c_n &= \lim_{x \rightarrow x_0} \frac{f(x) - (c_0 + \dots + c_{n-1}(x - x_0)^{n-1})}{(x - x_0)^n} \end{aligned}$$

We formalize the order of these approximations by analyzing their error bound.

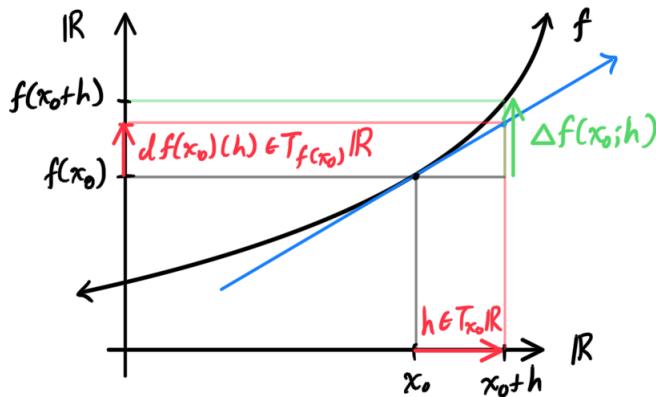
Definition 5.4 (nth Order Contact). If $f, g : E \rightarrow \mathbb{R}$ are continuous at point x_0 and $(f - g)(x) = o((x - x_0)^n)$ as $x \rightarrow x_0$, then we say that f and g have *n*th order contact at x_0 , or more precisely, *contact of order at least n*.

The following visual shows approximations g of an arbitrary function f that have 0th (left), 1st (middle), and 2nd (right) order contact at x_0 .



5.2.4 The Real Tangent Space

Definition 5.5 (Tangent Space). Given function $f : E \rightarrow \mathbb{R}$ and a point $x_0 \in E$, the increment of the argument $h = x - x_0$ can be regarded as a vector attached to the point x_0 and defining the transition from x_0 to $x_0 + h$. h is called a *tangent vector*, and the set of all such vectors as $T_{x_0}\mathbb{R}$. Similarly, we denote $T_{y_0}\mathbb{R}$ the set of all displacement vectors from the point y_0 along the y -axis.



Then, we can see that the differential is a mapping

$$df(x_0) : T_{x_0}\mathbb{R} \longrightarrow T_{f(x_0)}\mathbb{R}$$

Note that there are two functions to pay attention to here:

1. The true increment of f , defined $h \mapsto f(x_0 + h) - f(x_0) = \Delta f(x_0; h)$ (labeled in green).
2. The differential $h \mapsto f'(x_0)h = df(x_0)(h)$, which gives the increment of the tangent to the graph for increment h in the argument (labeled in red).

Example 5.1. Let $f(x) = \sin x$. Then we will show that $f'(x) = \cos x$.

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin(x)}{h} &= \lim_{h \rightarrow 0} \frac{2 \sin\left(\frac{h}{2}\right) \cos\left(x + \frac{h}{2}\right)}{h} \\ &= \lim_{h \rightarrow 0} \cos\left(x + \frac{h}{2}\right) \cdot \lim_{h \rightarrow 0} \frac{\sin\left(\frac{h}{2}\right)}{\left(\frac{h}{2}\right)} = \cos(x) \end{aligned}$$

Here, we have used the theorem on the limit of a product, the continuity of the function $\cos(x)$, the equivalence $\sin t \sim t$ as $t \rightarrow 0$, and the theorem on the limit of a composite function.

Example 5.2. We will show that $\cos'(x) = -\sin(x)$.

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\cos(x+h) - \cos(x)}{h} &= \lim_{h \rightarrow 0} \frac{-2 \sin\left(\frac{h}{2}\right) \sin\left(x + \frac{h}{2}\right)}{h} \\ &= -\lim_{h \rightarrow 0} \sin\left(x + \frac{h}{2}\right) \cdot \lim_{h \rightarrow 0} \frac{\sin\left(\frac{h}{2}\right)}{\left(\frac{h}{2}\right)} = -\sin(x) \end{aligned}$$

5.3 Rules of Differentiation over \mathbb{R}

5.3.1 Basic Properties; Derivatives of Composite, Inverse Functions

Theorem 5.1 (Arithmetic Properties of Differentiation over \mathbb{R}). If functions $f, g : E \rightarrow \mathbb{R}$ are differentiable at a point $x \in E$, then

1. their sum is differentiable at x , and

$$d(f+g)(x) = df(x) + dg(x) \iff (f+g)'(x) = (f'+g')(x)$$

2. their product is differentiable at x , and

$$d(f \cdot g)(x) = g(x)df(x) + f(x)dg(x) \iff (f \cdot g)'(x) = f'(x) \cdot g(x) + f(x) \cdot g'(x)$$

3. their quotient is differentiable at x if $g(x) \neq 0$, and

$$d\left(\frac{f}{g}\right)(x) = \frac{g(x)df(x) - f(x)dg(x)}{g^2(x)} \iff \left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$

It is clear that $c \cdot df(x) = d(cf)(x)$, it is clear that the derivative is a linear operator from the space of all functions differentiable at x_0 the space of all functions.

Proof. Since f and g are differentiable at x , there exists the differential $df(x)(h) = f'(x)h$ and $dg(x) = g'(x)h$ where

$$\begin{aligned} f(x+h) &= f(x) + df(x)(h) + o(h) = f(x) + f'(x)h + o(h) \\ g(x+h) &= g(x) + dg(x)(h) + o(h) = g(x) + g'(x)h + o(h) \end{aligned}$$

From this relation, we can clearly see that a certain property of the differential automatically implies the same property of the derivative. (Remember that $f'(x)$ and $g'(x)$ are not functions! They are scalars defined on fixed point x .)

1. Even though this derivation may be a bit long, every step is included to minimize ambiguity.

$$\begin{aligned} (f+g)(x+h) - (f+g)(x) &= (f(x+h) + g(x+h)) - (f(x) + g(x)) \\ &= (f(x+h) - f(x)) + (g(x+h) - g(x)) \\ &= (df(x)(h) + o(h)) + (dg(x)(h) + o(h)) \\ &= (f'(x)h + o(h)) + (g'(x)h + o(h)) \\ &= (f'(x) + g'(x))h + o(h) \\ &= (f' + g')(x)h + o(h) \\ &= d(f+g)(x)h + o(h) \end{aligned}$$

2. For the product rule, we have

$$\begin{aligned}(f \cdot g)(x+h) - (f \cdot g)(x) &= f(x+h)g(x+h) - f(x)g(x) \\ &= (f(x) + df(x)(h) + o(h))(g(x) + dg(x)(h) + o(h)) - f(x)g(x) \\ &= (f(x) + f'(x)h + o(h))(g(x) + g'(x)h + o(h)) - f(x)g(x)\end{aligned}$$

Expanding this gives

$$(f'(x)g(x) + f(x)g'(x))h + (f(x) + g(x))o(h) + f'(x)g'(x)h^2 + (f'(x) + g'(x))ho(h) + (o(h))^2$$

but note that since $f(x), g(x), f'(x), g'(x)$ are constants, we see that

(a) $(f(x) + g(x))o(h) = o(h)$ because

$$\lim_{h \rightarrow 0} \frac{(f(x) + g(x))o(h)}{h} = (f(x) + g(x)) \lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$$

(b) $f'(x)g'(x)h^2 = o(h)$ since

$$\lim_{h \rightarrow 0} \frac{f'(x)g'(x)h^2}{h} = f'(x)g'(x) \lim_{h \rightarrow 0} h = 0$$

(c) $(f'(x) + g'(x))ho(h) = o(h)$ because

$$\lim_{h \rightarrow 0} \frac{(f'(x) + g'(x))ho(h)}{h} = (f'(x) + g'(x)) \lim_{h \rightarrow 0} o(h) = 0$$

In fact, this term is of $o(h^2)$.

(d) We can see that $(o(h))^2 = o(h)$ since

$$\lim_{h \rightarrow 0} \frac{(o(h))^2}{h} = \lim_{h \rightarrow 0} \frac{o(h)}{h} \cdot \lim_{h \rightarrow 0} o(h) = 0 \cdot 0 = 0$$

In fact, $(o(h))^2 = o(h^2)$.

Therefore, the above simplifies to

$$(f \cdot g)(x+h) - (f \cdot g)(x) = (f'(x)g(x) + f(x)g'(x))h + o(h)$$

But this means that the differential (best approximation) $d(f \cdot g)(x)$ must be

$$(f \cdot g)'(x)(h) = (f \cdot g)'(x)h = (f'(x)g(x) + f(x)g'(x))h$$

3. Since the function $g(x) \neq 0$ at point x , then by continuity we can assume that there exists a neighborhood $U(x)$ where the image of that neighborhood does not vanish. That is, we can guarantee that $g(x+h) \neq 0$ for sufficiently small values of h . We assume h is small in the following computations.

$$\begin{aligned}\left(\frac{f}{g}\right)(x+h) - \left(\frac{f}{g}\right)(x) &= \frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)} \\ &= \frac{1}{g(x)g(x+h)} (f(x+h)g(x) - f(x)g(x+h)) \\ &= \left(\frac{1}{g^2(x)} + o(1)\right) \left((f(x) + f'(x)h + o(h))g(x) \right. \\ &\quad \left. - f(x)(g(x) + g'(x)h + o(h)) \right) \\ &= \left(\frac{1}{g^2(x)} + o(1)\right) ((f'(x)g(x) - f(x)g'(x))h + o(h)) \\ &= \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)} h + o(h)\end{aligned}$$

Note that here we have used the continuity of g at the point x and the fact that $g(x) \neq 0$ to deduce that

$$\lim_{h \rightarrow 0} \frac{1}{g(x)g(x+h)} = \frac{1}{g^2(x)} \iff \frac{1}{g(x)+g(x+h)} = \frac{1}{g^2(x)} + o(1)$$

where $o(1)$ is infinitesimal as $h \rightarrow 0$. ■

Theorem 5.2 (Chain Rule for Composite Functions over \mathbb{R}). *Let there be functions $f : E_1 \subset \mathbb{R} \rightarrow E_2 \subset \mathbb{R}$ is differentiable at a point $x \in E_1$ and the function $g : E_2 \subset \mathbb{R} \rightarrow \mathbb{R}$ is differentiable at point $y = f(x) \in E_2$, with respective differentials*

$$\begin{aligned} df(x) &: T_x \mathbb{R} \rightarrow T_y \mathbb{R} \\ dg(y) &: T_y \mathbb{R} \rightarrow T_{g(y)} \mathbb{R} \end{aligned}$$

Then the composite function $g \circ f : E_1 \rightarrow \mathbb{R}$ is differentiable at x , and $d(g \circ f)(x) : T_x \mathbb{R} \rightarrow T_{g \circ f(x)} \mathbb{R}$ is

$$d(g \circ f)(x) = dg(y) \circ df(x) \iff (g \circ f)'(x) = g'(f(x)) \circ f'(x)$$

Proof. We will denote the increment of the argument with the variables h and t . Then, by differentiability of f and g , we have

$$\begin{aligned} f(x+h) - f(x) &= f'(x)h + o(h) \text{ as } h \rightarrow 0 \\ g(y+t) - g(y) &= g'(y)t + o(t) \text{ as } t \rightarrow 0 \end{aligned}$$

Since the function $o(t)$ can be represented as $o(t) = \gamma(t)t$, where $\gamma = o(1)$ and hence is infinitesimal as $t \rightarrow 0$, meaning that we can assume $\gamma(0) = 0$ (since $o(t)$ is defined for $t = 0$).

We can think of the displacement of x as like a chain reaction: As $x \mapsto x+h$, $f(x) \mapsto f(x+h)$, which we could interpret as $y \mapsto y+t$ and hence means that $g(y) \mapsto g(y+t)$. So, setting $f(x) = y$ and $f(x+h) = y+t$, by differentiability and hence continuity of f at point x , we can conclude that $t \rightarrow 0$ as $h \rightarrow 0$. So, we have

$$\gamma(f(x+h) - f(x)) = \gamma((y+t) - y) = \gamma(t) = \alpha(h) \rightarrow 0 \text{ as } h \rightarrow 0$$

Thus, we get

$$\begin{aligned} o(t) &= \gamma(t)t = \gamma(f(x+h) - f(x))(f(x+h) - f(x)) \\ &= \alpha(h)(f'(x)h + o(h)) \\ &= \alpha(h)f'(x)h + \alpha(h)o(h) \\ &= o(h) + o(h) = o(h) \text{ as } h \rightarrow 0 \\ (g \circ f)(x+h) - (g \circ f)(x) &= g(f(x+h)) - g(f(x)) \\ &= g(y+t) - g(y) \\ &= g'(y)t + o(t) \\ &= g'(f(x))(f(x+h) - f(x)) + o(f(x+h) - f(x)) \\ &= g'(f(x))(f'(x)h + o(h)) + o(f(x+h) - f(x)) \\ &= g'(f(x))(f'(x)h) + g'(f(x))(o(h)) + o(f(x+h) - f(x)) \end{aligned}$$

Since $g'(f(x))(o(h))$ is really just a constant multiplied by a function that is $o(h)$, it is $o(h)$. $o(f(x+h) - f(x))$. As for $o(f(x+h) - f(x))$, we see that since $f(x+h) - f(x) = t$, a function that is $o(f(x+h) - f(x))$ becomes infinitesimal compared to t as $t \rightarrow 0$. As already stated before, we have

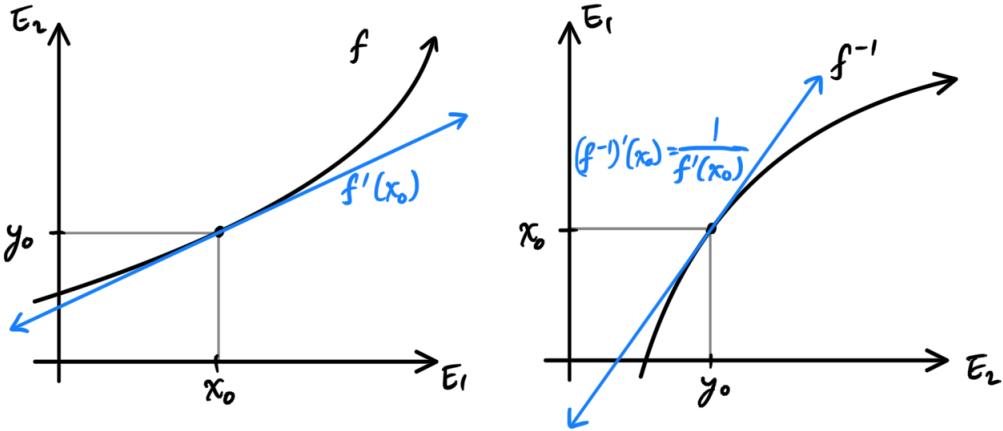
$$o(f(x+h) - f(x)) = o(h) \text{ as } h \rightarrow 0$$

and thus, we proved that

$$\begin{aligned} (g \circ f)(x+h) - (g \circ f)(x) &= g'(y)f'(x)h + o(h) \\ &= (dg(y) \circ df(x))(h) + o(h) \end{aligned}$$
■

Theorem 5.3 (Differentiation of Inverse Functions over \mathbb{R}). Let $E_1, E_2 \subset \mathbb{R}$, and $f : E_1 \rightarrow E_2$ and $f^{-1} : E_2 \rightarrow E_1$ be mutually inverse and continuous at points $x_0 \in E_1$ and $f(x_0) = y_0 \in E_2$. If f is differentiable at x_0 and $f'(x_0) \neq 0$, then f^{-1} also differentiable at the point y_0 , and

$$(f^{-1})^{-1}(y_0) = (f'(x_0))^{-1} \iff df^{-1}(y_0) = (df(x_0))^{-1}$$



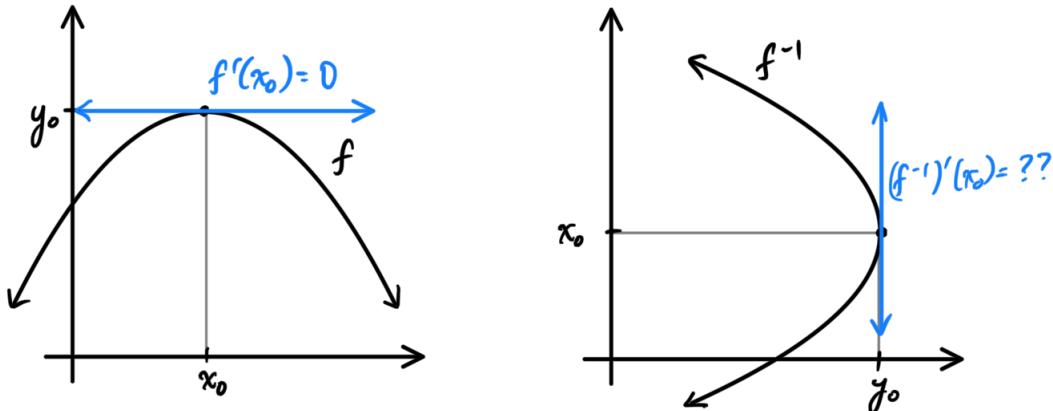
Note that if we knew in advance that f^{-1} was differentiable at y_0 (which is a stronger hypothesis), we can find immediately by the identity

$$(f^{-1} \circ f)(x) = x$$

and the theorem on the differentiation of a composite function that

$$(f^{-1})'(y_0) \cdot f'(x_0) = 1$$

Note that if the hypothesis was satisfied, but $f'(x_0) = 0$, then f^{-1} would not be differentiable since it would have an undefined differential.



5.3.2 Higher-Order Derivatives

Definition 5.6 (Global Derivative Function). If function $f : E \rightarrow \mathbb{R}$ is differentiable at every point $x \in E$, then a new function $f' : E \rightarrow \mathbb{R}$, whose value at a point $x \in E$ equals the derivative $f'(x)$ of the function f at the point.

Definition 5.7 (Second, Nth Derivative). This function f' may itself have a derivative $(f')' : E \rightarrow \mathbb{R}$, called the *second derivative* of the original function f , denoted

$$f''(x), \quad \frac{d^2 f(x)}{dx^2}$$

By induction, if the derivative $f^{(n-1)}(x)$ of order $n-1$ of f has been defined, then the *derivative of order n* is defined by the formula

$$f^{(n)}(x) \equiv (f^{(n-1)})'(x)$$

The set of function $f : E \rightarrow \mathbb{R}$ having continuous derivatives up to order n inclusive is denoted $C^n(E, \mathbb{R})$.

Lemma 5.4 (Leibniz' Formula). *Let $u(x)$ and $v(x)$ be functions having derivatives up to order n inclusive on a common set E . Then,*

$$(uv)^{(n)} = \sum_{m=0}^n \binom{n}{m} u^{(n-m)} v^{(m)}$$

This means that given a polynomial $P_n(x) = c_0 + c_1(x - x_0) + \dots + c_n(x - x_0)^n$, then

$$\begin{aligned} P_n(x_0) &= 0 \\ P'_n(x_0) &= 1!c_1 \\ P''_n(x_0) &= 2!c_2 \\ &\dots = \dots \\ P_n^{(n)}(x_0) &= n!c_n \\ P_n^{(k)}(x_0) &= 0 \text{ for } k > n \end{aligned}$$

and thus the polynomial $P_n(x)$ can be written as

$$P_n(x) = P_n^{(0)}(x_0) + \frac{1}{1!} P_n^{(1)}(x_0)(x - x_0) + \frac{1}{2!} P_n^{(2)}(x_0)(x - x_0)^2 + \dots + \frac{1}{n!} P_n^{(n)}(x_0)(x - x_0)^n$$

5.4 Theorems of Differential Calculus

5.4.1 Fermat's Lemma, Rolle's Theorem

Definition 5.8 (Local Extrema). A point $x_0 \in E \subset \mathbb{R}$ is called a *local maximum* (resp. *local minimum*) and the value of a function $f : E \rightarrow \mathbb{R}$ at that point a *local maximum value* (resp. *local minimum value*) if there exists a neighborhood $U_E(x_0)$ of x_0 in E such that at any point $x \in U_E(x_0)$ we have

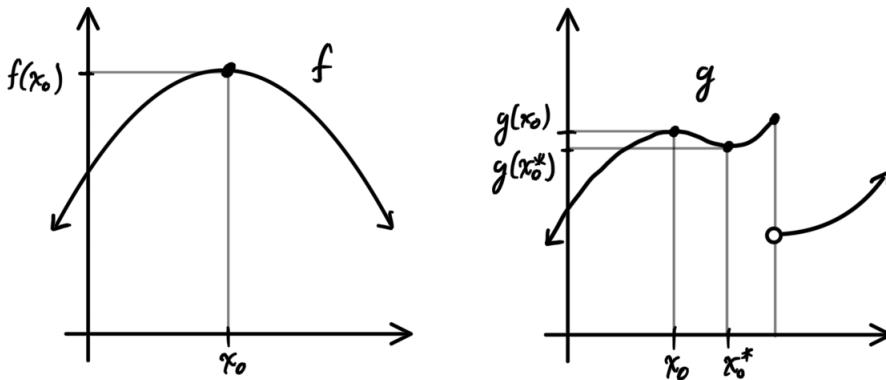
$$f(x) \leq f(x_0), \text{ (resp. } f(x) \geq f(x_0))$$

If this is a strict inequality

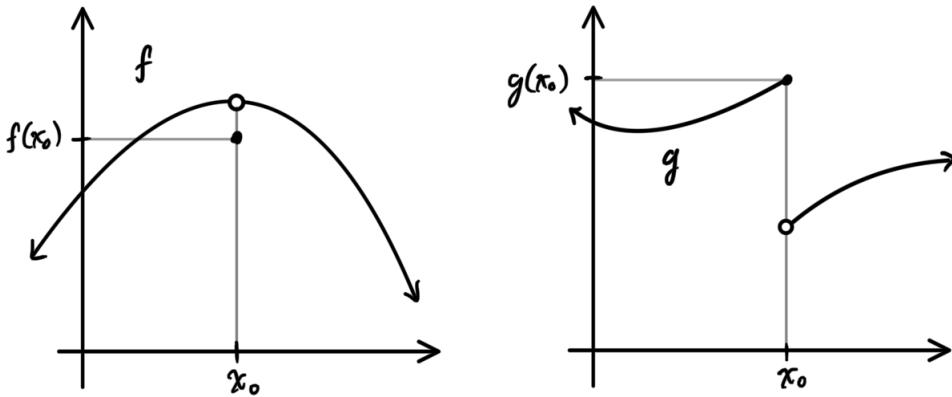
$$f(x) < f(x_0), \text{ (resp. } f(x) > f(x_0))$$

then x_0 is called a *strict local maximum* (resp. *strict local minimum*).

