Ch. 0 Mathematical Background

October 31, 2014

Mathematics is very much like poetry $\cdots$. What makes a good poem–a great poem–is that there is a large amount of thought expressed in a very few words. In this sense formula like

$$e^{\pi i} + 1 = 0 \quad or \quad \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

are poems.[1]

– Lipman Bers

---

[1] From Ok, E.F., 2007, *Real Analysis with Economic Application*, Princeton University Press.

1 Sets and Numbers

1.1 The concept of a set

The origin of the modern theory of sets can be traced back to the German mathematician Georg Cantor (1854-1918).

Definition (Set):

A set is any collection of well-defined and distinguishable objects. These objects are called elements, or members, of the set. Thus if x is an element of a set A, then this fact is denoted by writing $x \in A$. If, however, x is not an element of A, then we write $x \notin A$.

Curly brackets are usually used to describe the contents of a set. For example, if a set A consists of the element $x_1, x_2, ..., x_n$, then it can be represented as $A = \{x_1, x_2, ..., x_n\}$. If the event membership in set is determined by satisfaction of a certain property or a relationship, then the description of the satisfaction can be given with the curly bracket. For example, if A consists of all real numbers x such that $x^2 > 1$, then it can be expressed as $A = \{x|x^2 > 1\}$, where the bar $|$ is used simply to mean "such that".

Definition (Empty Set):

The set that contains no element is called the empty set and is denoted by $\varnothing$.

Definition (Subset):

A set A is a subset of another set B, written symbolically as $A \subset B$, if every element of A is an element of B. If B contains at least one element that is not in A, then A is said to be a proper subset of B.

Definition:

A set A and a set B are equal if $A \subset B$ and $B \subset A$. Thus, every element of A is an element of B and vice versa.

Definition (Universal Set):

The set that contains all set under consideration in a certain study is called the universal set and is denoted by Ω.

1.2 Set Operations

There are two basic operations for set that produce new sets from existing ones. They are the operations of union and intersection.

Definition (Union):

The union of two sets A and B, denoted by $A \cup B$, is the set of elements that belong to either A or B, that is

$$C = A \cup B = \{x|x \in A \textbf{ or } x \in B\}.$$

The definition can be extended to more than two sets. For example, if $A_1, A_2, ..., A_n$ are n given sets, then their union, denoted by $\bigcup_{i=1}^{n} A_i$, is a set such that x is an element of it if and only if x belongs to at least one of the $A_i, i = 1, 2, ..., n$.

Definition (Intersection):

The intersection of two sets A and B, denoted by $A \cap B$, is the set of elements that belong to both A and B, that is

$$C = A \cap B = \{x|x \in A \textbf{ and } x \in B\}.$$

The definition can be extended to more than two sets. As before, if $A_1, A_2, ..., A_n$ are n given sets, then their intersection, denoted by $\bigcap_{i=1}^{n} A_i$, is the set consisting all elements that belong to all the $A_i, i = 1, 2, ..., n$.

Definition (Disjoint):

Two sets A and B are disjoint if their intersection is the empty set, that is $A \cap B = \varnothing$.

Definition (Complement):

The complement of a set A, denoted by A^c (or $\bar{A}$), is the set consisting of all elements in the universal set that do not belong to A. In other words, $x \in A^c$ if and only if $x \notin A$.

Definition (Relative Complement):

The complement of A with respect to s set B is the set $B - A$ which consists of the element of B that do not belong to A. This complement is called the relative complement of A with respect to B.

From the definition above, the following results can be concluded:

Results 1: The empty set $\varnothing$ is a subset of every set.

Results 2: The empty set $\varnothing$ is unique.

Results 3: The complement of $\varnothing$ is Ω. Vice versa, the complement of Ω is $\varnothing$.

Results 4: The complement of A^c is A.

Results 5: For any set A, $A \cup A^c = \Omega$ and $A \cap A^c = \varnothing$.

Results 6: $A - B = A - (A \cap B)$.

Results 7: $A \cup (B \cup C) = (A \cup B) \cup C)$.

Results 8: $A \cap (B \cap C) = (A \cap B) \cap C$.

Results 9: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Results 10: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Results 11: $(A \cup B)^c = A^c \cap B^c$.[2] More generally, $(\bigcup_{i=1}^{n} A_i)^c = \bigcap_{i=1}^{n} A_i^c$.