Definition 5.9 (Interior Extrema). An extremum $x_0 \in E$ of the function $f : E \rightarrow \mathbb{R}$ is called an *interior extremum* if x_0 is not on the boundary of E , or more rigorously, x_0 is a limit point of both sets $E_- = \{x \in E \mid x < x_0\}$ and $E_+ = \{x \in E \mid x > x_0\}$. For example, the graphs below are interior extrema at x_0, x_0^* .



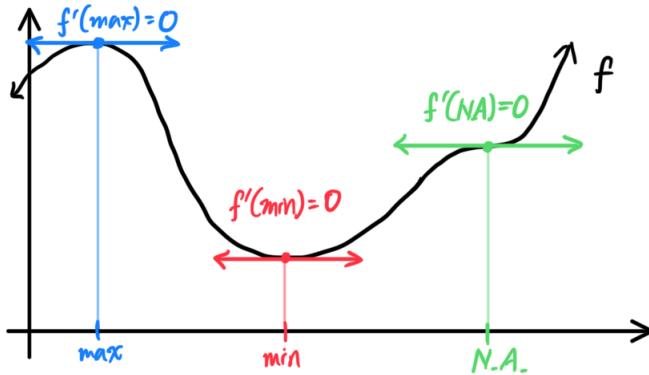
But the following graphs show extrema (at x_0) that are not interior extrema.



Lemma 5.5 (Fermat). If a function $f : E \rightarrow \mathbb{R}$ is differentiable at an interior extremum $x_0 \in E$, then its derivative at x_0 is 0. That is,

$$f'(x_0) = 0$$

Thus, this lemma gives a necessary condition for an interior extremum of a differentiable function. But for non-interior extrema, it is generally not true that $f'(x_0) = 0$ and so the converse does not hold (labeled in green).



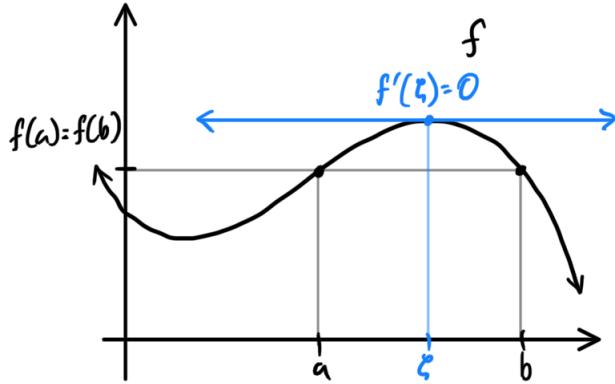
Proof. By definition of differentiability at x_0 we get

$$\begin{aligned} f(x_0 + h) - f(x_0) &= f'(x_0)h + o(h) \\ &= f'(x_0)h + \alpha(x_0; h)h \\ &= (f'(x_0) + \alpha(x_0; h))h \end{aligned}$$

where we know that $o(h)$ can be written as $o(1)h$ for some infinitesimal $o(1)$ as $h \rightarrow 0$. If $f'(x_0) \neq 0$, then for h sufficiently close to 0 the quantity $f'(x_0) + \alpha(x_0; h)$ would have the same sign as $f'(x_0)$, since $\alpha(x_0; h) \rightarrow 0$ as $h \rightarrow 0$. But the value of h can be both positive or negative, given that x_0 is an interior extremum. This contradiction must imply that $f'(x_0) = 0$. ■

Geometrically, Fermat's lemma is obvious, since it asserts that at an extremum of a differentiable function, the tangent to its graph is horizontal. Physically, this lemma means that in motion along a line the velocity must be zero at the instant then the direction reverses.

Theorem 5.6 (Rolle's Theorem). If a function $f : [a, b] \rightarrow \mathbb{R}$ is continuous on a closed interval $[a, b]$ and differentiable on the open interval (a, b) and $f(a) = f(b)$, then there exists a point $\zeta \in (a, b)$ such that $f'(\zeta) = 0$.



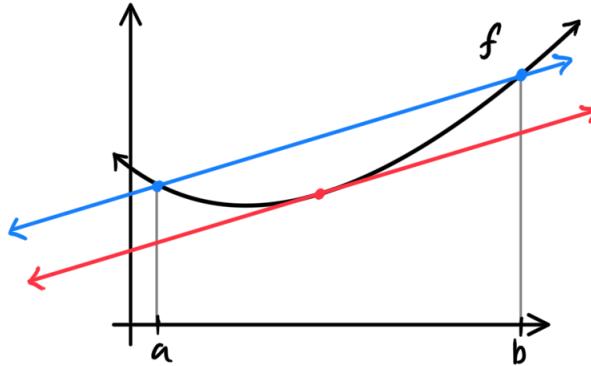
5.4.2 Mean Value Theorem, Cauchy's Finite-Increment Theorem

The following theorem is extremely useful in studying numerical valued functions.

Theorem 5.7 (Mean Value Theorem). *If a function $f : [a, b] \rightarrow \mathbb{R}$ is continuous on a closed interval $[a, b]$ and differentiable on the open interval (a, b) , there exists a point $\zeta \in (a, b)$ such that*

$$f'(\zeta) = \frac{f(b) - f(a)}{b - a} \iff f(b) - f(a) = f'(\zeta)(b - a)$$

Geometrically, this means that there exists a tangent line somewhere at $\zeta \in (a, b)$ that is parallel the secant line connecting the two points $(a, f(a))$ and $(b, f(b))$.



Some remarks:

1. Physically, if x is interpreted as time and $f(b) - f(a)$ as the amount of displacement over the time $b - a$ of a particle moving along the line, this theorem says that the velocity $f'(x)$ of the particle at some time $\zeta \in (a, b)$ is such that if the particle had moved with constant velocity $f'(\zeta)$ over the whole time interval, it would have been displaced by the same amount $f(b) - f(a)$. We call $f'(\zeta)$ the *average velocity* over the time interval $[a, b]$.
2. Note that the Mean Value Theorem is important in that it connects the increment of a function over a finite interval with the derivative of the function on that interval. Up to now, we have characterized only the local (infinitesimal) increment of a function in terms of the derivative or differential at a given point. MVT connects the increment of a function over a *finite* interval with the derivative of the function.

The MVT actually leads to multiple useful corollaries.

Corollary 5.7.1 (Derivative of a Monotonic Function). *If the derivative of a function is nonnegative (resp. positive) at every point of an open interval, then the function is nondecreasing (resp. increasing) on that interval.*

Proof. If $x_1 < x_2$ are two points of the interval, then the MVT

$$f(x_2) - f(x_1) = f'(\zeta)(x_2 - x_1)$$

shows that the sign of the left hand side must equal that of the right. \blacksquare

Corollary 5.7.2 (Derivative of a Constant Function). *A function that is continuous on a closed interval $[a, b]$ is constant on it if and only if its derivative equals 0 at every point of the interval $[a, b]$ or the open interval (a, b) .*

Therefore, if the derivatives $f'_1(x)$ and $f'_2(x)$ of two functions $f_1(x)$ and $f_2(x)$ are equal on some interval (that is, $f'_1(x) = f'_2(x)$ on the interval), then the difference

$$(f_1 - f_2)(x) = f_1(x) - f_2(x)$$

is constant.

Proof. Given constant function f , the MVT equation

$$0 = f(x_2) - f(x_1) = f'(\zeta)(x_2 - x_1)$$

implies that $f'(\zeta) = 0$ for all $x_1, x_2 \in E$. It follows that by the arithmetic properties of the derivative, given two functions f_1, f_2 with the same derivative on an interval, the derivative of their difference $(f_1 - f_2)' = 0$, and therefore must be constant on that interval. \blacksquare

The following proposition is a useful generalization of Lagrange's theorem.

Theorem 5.8 (Cauchy's Finite-Increment Theorem). *Let $x = x(t)$ and $y = y(t)$ be functions that are continuous on a closed interval $[\alpha, \beta]$ and differentiable on the open interval (α, β) . Then, there exists a point $\tau \in [\alpha, \beta]$ such that*

$$x'(\tau)(y(\beta) - y(\alpha)) = y'(\tau)(x(\beta) - x(\alpha))$$

If in addition $x'(t) \neq 0$ for each $t \in (\alpha, \beta)$, then $x(\alpha) \neq x(\beta)$ and we have the equality

$$\frac{y(\beta) - y(\alpha)}{x(\beta) - x(\alpha)} = \frac{y'(\tau)}{x'(\tau)}$$

5.4.3 Taylor's Formula

From the following results one may deduce that the more derivatives of two functions coincide (including the derivative of the 0th order) at a point, the better these functions approximate each other in a neighborhood of that point. Using Leibniz's rule, approximations up to a certain degree at a point can be expressed as a polynomial

$$P_n(x_0; x) = P_n(x_0) + \frac{P'_n(x_0)}{1!}(x - x_0) + \dots + \frac{P_n^{(n)}(x_0)}{n!}(x - x_0)^n$$

where each coefficient of the polynomial

Definition 5.10 (Taylor Polynomial). If a function $f : E \rightarrow \mathbb{R}$ has derivatives of all orders $n \in \mathbb{N}$ at a point x_0 , the unique series

$$P_n(x_0; x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

is the *Taylor polynomial of order n of $f(x)$ at x_0* . We can see that the derivatives of f and P_n coincide up to order n .

Definition 5.11 (Analytic Functions). We cannot assume that the Taylor series of an infinitely differentiable function converges to the function f within a neighborhood $U(x_0)$, nor can we assume that it converges at all! These types of "nice" functions that have a Taylor approximation within the neighborhood of x_0 are called *analytic functions* and can be written in the form

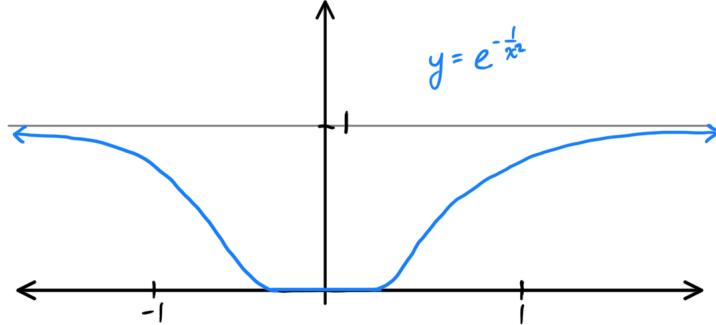
$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + r_n(x_0; x)$$

where r is called the *remainder term*.

Example 5.3 (Infinitely Differentiable, Non-Analytic Function). A example of a non-analytic function is

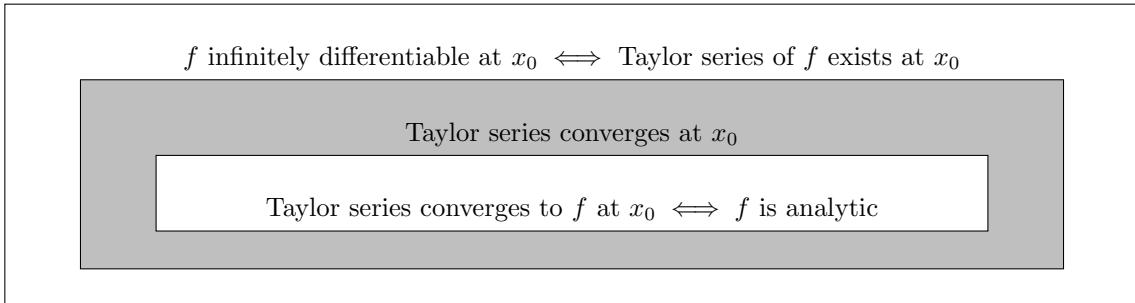
$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

which looks like



One can verify that the derivative $f^{(k)}(0) = 0$ for all k and hence the Taylor series is identically equal to 0, while $f(x) \neq 0$ if $x \neq 0$.

The relationship between these different conditions is nicely summarized in the figure.



The following lemma proves why Taylor Polynomials are considered a "good" approximations to analytic functions.

Lemma 5.9 (Infinitesimality of Functions with Vanishing Derivative up to Order n). Given a function $\varphi : E \rightarrow \mathbb{R}$ defined on a closed interval E with endpoint x_0 , let its derivatives vanish up to order n at x_0 . That is

$$\varphi(x_0) = \varphi'(x_0) = \dots = \varphi^{(n)}(x_0) = 0$$

Then, $\varphi = o((x - x_0)^n)$ as $x \rightarrow x_0$.

Proof. We prove by induction. For $n = 1$, the definition of differentiability states that

$$\varphi(x) = \varphi(x_0) + \varphi'(x - x_0) + o(x - x_0) \text{ as } x \rightarrow x_0$$

and so we have proved that

$$\varphi(x_0) = \varphi'(x_0) = 0 \implies \varphi(x) = o(x - x_0) \text{ as } x \rightarrow x_0$$

Now, suppose this assertion has been proved for order $n = k - 1 \geq 1$. That is, we have shown that

$$\varphi(x_0) = \dots = \varphi^{(k-1)}(x_0) = 0 \implies \varphi = o((x - x_0)^{k-1}) \text{ as } x \rightarrow x_0$$

Then we must show that this is valid for order $n = k \geq 2$. Assume that

$$\varphi(x_0) = \varphi'(x_0) = \dots = \varphi^{(k)}(x_0) = 0$$

We can see that this is equivalent to

$$(\varphi')'(x_0) = (\varphi')^{(2)}(x_0) = \dots = (\varphi')^{(k-1)}(x_0) = 0$$

and therefore by the induction assumption, we have

$$\varphi' = o((x - x_0)^{k-1}) \text{ as } x \rightarrow x_0$$

which means that we can put it in form

$$\varphi(x) = \alpha(x)(x - x_0)^{k-1} \text{ so that } \lim_{x \rightarrow x_0} \varphi(x) = \lim_{x \rightarrow x_0} \alpha(x) = 0$$

From the mean value theorem and substituting what we have above, we get

$$\begin{aligned} \varphi(x) &= \varphi(x) - \varphi(x_0) = \varphi'(\zeta)(x - x_0) \\ &= \varphi(\zeta)(\zeta - x_0)^{k-1}(x - x_0) \end{aligned}$$

where $\zeta \in (x_0, x)$. However, this implies that $|\zeta - x_0| < |x - x_0|$, and thus, as $x \rightarrow x_0$, $\zeta \rightarrow x_0$, which then makes $\alpha(\zeta) \rightarrow 0$. Since

$$|\varphi(x)| \leq |\alpha(\zeta)||x - x_0|^{k-1}|x - x_0| = |\alpha(\zeta)||x - x_0|^k$$

This means that $\varphi(x)$ is bounded by function $|\alpha(\zeta)||x - x_0|^k$, which is $o((x - x_0)^k)$, and so

$$\varphi = o((x - x_0)^k) \text{ as } x \rightarrow x_0$$

By induction, this works for all orders n . ■

Theorem 5.10 (Peano's Form of the Remainder). *Given analytic function $f : E \rightarrow \mathbb{R}$, a point $x_0 \in E$, and its n th order Taylor polynomial $P_n(x_0; x)$ around x_0 , P_n is a "good" approximation of f in the fact that its error term is $o((x - x_0)^n)$. That is,*

$$f(x) = P_n(x_0; x) + o((x - x_0)^n) \text{ as } x \rightarrow x_0$$

This equation where $r_n(x; x_0) = o((x - x_0)^n)$ is called the Peano's form of the remainder.

Proof. Since the Taylor polynomial $P_n(x_0; x)$ is constructed from the requirement that its derivatives up to order n inclusive must coincide with the corresponding derivatives of f at x_0 , it follows that

$$r_n(x_0; x_0) \equiv f^{(k)}(x_0) - P_n^{(k)}(x_0; x_0) = 0 \text{ for } k = 0, 1, \dots, n$$

Using the previous lemma, a this means that $r_n(x; x_0) = o((x - x_0)^n)$ as $x \rightarrow x_0$. ■

Theorem 5.11 (Lagrange Form of the Remainder). *If $f : E \rightarrow \mathbb{R}$ has derivatives of order $n+1$ on the open interval with endpoints x_0 and x , then*

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + r_n(x; x_0)$$

where

$$r_n(x; x_0) = \frac{f^{(n+1)}(\zeta)}{(n+1)!}(x - x_0)^{n+1}$$

This form is called Taylor's formula with the Lagrange form of the remainder. Furthermore, this form says that if $f^{(n+1)}(x)$ is bounded in a neighborhood of x_0 , it also implies the formula

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + O((x - x_0)^{n+1})$$

Therefore, we can use this boundedness of $f^{(n+1)}$ to find the maximum error bound

$$|r_n(x; x_0)|$$

of $P_n(x; x_0)$.

Proof. It is a direct result from the lemma. This is actually a generalization of the mean value theorem but for higher orders. ■

Corollary 5.11.1 (Table of Asymptotic Formulas for Elementary Functions). *We write the Maclaurin series (Taylor series around $x = 0$) for elementary functions. Note that these error terms are $O(x^{n+1})$ (bounded compared to x^{n+1}) and $o(x^n)$ (infinitesimal compared to x^n).*

$$\begin{aligned} e^x &= 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \dots + \frac{1}{n!}x^n + O(x^{n+1}) \\ \cos x &= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \dots + \frac{(-1)^n}{(2n)!}x^{2n} + O(x^{2n+2}) \\ \sin x &= x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots + \frac{(-1)^n}{(2n+1)!}x^{2n+1} + O(x^{2n+3}) \\ \cosh x &= 1 + \frac{1}{2!}x^2 + \frac{1}{4!}x^4 + \dots + \frac{1}{(2n)!}x^{2n} + O(x^{2n+2}) \\ \sinh x &= x + \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \dots + \frac{1}{(2n+1)!}x^{2n+1} + O(x^{2n+3}) \\ \ln(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \dots + \frac{(-1)^n}{n}x^n + O(x^{n+1}) \\ (1+x)^\alpha &= 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \dots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}x^n + O(x^{n+1}) \end{aligned}$$

5.5 The Study of Functions using Differential Calculus

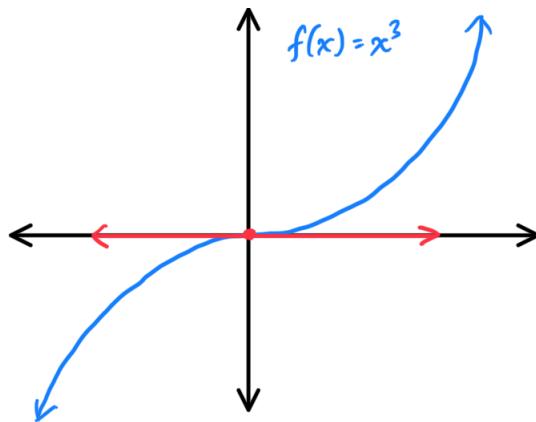
5.5.1 Conditions for Monotonicity of Functions

We can now connect the concepts of derivatives and monotonicity.

Theorem 5.12 (Derivative \Rightarrow Monotonicity). *Given function $f : E \rightarrow \mathbb{R}$ that is differentiable on an open interval $(a, b) = E$,*

$$\begin{aligned} f'(x) > 0 &\implies f \text{ is increasing} \\ f'(x) \geq 0 &\iff f \text{ is nondecreasing} \\ f'(x) \equiv 0 &\iff f \text{ is constant} \\ f'(x) \leq 0 &\iff f \text{ is nonincreasing} \\ f'(x) < 0 &\implies f \text{ is decreasing} \end{aligned}$$

Note that if f is strictly increasing (resp. decreasing), we cannot determine that $f'(x) \geq 0$ (resp. $f'(x) \leq 0$). For example, take the function $f(x) = x^3$, which is strictly increasing, but has derivative $f'(0) = 0$ at $x = 0$.



It is clearly strictly increasing within a neighborhood $U(0)$, so we can see that

$$\begin{aligned} f \text{ is increasing} &\implies f'(x) \geq 0 \\ f \text{ is decreasing} &\implies f'(x) \leq 0 \end{aligned}$$

5.5.2 Conditions for Extrema of Functions

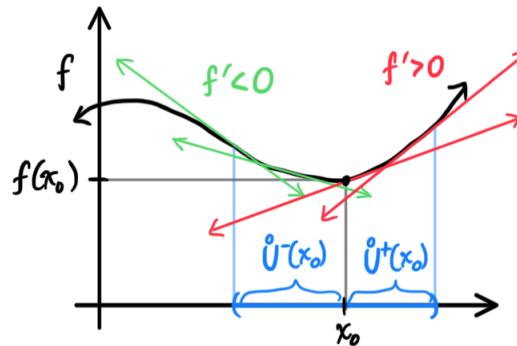
Similarly, we can connect the concepts of extrema and derivatives.

Theorem 5.13 (First Derivative Test). *Let function $f : E \rightarrow \mathbb{R}$ be defined in a neighborhood $U(x_0)$ of point x_0 , which is continuous at x_0 and differentiable in $\dot{U}(x_0)$, a deleted neighborhood of x_0 . (Note that this is broader hypothesis than just assuming that f be differentiable at x_0 .) Let*

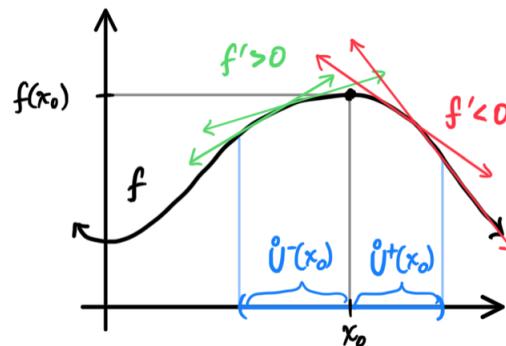
$$\dot{U}^-(x_0) \equiv \{x \in U(x_0) \mid x < x_0\}, \quad \dot{U}^+(x_0) \equiv \{x \in U(x_0) \mid x > x_0\}$$

That is, $\dot{U}^-(x_0)$ is the left portion of $U(x_0)$ and $\dot{U}^+(x_0)$ is the right portion of $U(x_0)$. Then,

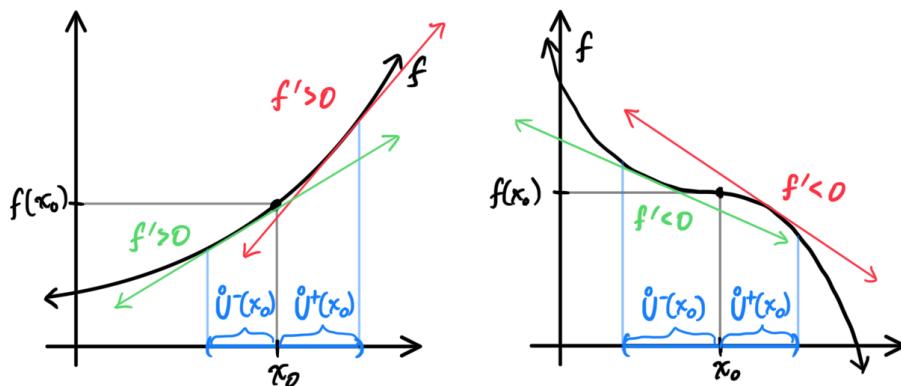
1. $(x_0, f(x_0))$ is strict local minimum if $f'(x) < 0$ in $\dot{U}^-(x_0)$ and $f'(x) > 0$ in $\dot{U}^+(x_0)$.



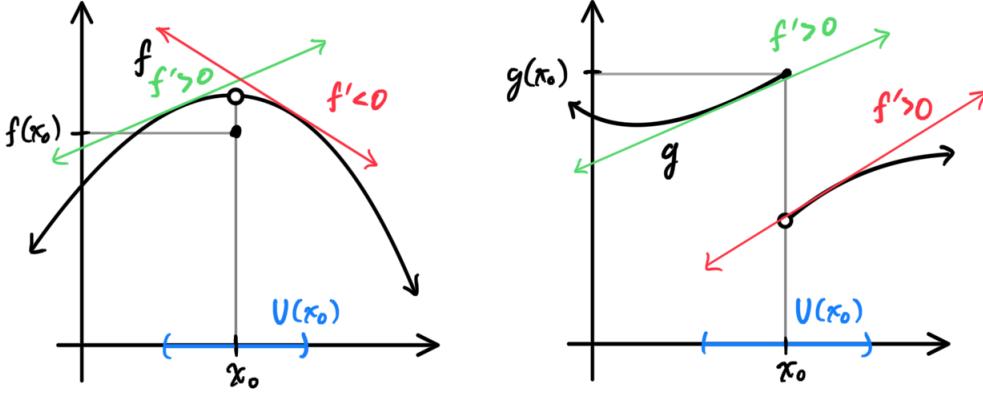
2. $(x_0, f(x_0))$ is strict local maximum if $f'(x) > 0$ in $\dot{U}^-(x_0)$ and $f'(x) < 0$ in $\dot{U}^+(x_0)$.



3. $(x_0, f(x_0))$ has no extremum at x_0 if $f'(x) > 0$ in both $\dot{U}^-(x_0), \dot{U}^+(x_0)$, or if $f'(x) < 0$ in both $\dot{U}^-(x_0), \dot{U}^+(x_0)$.



Note that if there is a discontinuity at a point x_0 , then this theorem does not apply. For example, $(x_0, f(x_0))$ in the graph below is a local minimum, even though the derivatives to the left of x_0 are positive and those to the right of x_0 are negative (within neighborhood $U(x_0)$). Similarly, $(x_0, g(x_0))$ is a local maximum, even though the derivative to the left and to the right of x_0 are both positive.



Proposition 5.14 (2nd, n th Derivative Test). *Let function $f : E \rightarrow \mathbb{R}$ be defined on a neighborhood $U(x_0)$ of x_0 has derivatives of order up to n inclusive at x_0 . If its derivatives up to the $(n-1)$ th order vanishes*

$$f'(x_0) = f''(x_0) = \dots = f^{(n-1)}(x_0) = 0$$

*but the n th derivative at x_0 does **not** vanish*

$$f^{(n)}(x_0) \neq 0$$

then

1. *n is odd \implies there is no local extremum at x_0*
2. *n is even \implies there is a local extremum at x_0*
 - (a) *$f^{(n)}(x_0) > 0 \implies$ it is a strict local minimum*
 - (b) *$f^{(n)}(x_0) < 0 \implies$ it is a strict local maximum*

5.5.3 Important Algebraic Inequalities

We also introduce various inequalities that may be useful for producing future results. The following lemmas can be proved with elementary algebra.

Lemma 5.15 (Young's Inequalities). *If $a > 0$ and $b > 0$, and the numbers p and p are such that $p \neq 0, 1$ and $q \neq 0, 1$, and $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\begin{aligned} a^{\frac{1}{p}} b^{\frac{1}{q}} &\leq \frac{1}{p}a + \frac{1}{q}b \text{ if } p > 1 \\ a^{\frac{1}{p}} b^{\frac{1}{q}} &\geq \frac{1}{p}a + \frac{1}{q}b \text{ if } p < 1 \end{aligned}$$

and equality holds in both cases if and only if $a = b$.

Lemma 5.16 (Holder's Inequalities). *Let $x_i \geq 0, y_i \geq 0$ for $i = 1, 2, \dots, n$, and let $\frac{1}{p} + \frac{1}{q} = 1$. Then,*

$$\begin{aligned} \sum_{i=1}^n x_i y_i &\leq \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n y_i^q \right)^{\frac{1}{q}} \text{ for } p > 1 \\ \sum_{i=1}^n x_i y_i &\geq \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n y_i^q \right)^{\frac{1}{q}} \text{ for } p < 1, p \neq 0 \end{aligned}$$

Lemma 5.17 (Minkowski's Inequalities). Let $x_i \geq 0, y_i \geq 0$ for $i = 1, 2, \dots, n$. Then,

$$\left(\sum_{i=1}^n (x_i + y_i)^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n y_i^p \right)^{\frac{1}{p}} \text{ when } p > 1$$

$$\left(\sum_{i=1}^n (x_i + y_i)^p \right)^{\frac{1}{p}} \geq \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^n y_i^p \right)^{\frac{1}{p}} \text{ when } p < 1, p \neq 0$$

5.5.4 Conditions for a Function to be Convex

Definition 5.12 (Convex, Concave Functions). A function $f : (a, b) \rightarrow \mathbb{R}$ defined on an open interval $(a, b) \subset \mathbb{R}$ is *convex* if the inequality

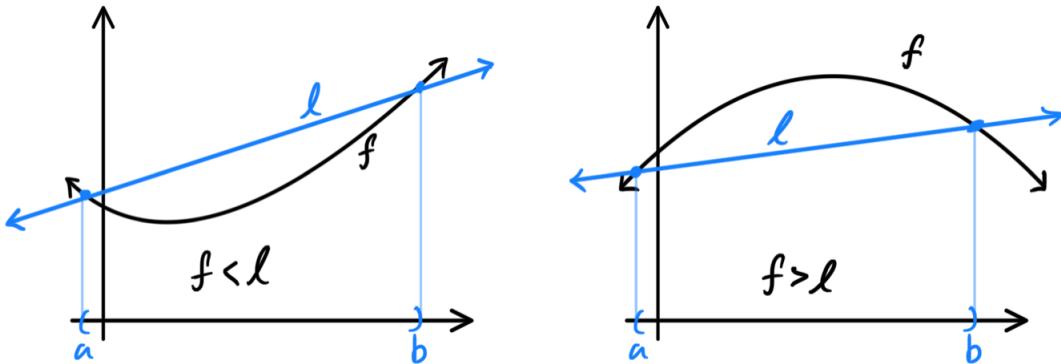
$$f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2)$$

holds and *concave*, or *convex upward*, if the inequality

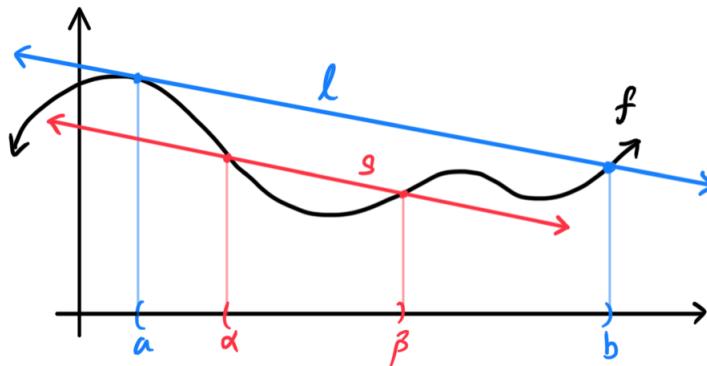
$$f(\alpha_1 x_1 + \alpha_2 x_2) \geq \alpha_1 f(x_1) + \alpha_2 f(x_2)$$

holds for all pairs of points $x_1, x_2 \in (a, b)$ and any numbers $\alpha_1, \alpha_2 \geq 0$ such that $\alpha_1 + \alpha_2 = 1$. If this inequality is strict whenever $x_1 \neq x_2$ and $\alpha_1 \alpha_2 \neq 0$, the function is said to be *strictly convex* and *strictly concave*, respectively.

Visually, this just means that given any two points a, b , the graph of a convex function (left) in (a, b) always lies underneath the secant line formed by the two points and the graph of a concave function (right) in (a, b) lies over the secant line formed by the two points.

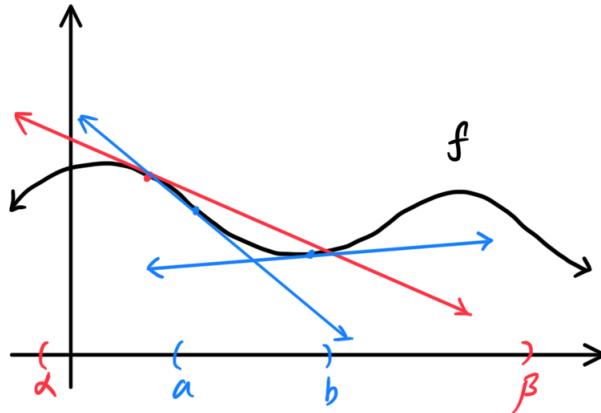


However, this is only a visual aid. In reality, it is actually not only the secant line formed by the two endpoints, but every pairs of points within that interval. For example, even though the secant line l formed by the endpoints a, b is above the whole graph in (a, b) , f is not convex over (a, b) since the secant line formed by points α, β do not lie completely underneath f .



The following is also another equivalent condition for a function to be convex over (a, b) .

Proposition 5.18. A function $f : (a, b) \rightarrow \mathbb{R}$ that is differentiable on the open interval (a, b) is convex on (a, b) if and only if its graph contains no points below any tangent drawn to it.



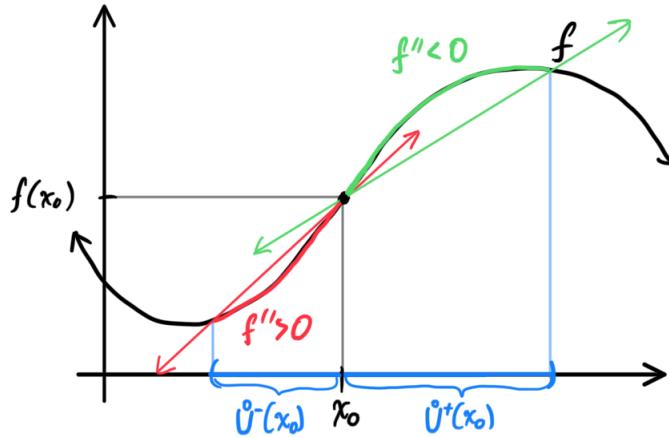
Theorem 5.19 (2nd Derivatives of Convex Functions). Given a function $f : (a, b) \rightarrow \mathbb{R}$ that is differentiable in its domain,

1. f is convex $\iff f'$ is nondecreasing on $(a, b) \iff f'' \geq 0$ on (a, b)
2. f is strictly convex $\iff f'$ is increasing on $(a, b) \iff f'' > 0$ on (a, b)
3. f is concave $\iff f'$ is nonincreasing on $(a, b) \iff f'' \leq 0$ on (a, b)
4. f is strictly concave $\iff f'$ is decreasing on $(a, b) \iff f'' < 0$ on (a, b)

Definition 5.13 (Inflection Point). Let $f : E \rightarrow \mathbb{R}$ be a function defined and differentiable on a neighborhood $U(x_0)$. If the function is convex downward (resp. upward) on the set $\overset{\circ}{U}^-(x_0) = \{x \in U(x_0) \mid x < x_0\}$ and convex upward (resp. downward) on $\overset{\circ}{U}^+(x_0) = \{x \in U(x_0) \mid x > x_0\}$, then the point

$$(x_0, f(x_0))$$

is called a *inflection point of the graph*.



Proposition 5.20 (Jensen's Inequality). If $f : (a, b) \rightarrow \mathbb{R}$ is a convex function, x_1, \dots, x_n are points of (a, b) , and $\alpha_1, \dots, \alpha_n$ are nonnegative numbers such that $\alpha_1 + \dots + \alpha_n = 1$, then

$$f(\alpha_1 x_1 + \dots + \alpha_n x_n) \leq \alpha_1 f(x_1) + \dots + \alpha_n f(x_n)$$

5.5.5 L'Hopital's Rule

Theorem 5.21 (L'Hopital's Rule). Let c be an extended real number (i.e. $c \in \mathbb{R} \cup \{+\infty, -\infty\}$) and let (a, b) be an open interval containing c (for a two-sided limit) or an open interval with endpoint c (for a one-sided limit, or a limit at infinity if c is infinite). Assume that f and g are assumed to be differentiable on $(a, b) \setminus c$, and additionally $g'(x) \neq 0$ on $(a, b) \setminus c$. If either

$$\lim_{x \rightarrow c} f(x) = \lim_{x \rightarrow c} g(x) = 0 \text{ or } \lim_{x \rightarrow c} |f(x)| = \lim_{x \rightarrow c} |g(x)| = \infty$$

then

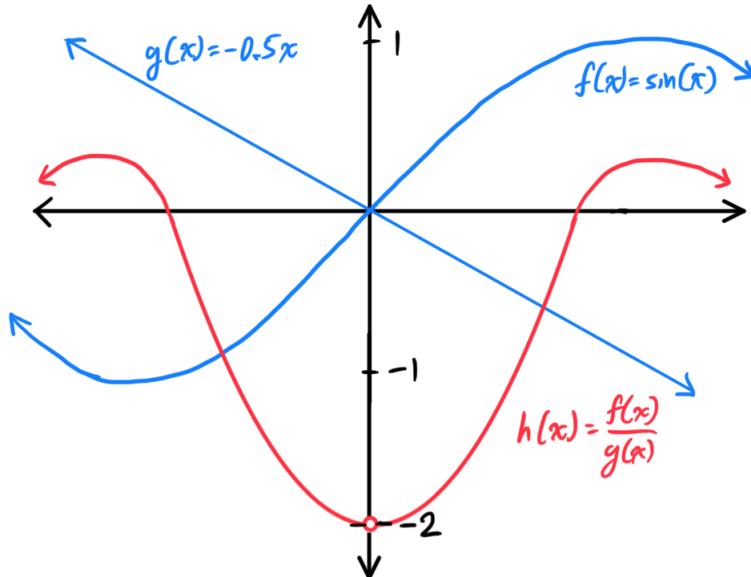
$$\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)}$$

L'Hopital's rule can be stated colloquially, but not quite accurately, as follows: The limit of a ratio of functions equals the limit of the ratio of their derivatives if their derivatives exist.

Example 5.4. Let $f(x) = \sin x$ and $g(x) = -0.5x$. Then, the function

$$h(x) = \frac{f(x)}{g(x)} = \frac{\sin x}{-0.5x}$$

is clearly undefined at $x = 0$.

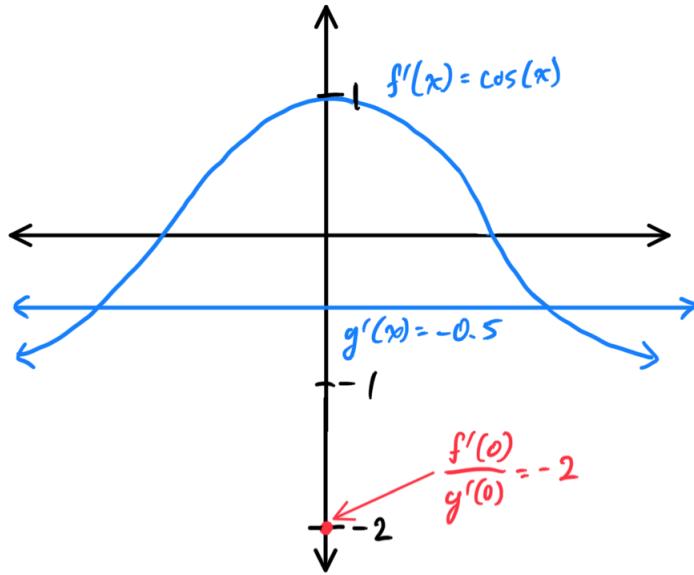


However, we can solve the limit using L'Hopital's rule to get

$$\lim_{x \rightarrow 0} \frac{\sin x}{-0.5x} = \lim_{x \rightarrow 0} \frac{\cos x}{-0.5} = -2$$

Therefore, $h : \mathbb{R} \setminus 0 \rightarrow \mathbb{R}$ can be completed to continuous function on all of \mathbb{R} by defining the extension:

$$H(x) \equiv \begin{cases} h(x), & x \neq 0 \\ -2, & x = 0 \end{cases}$$



5.6 Complex Analysis: An Introduction

Just as the equation $x^2 = 2$ has no solutions in the domain \mathbb{Q} of rational numbers, the equation $x^2 = -1$ has no solutions in the domain \mathbb{R} of real numbers. Just as we adjoin the symbol $\sqrt{2}$ as a solution of $x^2 = 2$ and connect it with rational numbers to get new numbers of the form

$$r_1 + r_2\sqrt{2}, \quad r_1, r_2 \in \mathbb{Q}$$

we introduce the symbol i as a solution of $x^2 = -1$ and attach this number to real numbers.

One feature of this enlargement of the field \mathbb{R} of real numbers into the resulting field \mathbb{C} of complex numbers, every algebraic equation with real or complex coefficients now has a solution.

5.6.1 Algebraic Extension of Field \mathbb{R}

We introduce the number i , called the *imaginary unit*, such that $i^2 = -1$. We may multiply real numbers y to i to get yi , and we can add real numbers to such numbers, to get numbers of the form

$$x + yi, \quad x, y \in \mathbb{R}$$

We then define all objects of the form $x + iy$ as the *complex numbers*, with addition defined

$$(x_1 + iy_1) + (x_2 + iy_2) \equiv (x_1 + x_2) + i(y_1 + y_2)$$

and multiplication defined

$$(x_1 + iy_1) \cdot (x_2 + iy_2) \equiv (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)$$

As expected, this makes $+$ and \cdot commutative operations. Furthermore, two complex numbers $z = x_1 + iy_1$ and $w = x_2 + iy_2$ are equal if and only if $x_1 = x_2$ and $y_1 = y_2$.

One nontrivial property of field \mathbb{C} is that every element $z \in \mathbb{C}$ has a multiplicative inverse z^{-1} . To find this, we must define the following.

Definition 5.14 (Complex Conjugate). Given complex number $z = x + iy$, its *complex conjugate* is

$$\bar{z} = \overline{x + iy} = x - iy$$

Note that

$$z \cdot \bar{z} = x^2 + y^2 \neq 0 \text{ iff } z \neq 0$$

Thus, given z ,

$$z^{-1} = \frac{1}{z \cdot \bar{z}} \cdot \bar{z} \iff (x + yi)^{-1} = \frac{x}{x^2 + y^2} - i \frac{y}{x^2 + y^2}$$

5.6.2 Geometric Interpretation of \mathbb{C}

Once the algebraic operations $+$ and \cdot has been introduced, the symbol i is no longer needed. That is, we can define a new set $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ with the operations $+_{\mathbb{R}}, \cdot_{\mathbb{R}} : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined

$$\begin{aligned}(x_1, y_1) +_{\mathbb{R}} (x_2, y_2) &\equiv (x_1 + x_2, y_1 + y_2) \\ (x_1, y_1) \cdot_{\mathbb{R}} (x_2, y_2) &\equiv (x_1 x_2 - y_1 y_2, x_1 y_2 + x_2 y_1)\end{aligned}$$

We can check that this new set $(\mathbb{R}^2, +_{\mathbb{R}}, \cdot_{\mathbb{R}})$ is isomorphic to $(\mathbb{C}, +, \cdot)$ as fields, and therefore one can identify complex numbers with vectors $z = (x, y)$ of the plane \mathbb{R}^2 , where $x = \operatorname{Re} z$ is called the *real part* and $y = \operatorname{Im} z$ is called the *imaginary part*.

Definition 5.15 (Norm, Metric of \mathbb{C}). Moreover, the isomorphism

$$\gamma : \mathbb{C} \rightarrow \mathbb{R}^2, \quad \gamma(x + yi) = (x, y)$$

induces additional structures on \mathbb{C} , such as the norm and metric.

1. The norm of $z = x + iy \in \mathbb{C}$ is defined as the norm of $\gamma(z) = (x, y) \in \mathbb{R}^2$. That is,

$$|z| = |x + yi| = |(x, y)| = \sqrt{x^2 + y^2}$$

Or more simply,

$$|z| = z \cdot \bar{z}$$

2. The metric of two complex numbers $z_1, z_2 \in \mathbb{C}$ is defined

$$|z_1 - z_2| = |(x_1, y_1) - (x_2, y_2)| = |(x_1 - x_2, y_1 - y_2)| = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Or more simply,

$$|z_1 - z_2| = (z_1 - z_2) \cdot \overline{(z_1 - z_2)}$$

Definition 5.16 (Polar Coordinates of \mathbb{C}). Given the basis transformation of polar coordinates $(r, \varphi) \mapsto p(r, \varphi) = (x, y)$ where

$$p \begin{pmatrix} r \\ \varphi \end{pmatrix} = \begin{pmatrix} r \cos \varphi \\ r \sin \varphi \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}$$

the isomorphism $\mathbb{C} \simeq \mathbb{R}^2$ induces a similar polar transformation in \mathbb{C}

$$\rho = \gamma^{-1} \circ p \circ \gamma : \mathbb{C}_{(r, \theta)} \rightarrow \mathbb{C}_{(x, y)}, \quad \rho(r + \theta i) = r \cos \theta + r \sin \theta i = x + yi$$

as shown in the commutative diagram.

$$\begin{array}{ccc}\mathbb{C}_{(r, \theta)} & \xrightarrow{\rho} & \mathbb{C}_{(x, y)} \\ \downarrow \gamma & & \downarrow \gamma \\ \mathbb{R}^2_{(r, \theta)} & \xrightarrow{p} & \mathbb{R}^2_{(x, y)}\end{array}$$

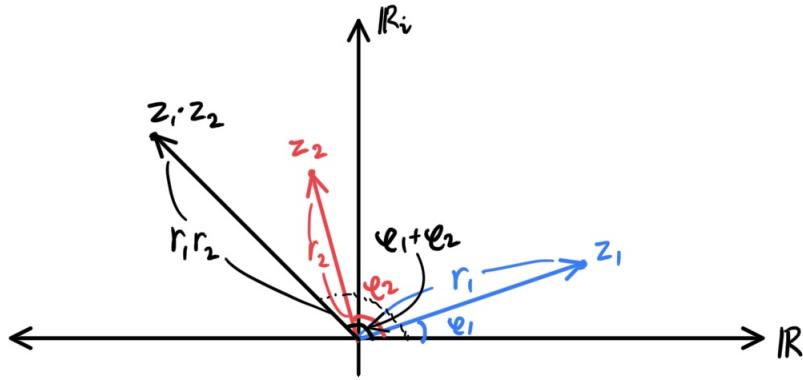
Therefore, we can write

$$z = r(\cos \varphi + i \sin \varphi)$$

where $r = |z|$ is called the *magnitude* of z , and $\varphi = \operatorname{Arg} z$ is called the *argument* of z .

Lemma 5.22 (Multiplication of Complex Numbers in Polar Form). *It turns out that multiplication is a lot easier in polar coordinates than in rectangular ones:*

$$\begin{aligned}z_1 \cdot z_2 &= (r_1 \cos \varphi_1 + ir_1 \sin \varphi_1)(r_2 \cos \varphi_2 + ir_2 \sin \varphi_2) \\ &= \dots \\ &= r_1 r_2 (\cos(\varphi_1 + \varphi_2) + i \sin(\varphi_1 + \varphi_2))\end{aligned}$$



Theorem 5.23 (De Moivre's Formula). *By induction using the previous lemma, we get*

$$z = r(\cos \varphi + i \sin \varphi) \implies z^n = r^n(\cos n\varphi + i \sin n\varphi)$$

Corollary 5.23.1 (Roots of Unity). *The n complex solutions of the equation*

$$z^n = a$$

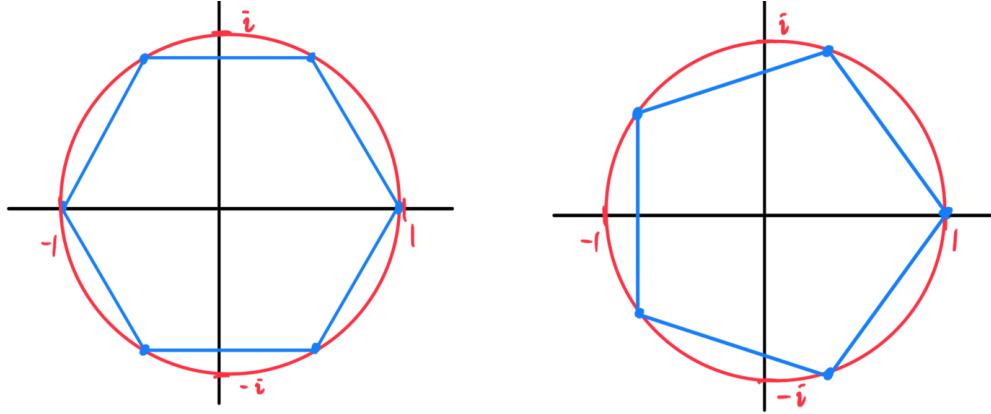
where $a = \rho(\cos \psi + i \sin \psi)$ is

$$z_k = \sqrt[n]{\rho} \left(\cos \left(\frac{\psi + 2\pi k}{n} \right) + i \sin \left(\frac{\psi + 2\pi k}{n} \right) \right), \quad k = 0, 1, 2, \dots, n-1$$

Moreover, if $a = 1$, then the n complex solutions are called the n th roots of unity, defined

$$z_k = \cos \left(\frac{2\pi k}{n} \right) + i \sin \left(\frac{2\pi k}{n} \right), \quad k = 0, 1, 2, \dots, n-1$$

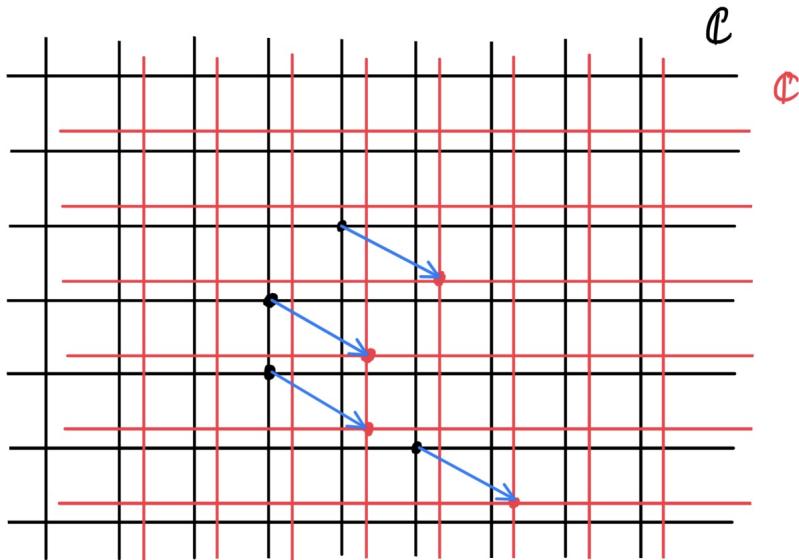
which shows that the n th roots of unity are at the vertices of a regular n -sided polygon inscribed in the unit circle, with one vertex at 1, within the complex plane. The 5th and 6th roots of unity are shown below.



Finally, we can visualize certain transformations in \mathbb{C} . For a fixed $b \in \mathbb{C}$, the sum $z + b$ can be interpreted as the mapping of \mathbb{C} onto itself given by the formula

$$z \mapsto z + b$$

This mapping is a translation of the plane by the vector b .



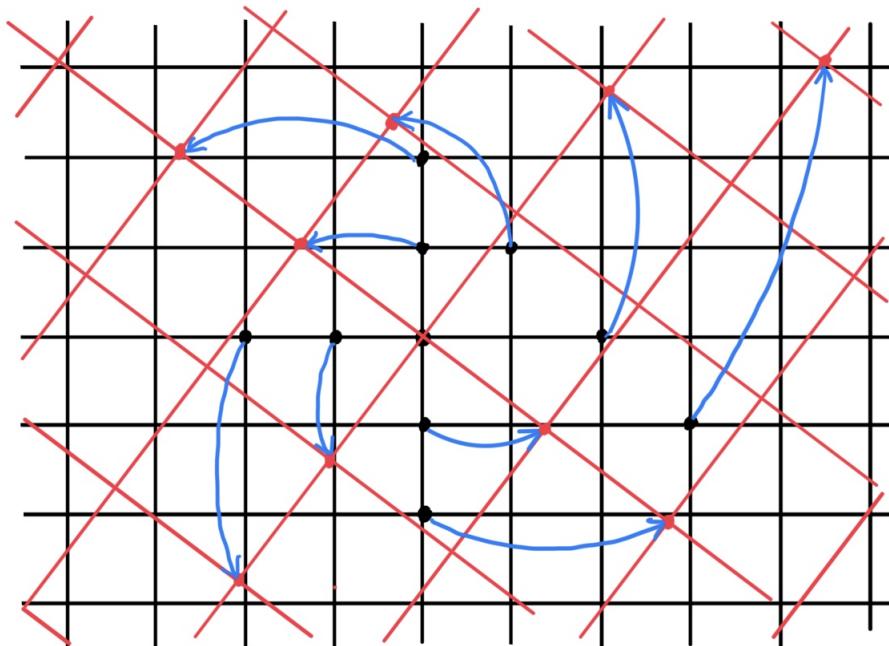
Visualizing multiplication is a bit harder. Given a

$$a = |a|(\cos \varphi + i \sin \varphi) \neq 0$$

the product az can be interpreted as the mapping of \mathbb{C} onto itself given by the formula

$$z \mapsto az$$

which is the composition of a dilation by a factor of $|a|$ and a rotation through the angle $\varphi \in \text{Arg } a$.

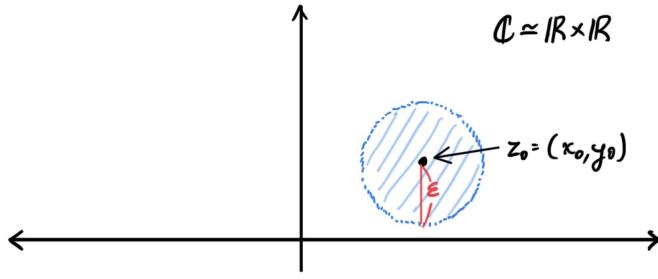


5.6.3 Sequences and Series in \mathbb{C}

Our previous construction of a metric within \mathbb{C} enables to define the ϵ -neighborhood of a number $z_0 \in \mathbb{C}$ as the set

$$U_\epsilon(z_0) \equiv \{z \in \mathbb{C} \mid |z - z_0| < \epsilon\}$$

which can be visualized as an open disk of radius ϵ in \mathbb{R}^2 centered at point (x_0, y_0) if $z_0 = x_0 + iy_0$.



Definition 5.17 (Convergence of a Sequence in \mathbb{C}). A sequence $\{z_n\}$ of complex numbers *converges* to $z_0 \in \mathbb{C}$ if and only if

$$\lim_{n \rightarrow \infty} |z_n - z_0| = 0$$

It is clear from the inequality

$$\max\{|x_n - x_0|, |y_n - y_0|\} \leq |z_n - z_0| \leq |x_n - x_0| + |y_n - y_0|$$

that a sequence of complex numbers converges if and only if the sequences of its real and imaginary parts of the terms of the sequence both converge. That is,

$$\{z_n\} \text{ converges} \iff \{\operatorname{Re} z\} \text{ and } \{\operatorname{Im} z\} \text{ converges}$$

Lemma 5.24 (Convergence of Cauchy Sequences over \mathbb{C}). A sequence of complex numbers $\{z_n\}$ is called a *Cauchy sequence* if for every $\epsilon > 0$ there exists an index $N \in \mathbb{N}$ such that

$$|z_n - z_m| < \epsilon \text{ for all } n, m > N$$

It is also clear that

$$\{z_n\} \text{ is Cauchy} \iff \{\operatorname{Re} z\} \text{ and } \{\operatorname{Im} z\} \text{ is Cauchy}$$

and using the Cauchy criterion for sequences of real numbers, we can easily see that a sequence of complex numbers converges if and only if it is a Cauchy sequence.

Lemma 5.25 (Convergence of Cauchy Series over \mathbb{C}). Interpreting the sum of a series of complex numbers

$$z_1 + z_2 + \dots + z_n + \dots$$

as the limit of the sequence its partial sums $\{s_n\}$, where $s_n = z_1 + \dots + z_n$ as $n \rightarrow \infty$, we can see that the series converges if and only if for every $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that

$$|z_m + \dots + z_n| < \epsilon$$

for any natural numbers $n \geq m > N$.

Definition 5.18 (Absolute Convergence of \mathbb{C}). A series $z_1 + \dots + z_n + \dots$ of complex numbers is *absolutely convergent* if the series

$$|z_1| + |z_2| + \dots + |z_n| + \dots$$

converges. Clearly, if a series converges absolutely, then it converges due to the inequality

$$|z_m + \dots + z_n| \leq |z_m| + \dots + |z_n|$$

Example 5.5. The following complex series converges because they converges absolutely. That is,

$$\begin{aligned} 1 + \frac{1}{1!}|z| + \frac{1}{2!}|z^2| + \dots &\text{ converges } \forall \mathbb{C} \implies 1 + \frac{1}{1!}z + \frac{1}{2!}z^2 + \dots \text{ converges } \forall \mathbb{C} \\ |z| + \frac{1}{3!}|z^3| + \frac{1}{5!}|z^5| + \dots &\text{ converges } \forall \mathbb{C} \implies z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \dots \text{ converges } \forall \mathbb{C} \\ 1 + \frac{1}{2!}|z|^2 + \frac{1}{4!}|z|^4 + \dots &\text{ converges } \forall \mathbb{C} \implies 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \dots \text{ converges } \forall \mathbb{C} \end{aligned}$$

Definition 5.19 (Complex Power Series). Series of the form

$$\sum_{n=0}^{\infty} c_n(z - z_0)^n = c_0 + c_1(z - z_0) + \dots + c_n(z - z_0) + \dots$$

are called *complex power series*, or *power series over \mathbb{C}* .

But a power series is quite useless unless we know the domain in which it converges (again, note that it is not always guaranteed to converge onto the function f if its power series expansion does converge at all). To develop more sophisticated tests of convergence of a complex power series, we introduce the complex analogue of the root test for real power series.

Theorem 5.26 (Cauchy-Hadamard Theorem). *The complex power series*

$$c_0 + c_1(z - z_0) + \dots + c_n(z - z_0) + \dots$$

converges inside the disk $|z - z_0| < R$ with center at z_0 and radius given by the formula

$$R = \frac{1}{\overline{\lim}_{n \rightarrow \infty} \sqrt[n]{|c_n|}} = \frac{1}{\lim_{n \rightarrow \infty} \sup \sqrt[n]{|c_n|}}$$

Where $\overline{\lim}$ denotes the superior limit. Furthermore,

1. the power series diverges at any point exterior to the disk.
2. the power series converges absolutely at any point interior to the disk.
3. the power series is indeterminate at any point on the boundary of the disk.

Note that in the degenerate case when $R = 0$, the series converges only at the point $z = z_0$.

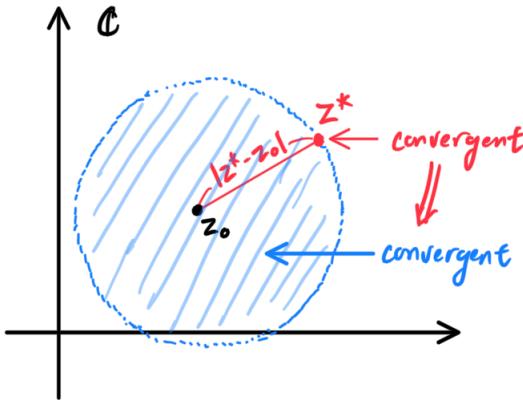
Corollary 5.26.1 (Abel's First Theorem on Power Series). *If the power series*

$$c_0 + c_1(z - z_0) + \dots + c_n(z - z_0) + \dots$$

converges at some value z^* , then it converges absolutely for any value of z satisfying

$$|z - z_0| < |z^* - z_0|$$

The values of z satisfying the inequality above can be intuitively visualized as the following region.



Theorem 5.27 (Product of Absolutely Convergent Series). *Let $a_1 + a_2 + \dots$ and $b_1 + b_2 + \dots$ be an absolutely convergent series such that*

$$\sum_{i=1}^{\infty} a_i = A \text{ and } \sum_{j=1}^{\infty} b_j = B$$

Then, the Cauchy product of the two series

$$\left(\sum_{i=1}^{\infty} a_i \right) \cdot \left(\sum_{j=1}^{\infty} b_j \right) = \sum_{k=0}^{\infty} c_k = AB, \text{ where } c_k = \sum_{l=0}^k a_l b_{k-l}$$

$a_1 b_1 + a_2 b_2 + \dots$ is absolutely convergent and

$$\sum_{i=1}^{\infty} a_i b_i = AB$$

Proof. To be done. ■

Example 5.6 (Convergence of the Cauchy Product of Absolutely Convergent Complex Series). *The two series*

$$\sum_{n=0}^{\infty} \frac{1}{n!} a^n \text{ and } \sum_{m=0}^{\infty} \frac{1}{m!} b^m$$

converges absolutely. Therefore, we can see that their Cauchy product can be nicely represented by grouping together all monomials of the form $a^n b^m$ having the same total degree $m+n = k$.

$$\left(\sum_{n=0}^{\infty} \frac{1}{n!} a^n \right) \cdot \left(\sum_{m=0}^{\infty} \frac{1}{m!} b^m \right) = \sum_{k=0}^{\infty} \left(\sum_{n+m=k} \frac{1}{n!} a^n \frac{1}{m!} b^m \right)$$

But we can simplify

$$\sum_{m+n=k} \frac{1}{n!m!} a^n b^m = \frac{1}{k!} \sum_{n=0}^k \frac{k!}{n!(k-n)!} a^n b^{k-n} = \frac{1}{k!} (a+b)^k$$

and therefore we find that

$$\left(\sum_{n=0}^{\infty} \frac{1}{n!} a^n \right) \cdot \left(\sum_{m=0}^{\infty} \frac{1}{m!} b^m \right) = \sum_{k=0}^{\infty} \frac{1}{k!} (a+b)^k$$

5.6.4 Euler's Formula

Definition 5.20 (Complex Taylor Expansions of Transcendental Functions). Since we have determined absolute convergence, and therefore convergence, of all these series in all of \mathbb{C} , it is natural to extend the definitions of

$$\exp, \cos, \sin : \mathbb{R} \longrightarrow \mathbb{R}$$

to the complex field

$$\exp, \cos, \sin : \mathbb{C} \longrightarrow \mathbb{C}$$

by defining them as

$$\begin{aligned} e^z &\equiv 1 + \frac{1}{1!} z + \frac{1}{2!} z^2 + \frac{1}{3!} z^3 + \dots \\ \cos z &\equiv 1 - \frac{1}{2!} z^2 + \frac{1}{4!} z^4 - \frac{1}{6!} z^6 + \dots \\ \sin z &\equiv z - \frac{1}{3!} z^3 + \frac{1}{5!} z^5 - \frac{1}{7!} z^7 + \dots \end{aligned}$$

Notice that even in the complex field, $\cos z$ is an even function and $\sin z$ is an odd function.

$$\begin{aligned} \cos(-z) &= \cos(z) \\ \sin(-z) &= -\sin(z) \end{aligned}$$

In fact, the last example in the previous subsection just proves the following.

Lemma 5.28 (Exponential Map as a Group Homomorphism). *The exponential map $\exp : \mathbb{C} \longrightarrow \mathbb{C} \setminus \{0\}$ satisfies the following*

$$\exp(z_1 + z_2) = \exp(z_1) \cdot \exp(z_2)$$

That is, \exp is a group homomorphism from $(\mathbb{C}, +)$ to $(\mathbb{C} \setminus \{0\}, \cdot)$.

Definition 5.21 (Euler's Formula). By making the substitution $z = yi$ in the series expansion of e^z (where y is an arbitrary complex number), we get

$$\begin{aligned} e^{iy} &= 1 + \frac{1}{1!}(iy) + \frac{1}{2!}(iy)^2 + \frac{1}{3!}(iy)^3 + \frac{1}{4!}(iy)^4 + \dots \\ &= \left(1 - \frac{1}{2}y^2 + \frac{1}{4!}y^4 - \dots \right) + i \left(\frac{1}{1!}y - \frac{1}{3!}y^3 + \frac{1}{5!}y^5 - \dots \right) \end{aligned}$$

which brings us the identity

$$e^{iy} = \cos y + i \sin y$$

Since cos is even and sin is odd, we can add the two identities

$$\begin{aligned} e^{iz} &= \cos z + i \sin z \\ e^{-iz} &= \cos z - i \sin z \end{aligned}$$

to get

$$\begin{aligned} \cos z &= \frac{1}{2}(e^{iz} + e^{-iz}) \\ \sin z &= \frac{1}{2i}(e^{iz} - e^{-iz}) \end{aligned}$$

This gives us a very elegant connection between these three transcendental functions.

Definition 5.22 (Hyperbolic Functions). Likewise, the following series are convergent (since they are absolutely convergent) and therefore we can define the extension of cosh and sinh into the complex field as

$$\begin{aligned} \cosh z &\equiv 1 + \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \frac{1}{6!}z^6 + \dots \\ \sinh z &\equiv z + \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \frac{1}{7!}z^7 + \dots \end{aligned}$$

The following identities immediately follow

$$\begin{aligned} \cosh z &= \frac{1}{2}(e^z + e^{-z}) \\ \sinh z &= \frac{1}{2}(e^z - e^{-z}) \end{aligned}$$

Lemma 5.29 (Trigonometric, Hyperbolic Identities over \mathbb{C}). *Common identities, which are exactly the same as their real analogues, are listed.*

1. $\cos^2 z + \sin^2 z = 1$
2. $\cosh^2 z - \sinh^2 z = 1$
3. $e^{i(z_1+z_2)} = (\cos z_1 \cos z_2 - \sin z_1 \sin z_2) + i(\sin z_1 \cos z_2 + \cos z_1 \sin z_2)$
4. $\cos(z_1 + z_2) = \cos z_1 \cos z_2 - \sin z_1 \sin z_2$
5. $\sin(z_1 + z_2) = \sin z_1 \cos z_2 + \cos z_1 \sin z_2$
6. $\cosh z = \cos iz$
7. $\sinh z = -i \sin iz$

However, to obtain even such geometrically obvious facts as the equality

$$\sin \pi = 0 \text{ or } \cos z + 2\pi = \cos z$$

from the power series definitions of cos and sin is extremely difficult. What the properties actually do is present the remarkable unity of these seemingly different trigonometric and hyperbolic functions, which would have been impossible to detect without going into the domain of complex numbers.

If we just take the following identities

$$\begin{aligned} \cos x &= \cos(x + 2\pi) \\ \sin x &= \sin(x + 2\pi) \\ \cos 0 &= 1 \\ \sin 0 &= 0 \end{aligned}$$

then we get the following identity.

Theorem 5.30 (Euler's Identity). *The following relation is true.*

$$e^{i\pi} + 1 = 0$$

which immediately implies

$$\exp(z + 2\pi i) = \exp z$$

That is, the exponential function is a periodic function on \mathbb{C} with the purely imaginary period $T = 2\pi i$.

Corollary 5.30.1 (Trigonometric Notation of Complex Number). *With Euler's formula and the periodic relation of $\exp z$, the trigonometric form of a complex number can be presented as*

$$z = r(\cos \varphi + i \sin \varphi) = re^{i\varphi}$$

We can rewrite DeMoivre's formula as

$$z^n = r^n e^{n\varphi i}$$

5.6.5 Visualizing Complex Functions

5.6.6 Continuity, Differentiability, Analyticity of Complex Functions

The definitions of continuity and differentiability are the same, just under a different field.

Definition 5.23 (Limit of a Complex Function). The function $f : E \subset \mathbb{C} \rightarrow \mathbb{C}$ tends to $A \in \mathbb{C}$ as $z \rightarrow a$, or that

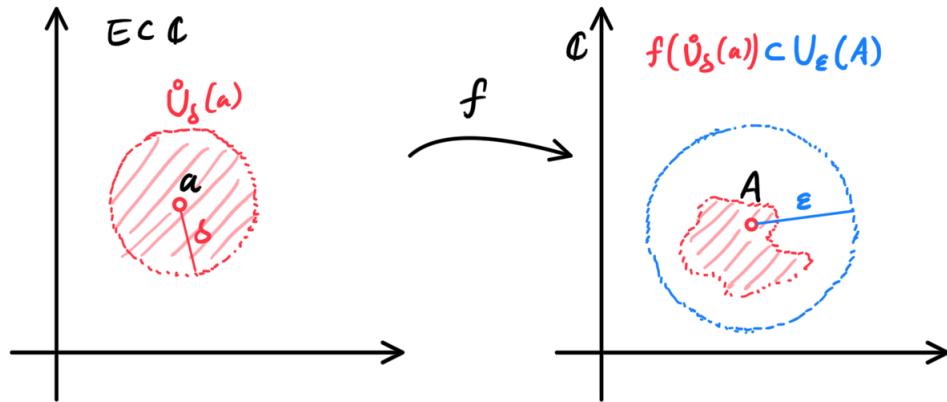
$$\lim_{z \rightarrow a} f(z) = A$$

if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$0 < |z - a| < \delta \implies |f(z) - A| < \epsilon$$

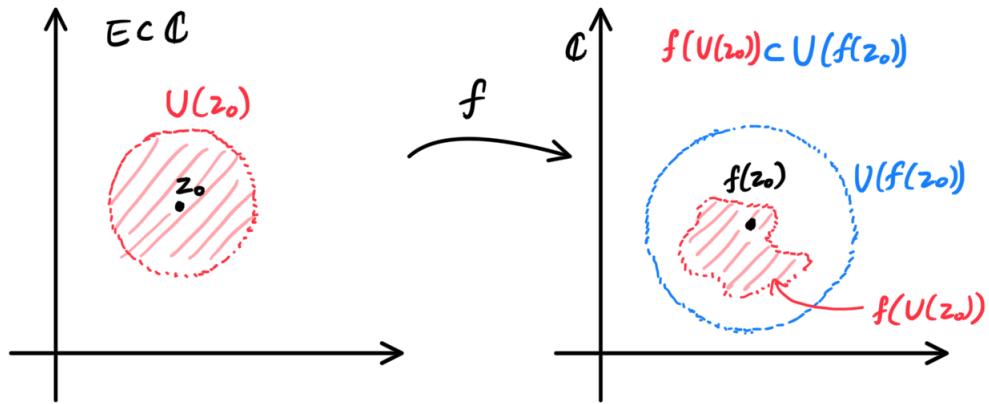
Note that we set $0 < |z - a|$ to ensure that $z \neq a$.

Therefore, in other words, for any arbitrarily small $\epsilon > 0$, we can find a $\delta > 0$ such that the image of the deleted δ -neighborhood of a , denoted $\mathring{U}_\delta(a)$, is completely within the ϵ -neighborhood $U_\epsilon(A)$.



Definition 5.24 (Continuity of a Complex Function). A function $f : E \subset \mathbb{C} \rightarrow \mathbb{C}$ is *continuous* at a point $z_0 \in E$ if for any neighborhood $U(f(z_0))$ there exists a neighborhood $U(z_0)$ such that its image is contained in $U(f(z_0))$. In short,

$$\lim_{z \rightarrow z_0} f(z) = f(z_0)$$



Definition 5.25 (Differentiability of a Complex Function). The *derivative* of a function $f : E \subset \mathbb{C} \rightarrow \mathbb{C}$ is defined

$$f'(z_0) = \lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

if this limit exists. f *differentiable* at z_0 means that a differential function

$$df(z_0) : T_{z_0} \mathbb{C} \rightarrow T_{f(z_0)} \mathbb{C}, \quad h \mapsto df(z_0)(h)$$

exists such that

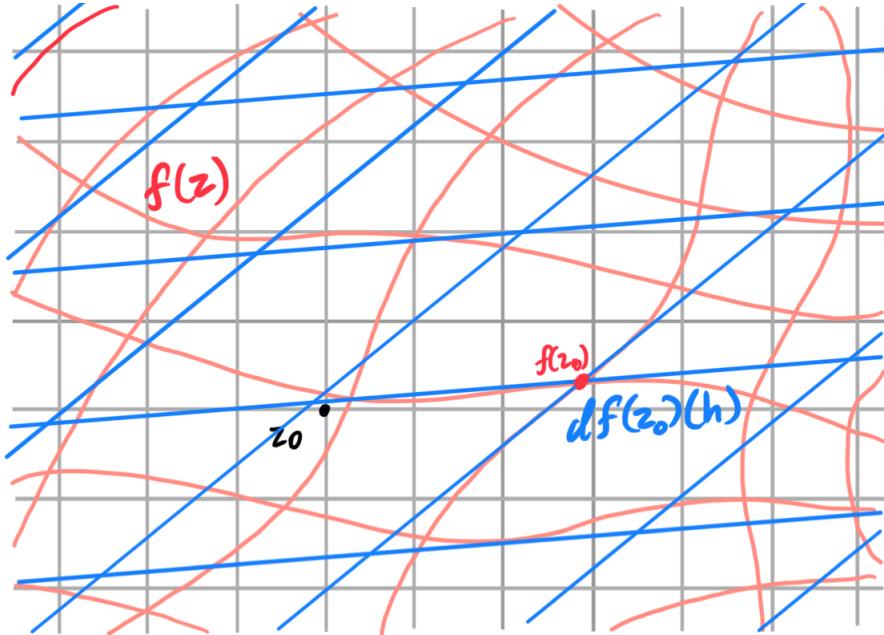
$$f(z) = f(z_0) + df(z_0)(h) + o(h)$$

where $h = z - z_0$ is the increment of the argument. Just like the real case, it turns out that $df(z_0)(h) = f'(z_0)h$, and

$$f(z) - f(z_0) = f'(z_0)(z - z_0) + o(z - z_0)$$

which elegantly weaves together the two concepts of differentiability and the derivative.

Visualizing this, we can see that for whatever function $f : \mathbb{C} \rightarrow \mathbb{C}$ there is a linear function that transforms the entire space as such at z_0 (along with a given point $z_0 \in \mathbb{C}$),



The differential $df(z_0)$ at the point z_0 is a linear mapping that "best" approximates f , with an error of $o(h) = o(z - z_0)$.

Lemma 5.31 (Arithmetic Properties of Differentiation over \mathbb{C}). *If functions $f, g : E \subset \mathbb{C} \rightarrow \mathbb{C}$ are differentiable at a point $z \in E$, then*

1. *their sum is differentiable at z , and*

$$d(f+g)(z) = df(z) + dg(z) \iff (f+g)'(z) = (f'+g')(z)$$

2. *their product is differentiable at z , and*

$$d(f \cdot g)(z) = g(z)df(z) + f(z)dg(z) \iff (f \cdot g)'(z) = f'(z)g(z) + f(z) \cdot g'(z)$$

3. *their quotient is differentiable at z if $g(z) \neq 0$, and*

$$d\left(\frac{f}{g}\right)(z) = \frac{g(z)df(z) - f(z)dg(z)}{g^2(z)} \iff \left(\frac{f}{g}\right)'(z) = \frac{f'(z)g(z) - f(z)g'(z)}{g^2(z)}$$

Just like the real case, the operation of taking the derivative is a linear operator.

Lemma 5.32 (Chain Rule for Composite Functions over \mathbb{C}). *Let there be functions $f : E_1 \subset \mathbb{C} \rightarrow \mathbb{C}$ differentiable at point $z \in E_1$ and $g : E_2 \subset \mathbb{C} \rightarrow \mathbb{C}$ differentiable at point $w = f(z) \in E_2$, with respective differentials*

$$\begin{aligned} df(z) &: T_z \mathbb{C} \rightarrow T_w \mathbb{C} \\ dg(w) &: T_w \mathbb{C} \rightarrow T_{g(w)} \mathbb{C} \end{aligned}$$

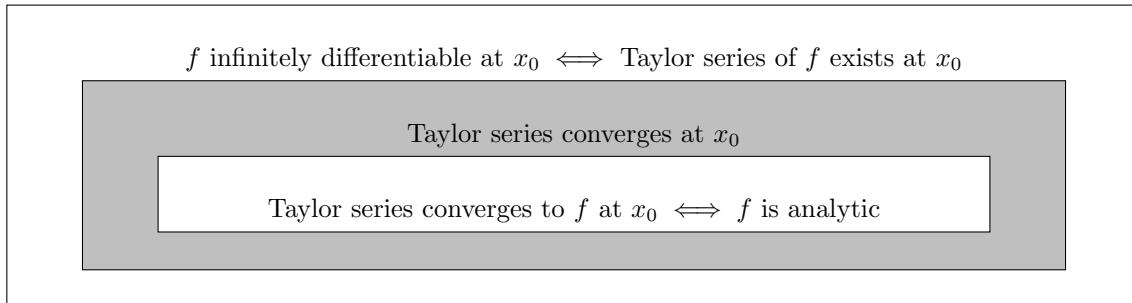
Then, the composite function $g \circ f : E_1 \rightarrow \mathbb{C}$ is differentiable at z , and $d(g \circ f)(z) : T_z \mathbb{C} \rightarrow T_{g(f(z))} \mathbb{C}$ is

$$d(g \circ f)(z) = dg(w) \circ df(z) \iff (g \circ f)'(z) = g'(f(z)) \circ f'(z)$$

5.6.7 Power Series Representation of a Function

Definition 5.26 (Holomorphic Function). *If function $f : E \subset \mathbb{C} \rightarrow \mathbb{C}$ is (complex) differentiable at a point $z_0 \in E$, then f is said to be *holomorphic at z_0* .*

We recall the diagram that summarizes the conditions of differentiability and analyticity of a function f over the field \mathbb{R} .

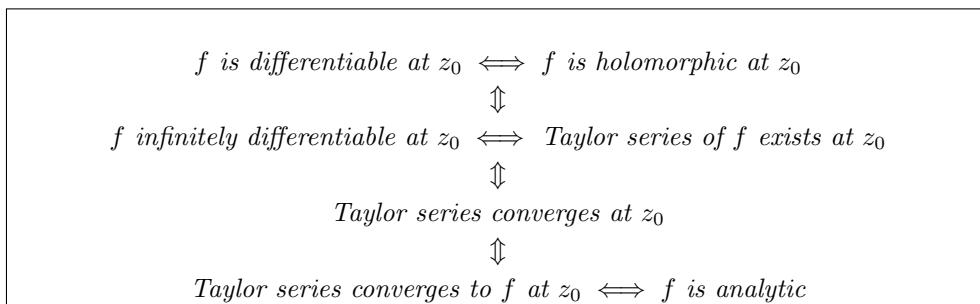


In the theory of functions of a complex variable we actually have a remarkable theorem that does not have an analogue for functions over \mathbb{R} .

Theorem 5.33 (Analyticity of Differentiable Functions over \mathbb{C}). *If a function $f : E \subset \mathbb{C} \rightarrow \mathbb{C}$ is differentiable in a neighborhood of a point $z_0 \in E$, then it is analytic at that point. In other words,*

$$f \text{ is holomorphic at } z_0 \implies f \text{ is analytic at } z_0$$

This means that the conditions in the diagram above all are equivalent! Visually,



This is certainly an amazing fact, since it then follows from the theorem that if a function $f(z)$ has one derivative $f'(z)$ in a neighborhood of a point, it also has derivatives of all orders in that neighborhood.

5.6.8 Algebraic Closedness of the Field \mathbb{C}

Definition 5.27 (Algebraically Closed Field). A field \mathbb{F} is *algebraically closed* if every nonconstant polynomial in $\mathbb{F}[x]$ (the polynomial ring with coefficients in \mathbb{F}) has a root in \mathbb{F} .

Theorem 5.34 (Fundamental Theorem of Algebra). \mathbb{C} is algebraically closed. That is, every polynomial

$$P(z) \equiv c_0 + c_1 z + c_2 z^2 + \dots + c_n z^n$$

of degree $n \geq 1$ with complex coefficients $c_i \in \mathbb{C}$ ($i = 0, 1, \dots, n$) has a root in \mathbb{C} . This immediately implies that every polynomial $P(z)$ admits a representation (unique up to the order of the factors) in the form

$$P(z) = c_n(z - z_1)(z - z_2) \dots (z - z_n)$$

where $z_1, \dots, z_n \in \mathbb{C}$ not necessarily all distinct.

We can also prove the interesting property about zeroes of polynomials in $\mathbb{R}[x]$.

Corollary 5.34.1 (Complex Conjugate Roots of Real Polynomials). Given a polynomial with real coefficients

$$P(z) \equiv a_0 + a_1 z + a_2 z^2 + \dots + a_n z^n$$

P , as we know, does not always have real roots (e.g. $P(x) = x^2 + 1$). However, we state that

$$\text{if } P(z_0) = 0, \text{ then } P(\bar{z}_0) = 0$$

Therefore, every polynomial P with real coefficients can be expanded as a product of linear and quadratic polynomial with real coefficients.

Proof. We can see from the properties of complex numbers that

$$\begin{aligned} \overline{(z_1 + z_2)} &= \overline{z_1} + \overline{z_2} \\ \overline{(z_1 \cdot z_2)} &= \overline{(r_1 e^{i\varphi_1} \cdot r_2 e^{i\varphi_2})} \\ &= \overline{r_1 r_2 e^{i(\varphi_1 + \varphi_2)}} = r_1 r_2 e^{-i(\varphi_1 + \varphi_2)} \\ &= r_1 e^{-i\varphi_1} \cdot r_2 e^{-i\varphi_2} = \bar{z}_1 \cdot \bar{z}_2 \end{aligned}$$

Thus, if $P(z_0) = 0$, then

$$0 = \overline{P(z_0)} = \overline{a_0 + \dots + a_n z_0^n} = \bar{a}_0 + \dots + \bar{a}_n \bar{z}_0^n = a_0 + \dots + a_n \bar{z}_0^n = P(\bar{z}_0)$$

and thus $P(\bar{z}_0) = 0$. ■

5.7 Primitives

Definition 5.28 (Primitive). A function $F(x)$ is a *primitive* of a function $f(x)$ on an interval if F is differentiable on the interval and satisfies the equation

$$F'(x) = f(x)$$

or equivalently, if their respective differentials satisfy

$$dF(x) = f(x) dx$$

Lemma 5.35. If $F_1(x)$ and $F_2(x)$ are two primitives of $f(x)$ on the same interval, then the difference $(F_1 - F_2)(x)$ is constant on that interval.

Example 5.7. Both

$$F_1(x) \equiv \arctan x \text{ and } F_2(x) \equiv \operatorname{arccot} \frac{1}{x}$$

are primitives of $f(x) = \frac{1}{1+x^2}$. Indeed, we can see by direct calculation that in the domain $\mathbb{R} \setminus 0$,

$$F_1(x) - F_2(x) = \arctan x - \operatorname{arccot} \frac{1}{x} = \begin{cases} 0, & x > 0 \\ -\pi, & x < 0 \end{cases}$$

which is supported by the lemma.

Notice how given a function $f(x)$, the operation of finding its differential, denoted with d , gives us a new function of h , called the differential

$$df(x)(h)$$

Similarly, the operation of finding a primitive of function $f(x)$, denoted with the symbol \int , gives us a new function.

Definition 5.29 (Indefinite Integration). The operation of finding a primitive of a certain function $f(x)$ is called *indefinite integration*, and the mathematical notation

$$\int f(x) dx$$

is called the *indefinite integral of $f(x)$* on a given interval (f called the *integrand* and $f(x) dx$ called the *differential form*).

1. It immediately follows from the lemma that if $F(x)$ is any particular primitive of $f(x)$ on the interval, then on that interval

$$\int f(x) dx = F(x) + C$$

2. If $F'(x) = f(x)$ (that is, F is a primitive of f on some interval), then we have

$$d \int f(x) dx = dF(x) = F'(x) dx$$

3. It also follows that

$$\int dF(x) = \int F'(x) dx = F(x) + C$$

Theorem 5.36 (Basic Methods of Indefinite Integration). *The definition of the indefinite integral has three basic properties that can be used to solve indefinite integrals.*

1. *Linearity of the indefinite integral.*

$$\int (\alpha u(x) + \beta v(x)) dx = \alpha \int u(x) dx + \beta \int v(x) dx + C$$

2. *Integration by parts.*

$$\int (uv)' dx = \int u'(x)v(x) dx + \int u(x)v'(x) dx + C$$

3. *Change of Variable, or U-substitution.* Given that $F'(x) = f(x)$ on an interval I_x and $\varphi : I_t \rightarrow I_x$ is a C^1 mapping of interval I_t into I_x , then

$$\int (f \circ \varphi)(t) \varphi'(t) dt = (F \circ \varphi)(t) + C$$

6 Integration

6.1 Construction of the Riemann Integral

We shall first define the integral using the familiar notation of Riemann sums.

Definition 6.1 (Partitions with Distinguished Points). A *partition* P of a closed interval $[a, b]$, $a < b$, is a finite system of points x_0, \dots, x_n of the interval such that

$$a = x_0 < x_1 < x_2 < \dots < x_n = b$$

The intervals $[x_{i-1}, x_i]$, $i = 1, 2, \dots, n$, are called the *intervals* of the partition P . The largest of the lengths of the intervals of the partition P , denoted $\lambda(P)$, is called the *mesh* of the partition.

A *partition with distinguished points* (P, ξ) on the closed interval $[a, b]$ is a partition P of $[a, b]$ along with the set of n points

$$\xi_1 \in [x_0, x_1], \xi_2 \in [x_1, x_2], \dots, \xi_n \in [x_{n-1}, x_n]$$

The n -tuple of ξ_i 's is denoted by the single letter ξ

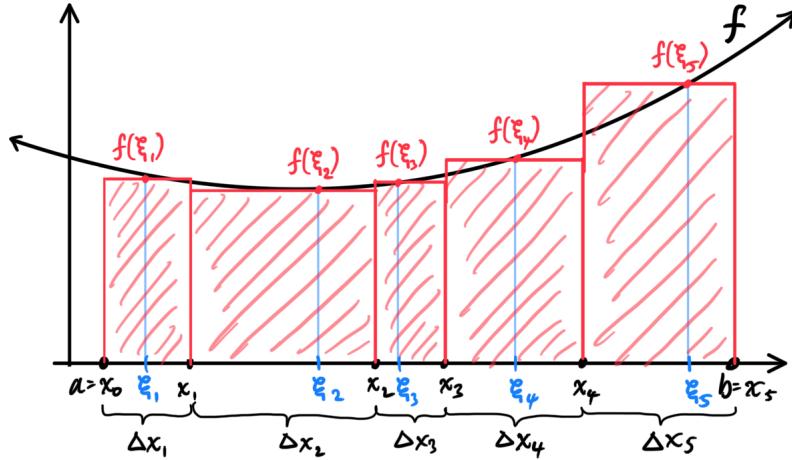
$$\xi = (\xi_1, \xi_2, \dots, \xi_n)$$

This naturally leads to the following construction.

Definition 6.2 (Riemann Sums). If a function f is defined on a closed interval $[a, b]$ and (P, ξ) is a partition with distinguished points on this closed interval, the sum

$$\sigma(f; P, \xi) \equiv \sum_{i=1}^n f(\xi_i) \Delta x_i, \text{ where } \Delta x_i = x_i - x_{i-1},$$

is the *Riemann sum* of the function f corresponding to the partition (P, ξ) with distinguished points on $[a, b]$.



Thus, when a function f is fixed, the Riemann sum $\sigma(f; P, \xi)$ is a mapping that takes in a partition with distinguished points $p = (P, \xi)$ on the closed interval $[a, b]$ and outputs a number representing the total area of the Riemann sums. That is, for a fixed f and some input $p = (P, \xi)$, we can define the function

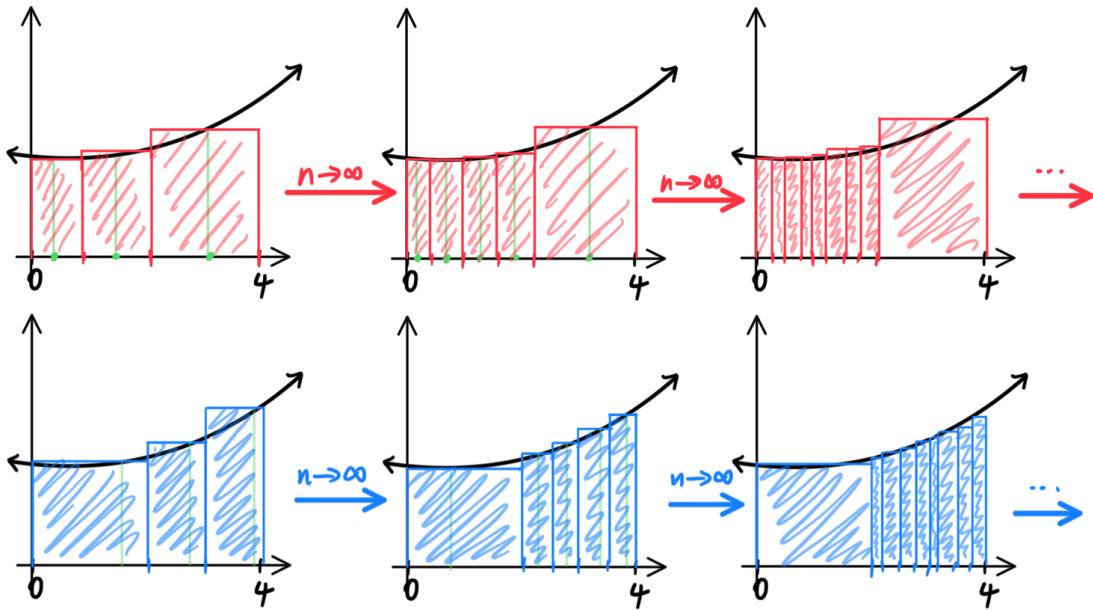
$$\Phi : \mathcal{P} \longrightarrow \mathbb{R}, \quad \Phi(p) \equiv \sigma(f; p) \equiv \sigma(f; (P, \xi))$$

that takes in a partition with distinguished points on $[a, b]$ and outputs the corresponding Riemann sum for that fixed f .

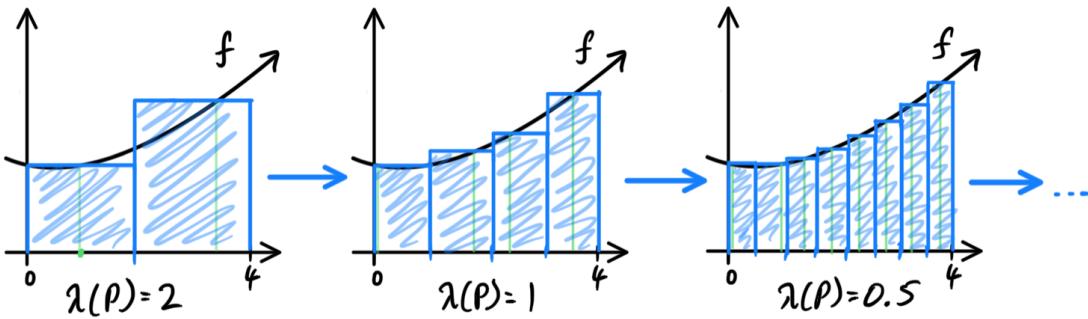
Definition 6.3 (Riemann Integral). The number $\int_a^b f(x) dx$ is the *Riemann integral* of the function f on the closed interval $[a, b]$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\left| \int_a^b f(x) dx - \sum_{i=1}^n f(\xi_i) \Delta x_i \right| < \epsilon$$

for any partition (P, ξ) with distinguished points on $[a, b]$ whose mesh $\lambda(P)$ is less than δ . We can view this as a limit where $n \rightarrow \infty$, but there is a problem since we can increase the partition within different subsets of $[a, b]$, leading to multiple values of convergence.



Rather, we can set the mesh $\lambda(P)$ to approach 0, which would take care of the problems. We can visualize this by imagining the lengths of the rectangles converging "uniformly."



Therefore, we can culminate by defining the Riemann integral of $f(x)$ over $[a, b]$ as

$$\int_a^b f(x) dx \equiv \lim_{\lambda(P) \rightarrow 0} \sum_{i=1}^n f(\xi_i) \lambda x_i$$

6.1.1 Conditions for Integrability

Definition 6.4 (Riemann Integrable Functions). A function f is *Riemann integrable* on the closed interval $[a, b]$ if

$$\int_a^b f(x) dx \equiv \lim_{\lambda(P) \rightarrow 0} \sum_{i=1}^n f(\xi_i) \lambda x_i$$

is defined, i.e. if the limit of the right-hand side of Riemann sums exists as $\lambda(P) \rightarrow 0$ (that is, the Riemann integral of f is defined).

Furthermore, the set of Riemann-integrable functions on a closed interval $[a, b]$ is denoted $\mathcal{R}[a, b]$.

Remember that the Riemann integral, as complicated as the formula is, is still a limit of a function. That means that we can apply the Cauchy criterion to it to determine convergence.

Lemma 6.1 (Cauchy Criterion on Existence of Riemann Integral). *Given a function f , the integral of f over $[a, b]$, defined*

$$\int_a^b f(x) dx \equiv \lim_{\lambda(P) \rightarrow 0} \sum_{i=1}^n f(\xi_i) \lambda x_i$$

exists if and only if for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$|\sigma(f; P', \xi') - \sigma(f; P'', \xi'')| < \epsilon$$

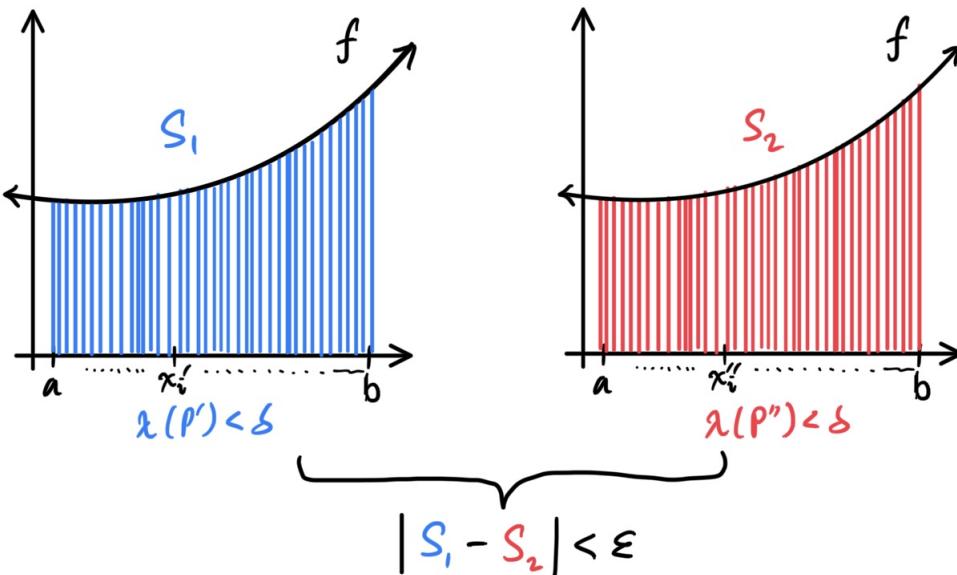
or, what is the same,

$$\left| \sum_{i=1}^{n'} f(\xi'_i) \Delta x'_i - \sum_{i=1}^{n''} f(\xi''_i) \Delta x''_i \right| < \epsilon$$

for any partitions (P', ξ') and (P'', ξ'') with distinguished points on the interval $[a, b]$ with

$$\lambda(P'), \lambda(P'') < \delta$$

In words, this means that for any $\epsilon > 0$ that we choose, there always exists a $\delta > 0$ such that **any** two Riemann sums with mesh size **both** smaller than δ will have an error difference of less than ϵ .



Theorem 6.2 (Necessary Condition for Integrability). *A necessary condition for f defined on a closed interval $[a, b]$ to be Riemann integrable on $[a, b]$ is that f be bounded on $[a, b]$. That is,*

$$f \in \mathcal{R}[a, b] \implies f \text{ is bounded on } [a, b]$$

We can clearly see the necessity of f being bounded by looking at the contrapositive of the following statement.

Theorem 6.3 (Refinement). *Given a partition P on interval $[a, b]$, recall that we have points x_0, \dots, x_n such that*

$$a = x_0 < x_1 < \dots < x_n = b$$

Here we introduce new notation:

1. Δ_i denotes the interval $[x_{i-1}, x_i]$
2. Δx_i denotes the difference $x_i - x_{i-1}$, i.e. the length of Δ_i

If a partition \tilde{P} of the closed interval $[a, b]$ is obtained from the partition P by the addition of new points to P , we call \tilde{P} a refinement of P .

When a refinement \tilde{P} of a partition P is constructed, some (perhaps all) of the closed intervals $\Delta_i = [x_{i-1}, x_i]$ of the partition P themselves undergo partitioning.

$$x_{i-1} = x_{i0} < x_{i1} < \dots < x_{in_i} = x_i$$

In that connection, it will be useful to label the points of \tilde{P} by double indices, where in the notation x_{ij} the first index i means that

$$x_{ij} \in \Delta_i = [x_{i-1}, x_i]$$

and the second index j is the ordinal number of the point on the closed interval $\Delta_i = [x_{i-1}, x_i]$. Therefore, it is natural to set the notations

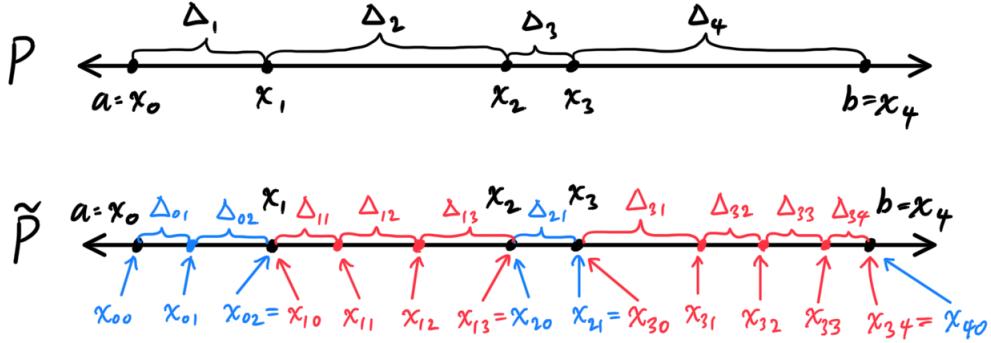
$$1. \Delta_{ij} = [x_{ij-1}, x_{ij}]$$

$$2. \Delta x_{ij} = x_{ij} - x_{ij-1}$$

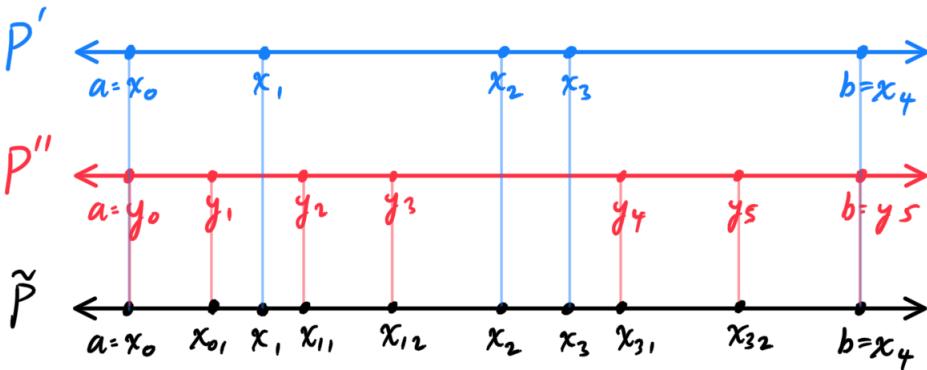
This means that

$$\Delta x_i = \Delta x_{i1} + \Delta x_{i2} + \dots + \Delta x_{in_i}$$

which can be visualized below



Example 6.1 (Union of Partitions as a Refinement). For some interval $[a, b]$, given partitions P' ($a = x_0 < \dots < x_n = b$) and P'' ($a = y_0 < \dots < y_n = b$), the union of the two partitions $\tilde{P} = P' \cup P''$ is a refinement of both P' and P'' .



Recall that $\omega(f; E)$ denotes the oscillation of the function f on the set E ; that is,

$$\omega(f; E) \equiv \sup_{x', x'' \in E} |f(x') - f(x'')|$$

In particular, $\omega(f; \Delta_i)$ is the oscillation of f on the closed interval Δ_i .

Theorem 6.4 (Sufficient Condition for Integrability). Let f be a bounded on a closed interval $[a, b]$ such that for every $\epsilon > 0$ there exists a number $\delta > 0$ such that

$$\sum_{i=1}^n \omega(f; \Delta_i) \Delta x_i < \epsilon$$

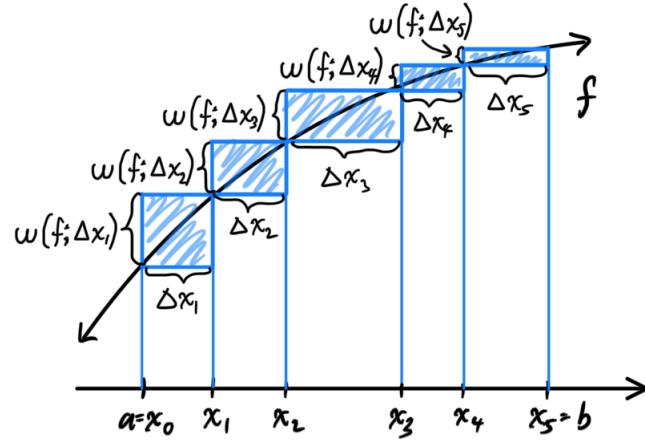
for any partition P of $[a, b]$ with mesh $\lambda(P) < \delta$. This is equivalent to saying that

$$\lim_{\lambda(P) \rightarrow 0} \sum_{i=1}^n \omega(f; \Delta_i) \Delta x_i = 0$$

Then, f is integrable. We can visualize

$$\sum_{i=1}^n \omega(f; \Delta_i) \Delta x_i$$

as the following sum of rectangles below.



What the theorem states, visually, is that as we make all the rectangles smaller and smaller (by putting a limit on the mesh $\lambda(P) < \delta$), we can make the sum of all these rectangles also arbitrarily small.

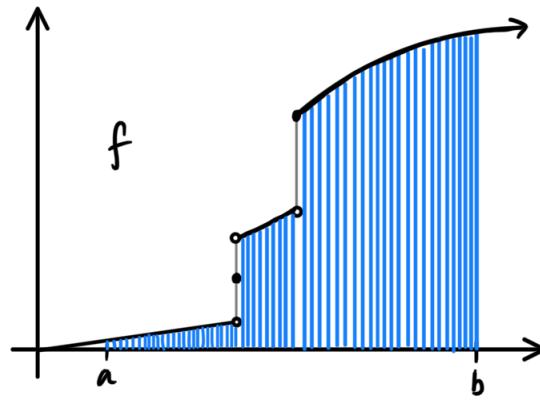
Corollary 6.4.1 (Integrability of Continuous Functions). *Every continuous function on a closed interval is integrable on that closed interval. That is,*

$$f \in C[a, b] \implies f \in \mathcal{R}[a, b]$$

We can actually make a stronger claim.

Corollary 6.4.2 (Integrability of Discontinuous Functions). *If a bounded function f on a closed interval $[a, b]$ is continuous everywhere except at a finite set of points, then $f \in \mathcal{R}[a, b]$.*

Corollary 6.4.3 (Integrability of Monotonic Functions). *A bounded monotonic function on a closed interval is integrable on that interval.*



Definition 6.5 (Upper, Lower Riemann Sums). Let $f : [a, b] \rightarrow \mathbb{R}$ be a real-valued function that is defined and bounded on the closed interval $[a, b]$, and let P be a partition of $[a, b]$, and let Δ_i ($i = 1, 2, \dots, n$) be the intervals of the partition P . Let

$$m_i = \inf_{x \in \Delta_i} f(x)$$

$$M_i = \sup_{x \in \Delta_i} f(x)$$

be the infimum and supremum of f over Δx_i . Then, the sums

$$s(f; P) \equiv \sum_{i=1}^n m_i \Delta x_i$$

$$S(f; P) \equiv \sum_{i=1}^n M_i \Delta x_i$$

are respectively called the *lower* and *upper Riemann sums* of the function f on the interval $[a, b]$ corresponding to the partition P of that interval.

Given an arbitrary partition (P, ξ) with distinguished points on $[a, b]$, it is clear that

$$s(f; P) = \inf_{\xi} \sigma(f; P, \xi) \leq \sigma(f; P, \xi) \leq \sup_{\xi} \sigma(f; P, \xi) = S(f; P)$$

Theorem 6.5. A bounded real-valued function $f : [a, b] \rightarrow \mathbb{R}$ is Riemann integrable on $[a, b]$ if and only if the following limits exist and are equal to each other.

$$\underline{I} \equiv \lim_{\lambda(P) \rightarrow 0} s(f; P) = \lim_{\lambda(P) \rightarrow 0} S(f; P) \equiv \bar{I}$$

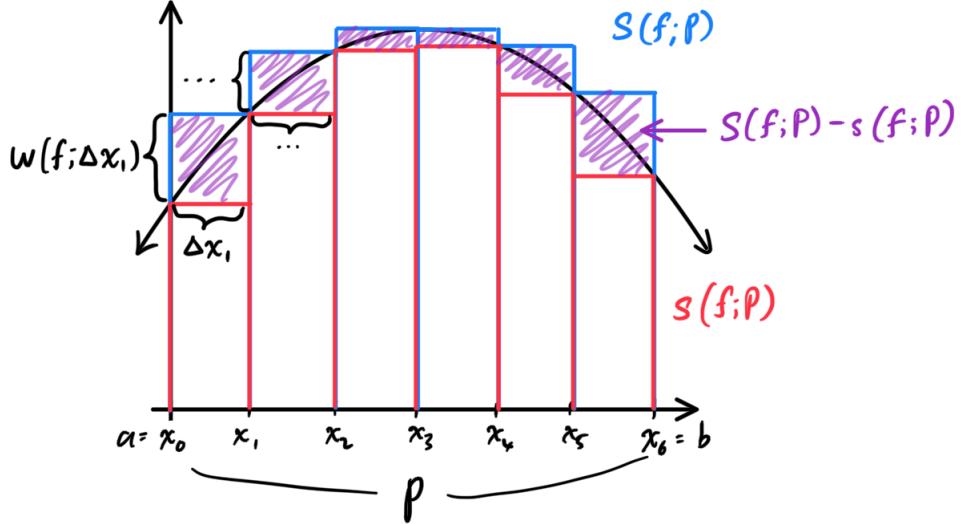
When the relation is true, then the integral is this common value.

$$\int_a^b f(x) dx = \underline{I} = \bar{I}$$

Note that this condition of the upper and lower Riemann sums converging to the same value and the condition that

$$\lim_{\lambda(P) \rightarrow 0} \sum_{i=1}^n \omega(f; \Delta_i) \Delta x_i = 0$$

are the same. For we can see that the rectangles visualized from the equation above are the exact same rectangles formed by $S(f; P) - s(f; P)$!



6.1.2 The Vector Space of Riemann Integrable Functions

Theorem 6.6 (The Vector Space of Integrable Functions). *The set of Riemann integrable functions $\mathcal{R}[a, b]$ over closed interval $[a, b]$ is a vector space. That is, given $f, g \in \mathcal{R}[a, b]$ and $\alpha \in \mathbb{R}$, then*

1. $(f + g) \in \mathcal{R}[a, b]$
2. $(\alpha f) \in \mathcal{R}[a, b]$

Furthermore,

1. $|f| \in \mathcal{R}[a, b]$
2. The restriction of f in any $[c, d] \subset [a, b]$, denoted $f|_{[c, d]}$, is in $\mathcal{R}[c, d]$
3. $(f \cdot g) \in \mathcal{R}[a, b]$

Proof. ■

6.1.3 Lebesgue's Criterion for Riemann Integrability

We give Lebesgue's version of an intrinsic description of a Riemann integrable function.

Definition 6.6 (Measure). A set $E \subset \mathbb{R}$ has (*Lebesgue*) measure zero if for every number $\epsilon > 0$ there exists a covering of the set E by an at most countable system $\{I_k\}$ of intervals, the sum of whose lengths

$$\sum_{k=1}^{\infty} |I_k| \leq \epsilon$$

This means that the above series summing up the lengths of the intervals is an absolutely convergent series.

Lemma 6.7. We can deduce measures of basic sets.

1. A finite number of points are sets of measure zero.
2. The union of a finite or countable number of sets of measure zero is a set of measure zero.
3. A subset of a set of measure zero is itself a set of measure zero.
4. A closed interval $[a, b]$ with $a < b$ is not a set of measure zero.

Definition 6.7. If a property holds at all points of a set X except possibly the points of a set of measure zero, we say that this property holds *almost everywhere on X* or *at almost every point of X* .

Now, we can state Lebesgue's criterion for integrability, which nicely summarizes what we have so far.

Theorem 6.8 (Lebesgue's Criterion for Integrability). *A function defined on a closed interval is Riemann integrable on that interval if and only if it is bounded and continuous at almost every point.*

Example 6.2 (Non-Integrability of the Dirichlet Function). *The Dirichlet function*

$$\mathcal{D}(x) \equiv \begin{cases} 1, & \text{for } x \in \mathbb{Q} \\ 0, & \text{for } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

on the interval $[0, 1]$ is not integrable on that interval. We state two different reasons why.

1. For any partition P of $[0, 1]$ we can find in each interval Δ_i both a rational point ξ'_i and an irrational point ξ''_i . Then, we can see that the lower and upper Riemann sums do not necessarily converge to each other since

$$\sigma(f; P, \xi') = \sum_{i=1}^n 1 \cdot \Delta x_i = 1 \text{ while } \sigma(f; P, \xi'') = \sum_{i=1}^n 0 \cdot \Delta x_i = 0$$

as $\lambda(P) \rightarrow 0$.

2. From the point of view of the Lebesgue criterion the nonintegrability of the Dirichlet function is obvious since $\mathcal{D}(x)$ is discontinuous at every point of $[0, 1]$, which is not a set of measure zero.

Notice that by the Lebesgue criterion, integrability is a weaker condition than continuity. That is,

$$f \text{ continuous} \implies f \text{ Riemann integrable}$$

but not necessarily the other way around. It turns out that this has consequences when determining the composition of functions.

Proposition 6.9 (Integrable + Continuous Composition). *Let $f : I_1 = [a, b] \rightarrow \mathbb{R}$ be a function that is integrable on $[a, b]$, with $\text{Im } f = [c, d] = I_2$. Define a continuous (remember, continuity is stronger than integrability) function $g : [c, d] \rightarrow \mathbb{R}$. Then the composition*

$$g \circ f : [a, b] \rightarrow \mathbb{R}$$

is clearly defined and continuous at all the points of $[a, b]$ where f is continuous. But since f is integrable, the union of all the discontinuities in $[a, b]$ must have measure zero, and so it follows that since $[a, b]$ is the same

$$g \circ f \in \mathcal{R}[a, b]$$

Therefore, we can find out that

$$f \text{ integrable and } g \text{ continuous} \implies g \circ f \text{ integrable}$$

as visualized in the commutative diagram below.

$$\begin{array}{ccc} & g \circ f & \\ I_1 & \xrightarrow{f} & I_2 \xrightarrow{g} \mathbb{R} \end{array}$$

However, contrary to intuition,

$$f \text{ integrable and } g \text{ integrable} \not\implies g \circ f \text{ integrable}$$

We present a counterexample.

Example 6.3. Consider the functions

$$|\text{sgn}|(x) \equiv \begin{cases} 1 & x \neq 0 \\ 0 & x = 0 \end{cases}$$

and the Riemann function

$$\mathcal{R}(x) \equiv \begin{cases} \frac{1}{n} & x = \frac{m}{n} \in \mathbb{Q}, \gcd(m, n) = 1 \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

We can see that \mathcal{R} is continuous at all irrational points and discontinuous at all rational points except 0, meaning that it is integrable (\mathbb{Q} has measure zero). Then, the composition of these two functions is precisely the Dirichlet function

$$\mathcal{D}(x) = |\text{sgn}| \circ \mathcal{R}$$

which is not integrable.

6.2 Basic Properties of the Integral

One of the most basic properties of the integral is that it is a linear map.

Lemma 6.10 (Linearity of the Integral). *Given closed interval $[a, b] \subset \mathbb{R}$, the Riemann integration function*

$$\int_a^b : \mathcal{R}[a, b] \rightarrow \mathbb{R}$$

is a linear functional living within the dual space $\mathbb{R}^[a, b]$. That is, given $f, g \in \mathcal{R}[a, b]$, a linear combination of them $\alpha f + \beta g$ is also integrable on $[a, b]$, and*

$$\int_a^b (\alpha f + \beta g)(x) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx$$

Proof. It is clear from basic algebraic transformation that the Riemann sums for the integral expressions on both sides are equal.

$$\sum_{i=1}^n (\alpha f + \beta g)(\xi_i) \Delta x_i = \alpha \sum_{i=1}^n f(\xi_i) \Delta x_i + \beta \sum_{i=1}^n g(\xi_i) \Delta x_i$$

Taking the limit as $\lambda(P) \rightarrow 0$ on both sides leads to the respective Riemann integrals. ■

The next property of the Riemann integral is its additive property *on the interval of integration*. Note that the value of the integral

$$\int_a^b f(x) dx \equiv \lim_{\lambda(P) \rightarrow 0} \sigma(f; P, \xi)$$

depends on both the integrand and the closed interval over which the integral is taken.

Lemma 6.11 (Properties of the Interval of Integration). *If $a < b < c$ and $f \in \mathcal{R}[a, c]$, then $f|_{[a,b]} \in \mathcal{R}[a, b]$, $f|_{[b,c]} \in \mathcal{R}[b, c]$, and the following equality holds*

$$\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx$$

From these we set

$$\int_a^b f(x) dx \equiv - \int_b^a f(x) dx$$

and

$$\int_a^a f(x) dx \equiv 0$$

Theorem 6.12 (Symmetry of the Riemann Integral). *Let $a, b, c \in \mathbb{R}$ and let f be integrable over the largest closed interval having two of these points as endpoints. Then, the restriction of f to each of the other closed intervals is also integrable over those intervals and the following equality holds.*

$$\int_a^b f(x) dx + \int_b^c f(x) dx + \int_c^a f(x) dx = 0$$

This property can be abstractified to those of additive interval functions, which will be shown soon.

We finally end with an important property of the integral which, as seen later, allows us to define inner products on function spaces.

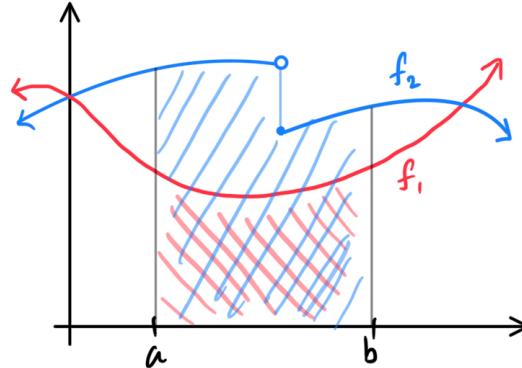
Theorem 6.13. *If $a \leq b$ and $f \in \mathcal{R}[a, b]$, then $|f| \in \mathcal{R}[a, b]$, and*

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f|(x) dx$$

6.2.1 Mean Value Theorem of the Integral

Lemma 6.14 (Monotonicity of the Integral). *If $a \leq b$, $f_1, f_2 \in \mathcal{R}[a, b]$, and $f_1(x) \leq f_2(x)$ for every $x \in [a, b]$, then*

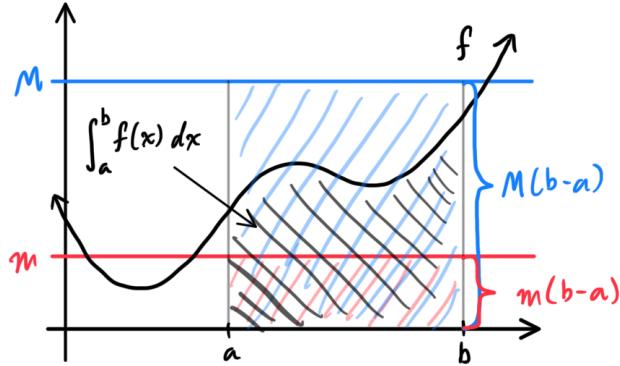
$$\int_a^b f_1(x) dx \leq \int_a^b f_2(x) dx$$



This immediately implies that given constants m, M such that $m \leq f(x) \leq M$ at each $x \in [a, b]$, we have

$$m \cdot (b - a) \leq \int_a^b f(x) dx \leq M \cdot (b - a)$$

This is very easily visualized below.



In particular, if $0 \leq f(x)$ on $[a, b]$, then

$$0 \leq \int_a^b f(x) dx$$

Theorem 6.15 (Mean Value Theorem of the Integral). Given $f \in \mathcal{R}[a, b]$, with

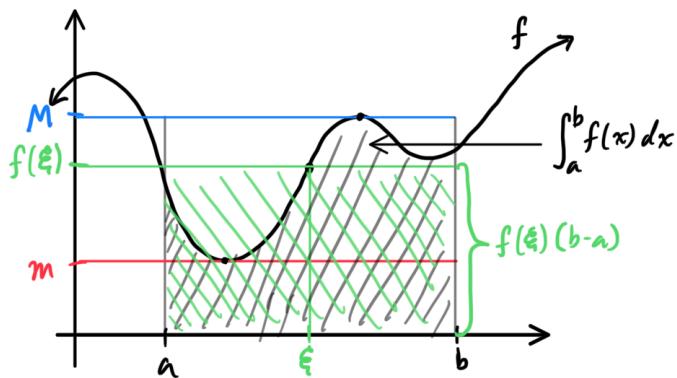
$$m = \inf_{x \in [a, b]} f(x) \text{ and } M = \sup_{x \in [a, b]} f(x)$$

then there exists a number $\mu \in [m, M]$ such that

$$\int_a^b f(x) dx = \mu \cdot (b - a)$$

Furthermore, if $f \in C[a, b]$ (that is, continuous on $[a, b]$), it immediately follows by the intermediate value theorem that there exists a point $\xi \in [a, b]$ such that

$$\int_a^b f(x) dx = f(\xi)(b - a)$$



Due to the length of the proof, we ask the reader to take it for granted the following theorem.

Theorem 6.16 (Bonnet's Formula). If $f, g \in \mathcal{R}[a, b]$ and g is a monotonic function on $[a, b]$, then there exists a point $\xi \in [a, b]$ such that

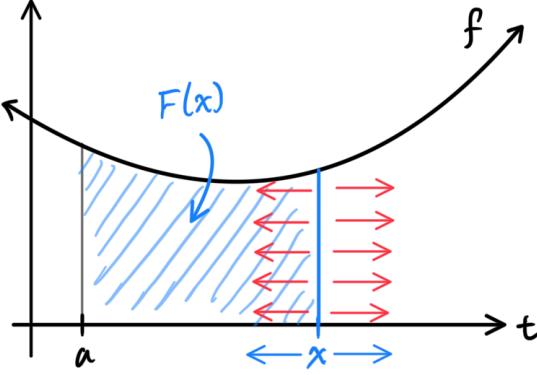
$$\int_a^b (f \cdot g)(x) dx = g(a) \int_a^\xi f(x) dx + g(b) \int_\xi^b f(x) dx$$

6.3 Connections between Integrals, Primitives, Derivatives

Definition 6.8 (Integral with Variable Upper Limit). Let $f \in \mathcal{R}[a, b]$, and let us choose an $x \in [a, b]$ in order to construct the function

$$F(x) \equiv \int_a^x f(t) dt$$

which is called an *integral with variable upper limit*. Note that since $[a, x] \subset [a, b]$, it follows that $f|_{[a, x]} \in \mathcal{R}[a, x]$ and therefore the function $x \mapsto F(x)$ is unambiguously defined for $x \in [a, b]$.



Furthermore, $F(x)$ is continuous on $[a, b]$. Since f is integrable on $[a, b]$, it is bounded by a constant C such that

$$|f(t)| \leq C \text{ on } [a, b]$$

It follows from the additive properties of the integral and boundedness theorem that

$$|F(x+h) - F(x)| \leq C|h|$$

if $x, x+h \in [a, b]$, as visualized. This means that for any δ -neighborhood of $F(x)$, we can find an arbitrary small h such that the $C|h|$ -neighborhood of $F(x)$ is completely contained in the δ -neighborhood. But by the inequality above, this means that there exists an $\epsilon = h$ -neighborhood of x such that its entire image is contained within the $C|h|$ -neighborhood, which itself is contained within the δ -neighborhood. This shows that F is continuous.

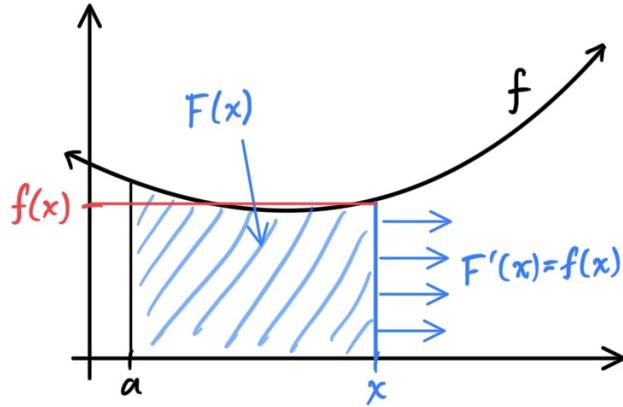
Theorem 6.17 (First Fundamental Theorem of Calculus). *Let $f \in \mathcal{R}[a, b]$ be continuous at point $x \in [a, b]$ (resp. continuous on closed interval $[a, b]$). Let F be the function, defined for all $x \in [a, b]$ by*

$$F(x) \equiv \int_a^x f(t) dt$$

Then, f is continuous and differentiable at x (resp. uniformly continuous on $[a, b]$ and differentiable on (a, b)),

$$F'(x) = f(x)$$

at x (resp. for all $x \in [a, b]$). This is an amazing fact, because visually, it tells us that the rate at which the integral F is increasing at x (represented by the increasing area under the curve of f) is equal to the value of f at the point x itself!



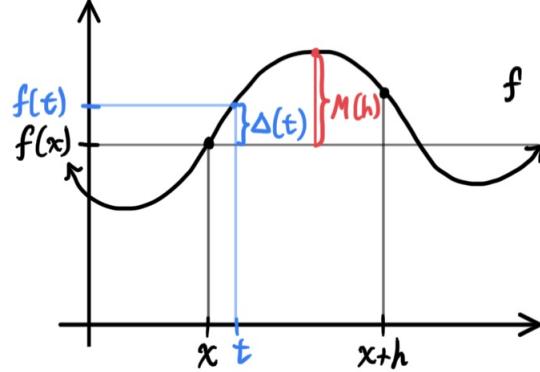
Proof. Let $x, x+h \in [a, b]$, and let us estimate the difference $F(x+h) - F(x)$. It follows from the continuity of f at x that $f(t) = f(x) + \Delta(t)$, where $\Delta(t) \rightarrow 0$ as $t \rightarrow x$. If point x is held fixed, the function

$$\Delta(t) = f(t) - f(x)$$

is integrable on $[a, b]$, being the difference of the integrable function $t \mapsto f(t)$ and the constant $f(x)$. Let us denote

$$M(h) \equiv \sup_{t \in [x, x+h]} |\Delta(t)|$$

which means that $M(h)$ is the largest difference between $f(x)$ and $f(t)$ in the interval $[x, x+h]$.



Clearly $M(h) \rightarrow 0$ as $h \rightarrow 0$. We can now find

$$\begin{aligned} F(x+h) - F(x) &= \int_a^{x+h} f(t) dt - \int_a^x f(t) dt \\ &= \int_x^{x+h} f(t) dt \\ &= \int_x^{x+h} (f(x) + \Delta(t)) dt \\ &= \int_x^{x+h} f(x) dt + \int_x^{x+h} \Delta(t) dt \\ &= f(x)h + \alpha(h)h \end{aligned}$$

where we have set

$$\int_x^{x+h} \Delta(t) dt = \alpha(h)h$$

where α is infinitesimal as $h \rightarrow 0$, since

$$\left| \int_x^{x+h} \Delta(t) dt \right| \leq \left| \int_x^{x+h} |\Delta(t)| dt \right| \leq \left| \int_x^{x+h} M(h) dt \right| = M(h)|h| = \alpha(h)|h|$$

Therefore, we have shown that if the function f is continuous at a point $x \in [a, b]$, then for displacements h from x such that $x + h \in [a, b]$, the following equality holds.

$$F(x + h) - F(x) = f(x)h + \alpha(h)h$$

where $\alpha(h) \rightarrow 0$ as $h \rightarrow 0$, and by definition, this means that $F(x)$ is differentiable on $[a, b]$ at the point $x \in [a, b]$ and that $F'(x) = f(x)$. \blacksquare

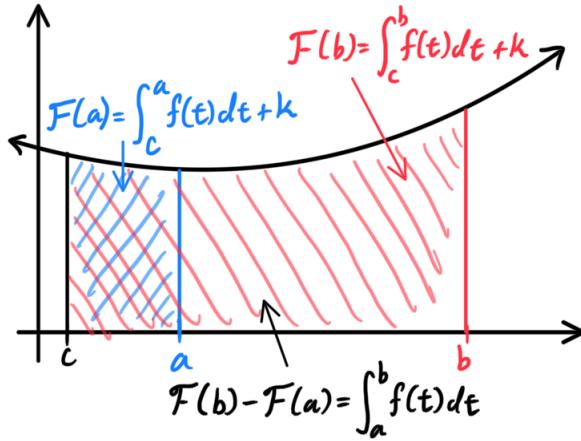
Corollary 6.17.1. *Every bounded function $f : [a, b] \rightarrow \mathbb{R}$ on the closed interval $[a, b]$ and has only a finite number of points of discontinuity has a primitive, and every primitive of f on $[a, b]$ has the form*

$$\mathcal{F}(x) \equiv \int_a^x f(t) dt + c$$

where c is a constant.

Theorem 6.18 (Second Fundamental Theorem of Calculus). *Let f be a real-valued function on a closed interval $[a, b]$ with \mathcal{F} any primitive of f on $[a, b]$. If f is Riemann-integrable (i.e. f bounded with finite points of Lebesgue measure zero) on $[a, b]$, then*

$$\int_a^b f(x) dx = \mathcal{F}|_a^b \equiv \mathcal{F}(b) - \mathcal{F}(a)$$



Proof. We already know that a bounded function on a closed interval having a finite number of discontinuities is integrable, and by the corollary, we are guaranteed an existence of a primitive $\mathcal{F}(x)$ of the function f on $[a, b]$ with the form

$$\mathcal{F}(x) \equiv \int_a^x f(t) dt + c$$

Setting $x = a$, we find that $c = \mathcal{F}(a)$, and so

$$\mathcal{F}(x) \equiv \int_a^x f(t) dt + \mathcal{F}(a)$$

Evaluating \mathcal{F} at $x = b$ gives

$$\int_a^b f(t) dt = \mathcal{F}(b) - \mathcal{F}(a)$$

\blacksquare

6.3.1 Integration by Parts and Taylor's Formula

Theorem 6.19 (Definite Integration by Parts). *If the functions $u(x)$ and $v(x)$ are continuously differentiable on a closed interval with endpoints a and b , then*

$$\int_a^b (u \cdot v')(x) dx = (u \cdot v)|_a^b - \int_a^b (v \cdot u')(x) dx$$

which is customarily written in the form as

$$\int_a^b u \, dv = u \cdot v \Big|_a^b - \int_a^b v \, du$$

Proof. By the product rule of differentiation, we have

$$(u \cdot v)'(x) = (u' \cdot v)(x) + (u \cdot v')(x)$$

where by hypothesis, $u' \cdot v, u \cdot v'$ are continuous and hence integrable on $[a, b]$. Using the linearity of the integral and the 2nd fundamental theorem of calculus, we get

$$(u \cdot v)(x) \Big|_a^b = \int_a^b (u' \cdot v)(x) \, dx + \int_a^b (u \cdot v')(x) \, dx$$

■

Theorem 6.20 (Integral Form of the Remainder). *If $f : E \rightarrow \mathbb{R}$ has continuous derivatives up to order n on the closed interval $[a, x]$, then Taylor's formula holds*

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \dots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1} + r_{n-1}(a; x)$$

where

$$r_{n-1}(a; x) = \frac{1}{(n-1)!} \int_a^x f^{(n)}(t)(x-t)^{n-1} \, dt$$

This form is called Taylor's formula with the integral form of the remainder.

Proof. Using the 2nd fundamental theorem and the definite integration by parts formula, we can carry out the following chain of transformations, assuming continuity and differentiability when needed.

$$\begin{aligned} f(x) - f(a) &= \int_a^x f'(t) \, dt \\ &= - \int_a^x f'(t)(x-t)' \, dt \\ &= -f'(t)(x-t) \Big|_a^x + \int_a^x f''(t)(x-t) \, dt \\ &= f'(a)(x-a) - \frac{1}{2} \int_a^x f''(t)((x-t)^2)' \, dt \\ &= f'(x-a) - \frac{1}{2} f''(t)(x-t)^2 \Big|_a^x + \frac{1}{2} \int_a^x f'''(t)(x-t)^2 \, dt \\ &= f'(a)(x-a) + \frac{1}{2} f''(a)(x-a)^2 - \frac{1}{2 \cdot 3} \int_a^x f'''(t)((x-t)^3)' \, dt \\ &= \dots \\ &= f'(a)(x-a) + \dots + \frac{1}{(n-1)!} f^{(n-1)}(a)(x-a)^{n-1} + r_{n-1}(a; x) \end{aligned}$$

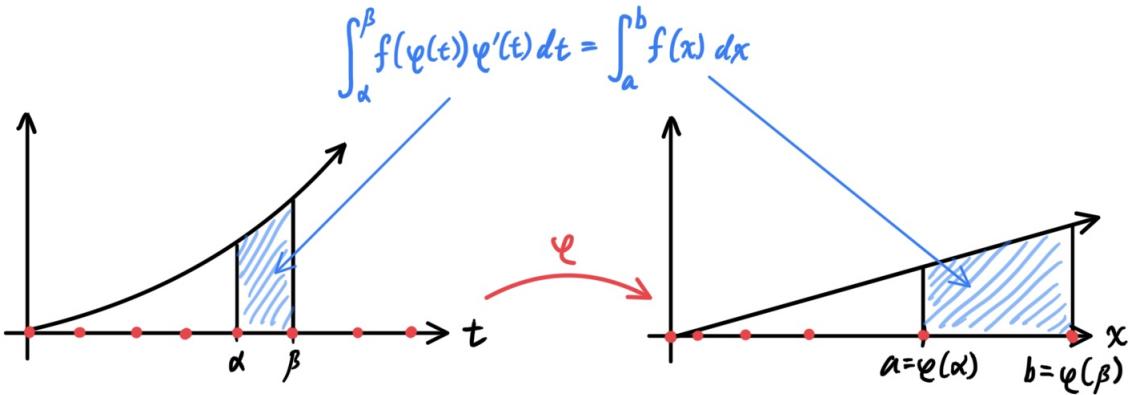
where $r_{n-1}(a; x)$ is given by the integral formula mentioned. ■

6.3.2 Change of Variables in Integration

We now show and prove the method what we call "u-substitution" for definite integration.

Theorem 6.21 (Change of Variable). *If $\varphi : [\alpha, \beta] \rightarrow [a, b]$ is a continuously differentiable mapping such that $\varphi(\alpha) = a$ and $\varphi(\beta) = b$, then for any continuous function $f(x)$ on $[a, b]$ the function $f(\varphi(t))\varphi'(t)$ is continuous on the closed interval $[\alpha, \beta]$ and*

$$\int_a^b f(x) \, dx = \int_\alpha^\beta f(\varphi(t))\varphi'(t) \, dt$$



Proof. We prove a slightly weaker form of the theorem with the additional hypothesis that φ is strictly monotonic. ■

6.3.3 Additive Interval Functions and the Integral

In this section we take a step back and construct the integral in a more abstract sense, using the concepts of an additive interval function.

Definition 6.9 (Additive Interval Function). An *additive (oriented) interval function* is a function

$$(\alpha, \beta) \mapsto I(\alpha, \beta) \in \mathbb{R}$$

that assigns a number $I(\alpha, \beta)$ to each ordered pair of points (α, β) of a fixed closed interval $[a, b]$ in such a way that the following equality holds for any triple of points $\alpha, \beta, \gamma \in [a, b]$.

$$I(\alpha, \gamma) = I(\alpha, \beta) + I(\beta, \gamma)$$

Notice that the integral holds this property, shown in the theorem on the symmetric property of the integral. It follows that all additive interval functions are anticommutative:

$$I(\alpha, \beta) + I(\beta, \alpha) = 0$$

which immediately results in

$$I(\alpha, \alpha) = 0$$

Lemma 6.22 (Generating Functions of Additive Interval Functions). *For any function $x \mapsto \mathcal{F}(x)$ that maps points on the interval $[a, b]$ to \mathbb{R} , we set*

$$\mathcal{F}(x) \equiv I(a, x)$$

and by additivity we have

$$I(\alpha, \beta) = I(\alpha, \beta) - I(a, \alpha) = \mathcal{F}(\beta) - \mathcal{F}(\alpha)$$

and thus, every additive oriented interval function has the form

$$I(\alpha, \beta) = \mathcal{F}(\beta) - \mathcal{F}(\alpha)$$

By constructing I in this manner, we say that the function \mathcal{F} generates the additive function I .

Example 6.4. If $f \in \mathcal{R}[a, b]$, the function $\mathcal{F} = \int_a^x f(t) dt$ generates the additive function

$$I(\alpha, \beta) = \mathcal{F}(\beta) - \mathcal{F}(\alpha) = \int_a^\beta f(t) dt - \int_a^\alpha f(t) dt = \int_\alpha^\beta f(t) dt$$

We conclude by stating a sufficient condition for an additive interval function to be generated by an integral.

Theorem 6.23. Suppose the additive function $I(\alpha, \beta)$ defined for points $\alpha, \beta \in [a, b]$ has the property that, for some known function $f \in \mathcal{R}[a, b]$,

$$\inf_{x \in [\alpha, \beta]} f(x)(\beta - \alpha) \leq I(\alpha, \beta) \leq \sup_{x \in [\alpha, \beta]} f(x)(\beta - \alpha)$$

holds for any closed interval $[\alpha, \beta] \subset [a, b]$ ($\alpha \leq \beta$). Then, the additive function I must be the definite integral

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

This theorem is extremely useful. It says that if we have any abstract additive interval function $I(\alpha, \beta)$ that satisfies the properties above, then it **must** be generated by an integral with variable upper limit, meaning that (by the previous example) I itself must be a definite integral!

6.3.4 Arc Length

When modeling systems in physics, one of the most fundamental tools we use are path functions that models the movement of a particle in \mathbb{R}^3 .

Definition 6.10 (Path). A *path* in \mathbb{R}^3 is a continuous mapping $r : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^3$ defined

$$t \mapsto (x(t), y(t), z(t))$$

of an interval of the real line into \mathbb{R}^3 defined by the (continuous) scalar functions x, y, z . The endpoints

$$A = (x(a), y(a), z(a)) \text{ and } B = (x(b), y(b), z(b))$$

in \mathbb{R}^3 are called the *initial point* and *terminal point* of the path. Furthermore, a path is *closed* if its initial and terminal points coincide.

Definition 6.11 (Support). If $\Gamma : I \rightarrow \mathbb{R}^3$ is a path, the image $\Gamma(I) \subset \mathbb{R}^3$ is called the *support* of the path.

Definition 6.12 (Simple Paths). A path $\Gamma : I \rightarrow \mathbb{R}^3$ that is injective is called a *simple path*, or a *paramaterized curve*, and its support is called a *curve* in \mathbb{R}^3 .

A closed path $\Gamma : [a, b] \rightarrow \mathbb{R}^3$ is called a *simple closed path/curve* if the path $\Gamma : [a, b] \rightarrow \mathbb{R}^3$ is simple.

Definition 6.13 (Smooth Paths). A path $\Gamma : [a, b] \rightarrow \mathbb{R}^3$ is C^k smooth if the functions $x(t), y(t), z(t)$ are C^k smooth. Γ is *piecewise smooth* if the closed interval $[a, b]$ can be partitioned into a finite number of closed intervals on each of which the corresponding restriction of Γ is smooth.

Now, we are ready to construct the length of a smooth path $\Gamma : [a, b] \rightarrow \mathbb{R}^3$. Our initial ideas about the length $l[a, b]$ of the path traversed during the time interval $\alpha \leq t \leq \beta$ are as follows:

1. If $\alpha < \beta < \gamma$, then l is an additive interval function.

$$l[\alpha, \gamma] = l[\alpha, \beta] + l[\beta, \gamma]$$

2. If $v(t) = (x'(t), y'(t), z'(t))$ is the velocity of the point at time t , then

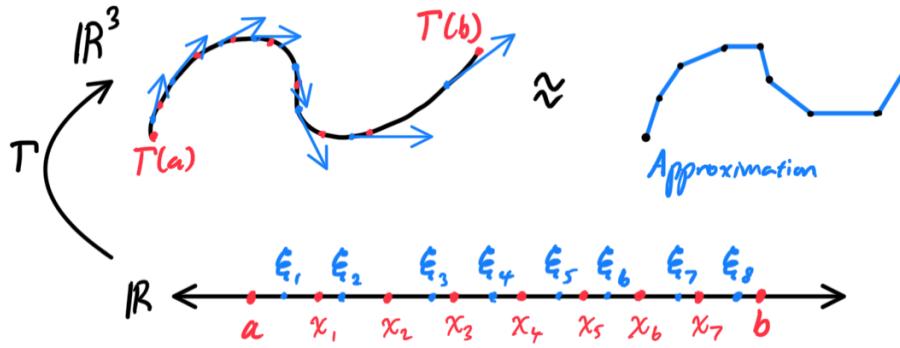
$$\int_{x \in [\alpha, \beta]} |v(t)|(\beta - \alpha) \leq l[\alpha, \beta] \leq \sup_{x \in [\alpha, \beta]} |v(t)|(\beta - \alpha)$$

Thus, if the functions x, y, z are continuously differentiable on $[a, b]$, this is sufficient condition (by the theorem in the previous subsection) that the additive function l is an integral.

Definition 6.14 (Arc Length Integral). The length of a smooth path $\Gamma : [a, b] \rightarrow \mathbb{R}^3$ is defined by

$$l[a, b] \equiv \int_a^b |\Gamma'(t)| dt \equiv \int_a^b \sqrt{x'^2(t) + y'^2(t) + z'^2(t)} dt$$

We can visualize this by partitioning the interval $[a, b]$ into the intervals Δ_i , each with point $\xi_i \in \Delta_i$. This would partition the path to $\Gamma(\Delta_i)$, each with points $\Gamma(\xi_i)$, and at each point $\Gamma(\xi_i)$, we can imagine the velocity vector of the curve. By taking the magnitude of this vector $\Gamma'(\xi_i)$, we multiply it by the length of the interval Δx_i to get one rectangle, creating an approximation for one partition of the path.



An immediate result of this formula is the formula for the length of a graph of a function $f : [a, b] \rightarrow \mathbb{R}$ in \mathbb{R}^2 , by looking at the parameterization $t \mapsto \Gamma(t) = (t, f(t))$.

$$l[a, b] \equiv \int_a^b \sqrt{1 + (f'(t))^2} dt$$

The question on the effect of parameterization on the integral now arises.

Definition 6.15 (Admissible Change of Parameter). The path $\tilde{\Gamma} : [\alpha, \beta] \rightarrow \mathbb{R}^3$ is obtained from $\Gamma : [a, b] \rightarrow \mathbb{R}^3$ by an *admissible change of parameter* if there exists a smooth mapping

$$T : [\alpha, \beta] \rightarrow [a, b]$$

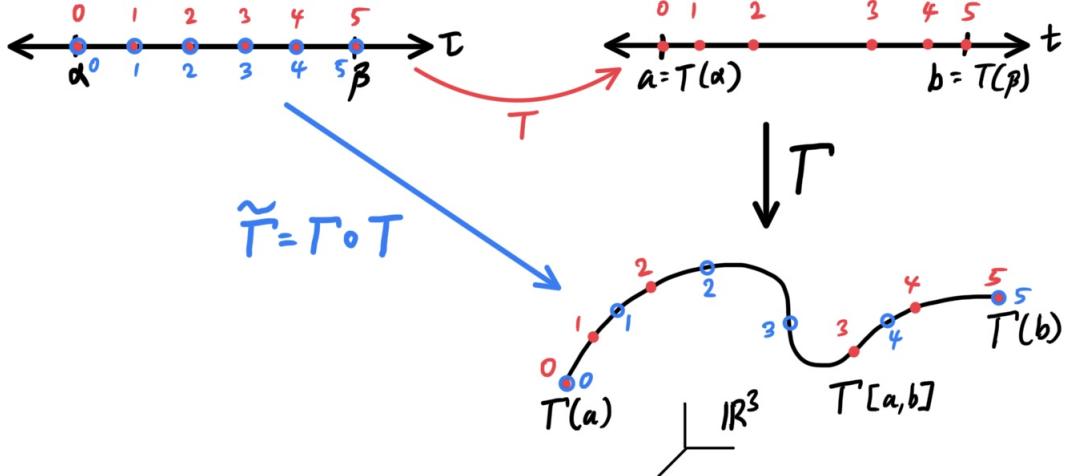
such that $T(\alpha) = a, T(\beta) = b, T'(\tau) > 0$ (that is, the reparameterization T is monotonic) on $[\alpha, \beta]$, and

$$\tilde{\Gamma} = \Gamma \circ T$$

The series of mappings can be represented with the following commutative diagram, where $I_{\alpha, \beta} = [\alpha, \beta] \subset \mathbb{R}$ and $I_{a, b} = [a, b] \subset \mathbb{R}$.

$$\begin{array}{ccc} I_{\alpha, \beta} & \xrightarrow{T} & I_{a, b} \\ & \searrow \tilde{\Gamma} & \downarrow \Gamma \\ & & \mathbb{R}^3 \end{array}$$

or with the more detailed visual below (Note that the points are labeled 0, 1, 2, 3, 4, 5 do not represent numerical values, but rather the order in which the points are parameterized. We can see from this ordering that T is monotonic.)



Theorem 6.24 (Invariance of Arclength Integral under Admissible Change of Parameters). *If a smooth path $\tilde{\Gamma} : [\alpha, \beta] \rightarrow \mathbb{R}^3$ is obtained from a smooth path $\Gamma : [a, b] \rightarrow \mathbb{R}^3$ by an admissible change of parameter, then the lengths of the two paths are equal. That is,*

$$\int_a^b |\Gamma'(t)| dt = \int_\alpha^\beta |\tilde{\Gamma}'(t)| dt \equiv \int_\alpha^\beta |(\Gamma \circ T)'(t)| dt$$

6.4 Improper Integrals

Due to some limitations of the Riemann integral, we cannot integrate over "singularities" where either the interval or the function is unbounded. We develop the tools of improper integration to deal with this problem; there are two types of improper integrals.

Definition 6.16 (Improper Integral of Unbounded Interval). Suppose the function $x \mapsto f(x)$ is defined on the interval $[a, +\infty)$ and is integrable on every closed interval $[a, b]$ contained in that interval. Then, we call the following term

$$\int_a^{+\infty} f(x) dx \equiv \lim_{b \rightarrow +\infty} \int_a^b f(x) dx$$

the *improper Riemann integral of f over the interval $[a, +\infty)$* and

$$\int_{-\infty}^b f(x) dx \equiv \lim_{a \rightarrow -\infty} \int_a^b f(x) dx$$

the *improper Riemann integral of f over the interval $(-\infty, b]$* . If the limit exists, then we say that the integral *converges* and *diverges* otherwise.

Definition 6.17 (Improper Integral of Unbounded Function). Suppose the function $x \mapsto f(x)$ is defined on the interval $[a, B)$ and integrable on any closed interval $[a, b] \subset [a, B)$. Then, we call the following term

$$\int_a^B f(x) dx \equiv \lim_{b \rightarrow B^-} \int_a^b f(x) dx$$

the *improper Riemann integral of f over interval $[a, B)$* and

$$\int_A^b f(x) dx \equiv \lim_{a \rightarrow A^+} \int_a^b f(x) dx$$

the *improper Riemann integral of f over interval $(A, b]$* .

For cohesiveness, we can combine these two definitions of improper integrals into the following one.

Definition 6.18 (Improper Integrals). Let $[a, \omega)$ be a finite or infinite interval and $x \mapsto f(x)$ a function defined on that interval and integrable over every closed interval $[a, b] \subset [a, \omega)$. Then, by definition

$$\int_a^\omega f(x) dx \equiv \lim_{b \rightarrow \omega} \int_a^b f(x) dx$$

if this limit exists as $b \rightarrow \omega, b \in [a, \omega)$. Similarly, given the finite or infinite interval $(\omega, b]$ with f integrable over every closed interval $[a, b] \subset (\omega, b]$, we have

$$\int_\omega^b f(x) dx \equiv \lim_{a \rightarrow \omega} \int_a^b f(x) dx$$

Note that if $\omega \in \mathbb{R}$ and $f \in \mathcal{R}[a, \omega]$, the improper integral is equivalent to the regular Riemann integral.

$$\int_a^\omega f(x) dx = \lim_{b \rightarrow \omega} \int_a^b f(x) dx$$

Lemma 6.25 (Properties of the Improper Integral). *Suppose f, g are functions defined on interval $[a, \omega)$ (without loss of generality, we let ω be the upper limit of integration) and integrable on every closed interval $[a, b] \subset [a, \omega)$. Suppose the improper integrals*

$$\int_a^\omega f(x) dx \text{ and } \int_a^\omega g(x) dx$$

are well-defined.

1. For any $\lambda_1, \lambda_2 \in \mathbb{R}$ the function $(\lambda_1 f + \lambda_2 g)(x)$ is integrable in the improper sense on $[a, \omega)$ and

$$\int_a^\omega (\lambda_1 f + \lambda_2 g)(x) dx = \lambda_1 \int_a^\omega f(x) dx + \lambda_2 \int_a^\omega g(x) dx$$

2. For any $c \in [a, \omega)$,

$$\int_a^\omega f(x) dx = \int_a^c f(x) dx + \int_c^\omega f(x) dx$$

3. If $\varphi : [\alpha, \gamma) \rightarrow [a, \omega)$ is a smooth strictly monotonic mapping with $\varphi(\alpha) = a$ and $\varphi(\beta) \rightarrow \omega$ as $\beta \rightarrow \gamma^-$, then the improper integral of the function $t \mapsto (f \circ \varphi)(t)\varphi'(t)$ over $[\alpha, \gamma)$ exists and

$$\int_a^\omega f(x) dx = \int_\alpha^\gamma (f \circ \varphi)(t)\varphi'(t) dt$$

6.4.1 Convergence of an Improper Integral

Note that by definition, an improper integral

$$\int_a^\omega f(x) dx \equiv \lim_{b \rightarrow \omega} \int_a^b f(x) dx$$

is a limit of the function

$$\mathcal{F}(b) \equiv \int_a^b f(x) dx$$

as $b \rightarrow \omega$. This means that we can use the Cauchy criterion to determine the convergence of this limit, and hence, existence of this improper integral.

Theorem 6.26 (Cauchy Criterion for Convergence of an Improper Integral). *If the function $x \mapsto f(x)$ is defined on the interval $[a, \omega)$ and integrable on every closed interval $[a, b] \subset [a, \omega)$, then the integral*

$$\int_a^\omega f(x) dx$$

converges if and only if for every $\epsilon > 0$ there exists $B \in [a, \omega)$ such that the relation

$$\left| \int_{b_1}^{b_2} f(x) dx \right| < \epsilon$$

holds for any $b_1, b_2 \in [a, \omega)$ satisfying $B < b_1$ and $B < b_2$.

Proof. We have

$$\int_{b_1}^{b_2} f(x) dx = \int_a^{b_2} f(x) dx - \int_a^{b_1} f(x) dx = \mathcal{F}(b_2) - \mathcal{F}(b_1)$$

and therefore the condition is simply the Cauchy criterion for the existence of a limit for the function $\mathcal{F}(b)$ as $b \rightarrow \omega$. ■

Definition 6.19 (Absolute Convergence of an Improper Integral). The improper integral

$$\int_a^\omega |f(x)| dx$$

converges absolutely if the integral

$$\int_a^\omega |f|(x) dx$$

converges. Clearly, the inequality

$$\left| \int_{b_1}^{b_2} f(x) dx \right| \leq \left| \int_{b_1}^{b_2} |f|(x) dx \right|$$

implies that if an improper integral converges absolutely, then it converges.

This study of absolute convergence reduces to the study of convergence of integrals of nonnegative functions. The following lemma is useful in determining convergence of such functions.

Lemma 6.27. *Let there be a function f defined on interval $[a, \omega)$ that is also integrable over every closed interval $[a, b] \subset [a, \omega)$. If $f(x) \geq 0$ on $[a, \omega)$, then the improper integral*

$$\int_a^\omega f(x) dx$$

exists if and only if the function

$$\mathcal{F}(b) \equiv \int_a^b f(x) dx$$

is bounded on $[a, \omega)$.

Proof. It is clear that

$$\int_a^\omega f(x) dx = \lim_{b \rightarrow \omega} \mathcal{F}(b)$$

If $f(x) \geq 0$, then the function $\mathcal{F}(b)$ is nondecreasing on $[a, \omega)$ and therefore has a limit as $b \rightarrow \omega$ only if it is bounded (since every monotonically increasing sequence that is bounded always converges). ■

This leads to the familiar integral test for convergence of a series.

Theorem 6.28 (Integral Test for Convergence of a Series). *If the function $x \mapsto f(x)$ is defined on the interval $[1, +\infty)$, nonnegative, nonincreasing, and integrable on each closed interval $[1, b] \subset [1, +\infty)$, then the series*

$$\sum_{n=1}^{\infty} f(n) = f(1) + f(2) + \dots$$

and the integral

$$\int_a^{+\infty} f(x) dx$$

either both converge or both diverge.

We can use the comparison test analogue to determine convergence of improper integrals.

Theorem 6.29 (Comparison Test for Convergence of Improper Integrals). *Suppose the functions $f(x), g(x)$ are defined on the interval $[a, \omega)$ and integrable on any closed interval $[a, b] \subset [a, \omega)$. If*

$$0 \leq f(x) \leq g(x)$$

on $[a, \omega)$, then

$$\int_a^\omega g(x) dx \text{ converges} \implies \int_a^\omega f(x) dx \text{ converges}$$

and the inequality

$$\int_a^\omega f(x) dx \leq \int_a^\omega g(x) dx$$

holds. Also,

$$\int_a^\omega f(x) dx \text{ diverges} \implies \int_a^\omega g(x) dx \text{ diverges}$$

6.4.2 Improper Integrals with Multiple Singularities

Definition 6.20 (Improper Integral with Both Limits as Singularities). Given singularities ω_1, ω_2 , the improper integral is defined

$$\int_{\omega_1}^{\omega_2} f(x) dx \equiv \int_{\omega_1}^c f(x) dx + \int_c^{\omega_2} f(x) dx$$

where c is an arbitrary point in (ω_1, ω_2) .

Example 6.5 (Gaussian Integral). *The integral*

$$\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$