[2] Read as "not either A or B"=neither A nor B=both not A and B.

Results 12: $(A \cap B)^c = A^c \cup B^c$.[3] More generally, $(\bigcap_{i=1}^n A_i)^c = \bigcup_{i=1}^n A_i^c$.

Another useful set operation is the Cartesian product defined below.

Definition (Cartesian Product):

Let A and B be two sets. Their Cartesian product, denoted by $A \times B$, is the set of **all ordered pairs** (a, b) such that $a \in A$ and $b \in B$, that is,

$$A \times B = \{(a, b)|a \in A \; and \; b \in B\}.$$

Example:

Let $A = \{1, 2\}$ and $B = \{\alpha, \beta, \gamma\}$. Then

$$A \times B = \{(1, \alpha), (1, \beta), (1, \gamma), (2, \alpha), (2, \beta), (2, \gamma)\}. \blacksquare$$

The preceding definition can be extended to more than two sets. If $A_1, A_2, ..., A_n$ are n given sets, then their Cartesian product is denoted by

$$\times_{i=1}^n A_i = \{(a_1, a_2, ..., a_n)|a_i \in A_i, i = 1, 2, ..., n\}.$$

In particular, if the A_i are equal to $A \; \forall \; i = 1, 2, ..., n$, the one writes A^n for $\times_{i=1}^n A$.

1.3 Class of Subsets

Definition (Power Set):

The set of all the subsets of A is called the power set of A, denoted 2^A. The power set of a set with n elements has 2^n elements, which accounts for its name.

When studying the subsets of a given set, particularly their measure-theoretic properties, the power set is often too big for anything very interesting or useful to said about it. The idea behind the following definitions is to specify subset of 2^A that are large enough to be interesting, but whose characteristics may be more tractable. We typically do this by choosing a *base* collection of sets with known properties, and

[3] Read as "not both A and B"=either not A or not B.

then specifying certain operations for creating new sets from existing ones. These operations permit an interesting diversity of class members to be generated, but important properties of the sets may be deduced from those of the base collection.

Definition (Sigma Fields):

A σ-field (σ-algebra) $\mathcal{F}$ is a class of subsets of A satisfying
(a). A and $\varnothing \in \mathcal{F}$.
(b). If $E \in \mathcal{F}$ then E^c (the complements of E in A) $\in \mathcal{F}$.
(c). If $\{E_n, n \in \mathbb{N}\}$ is a sequence of $\mathcal{F}$-sets, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$.

Definition (Borel Field):
The Borel field of $\mathbb{R}$, denoted by, $\mathcal{B}$, is the smallest σ-field of $\mathbb{R}$ that contains all the half-lines, the set of the form $(-\infty, x], \forall x \in \mathbb{R}$.

Applying the definition of a σ-field and the rule of set algebra, consider what sorts of sets $\mathcal{B}$ contains. By the de Morgan's laws it must contain the half-open intervals $(x_1, x_2]$ for $x_1 < x_2$ (intersections of half-lines). It also contains the open half lines, set of the form $(-\infty, x)$ for $x \in \mathbb{R}$ because

$$(-\infty, x) = \bigcup_{n=1}^{\infty} \left(-\infty, x - \frac{1}{n}\right]$$

where all the sets of the countable union are in $\mathcal{B}$. It therefore contains all open intervals, and also the singleton sets

$$(-\infty,] \cap (-\infty, x)^c = \{x\}$$

for any $x \in \mathbb{R}$. Also, any sets that can be formed from finite or countable infinity union, intersection and complements of theses sets. This is a rich enough collection for our needs to assign probability.

1.4 The Real Numbers

Let us first look at the particularly interested set–the real number set. The first questions may be: How the real numbers are created ? They are from the following procedures:

Natural Number ($\mathbb{N}$) $\overset{by\ addition,\ subtraction,\ multiplication,\ division}{\longrightarrow}$ Rationales,
Rationales + Irrationals $\overset{by\ Dedekind's\ Theorem(Completeness)}{\longrightarrow}$ Real Number ($\mathbb{R}$).

1.4.1 Geometry and the Number System

Geometrical language, with its highly suggestive power, can be very useful and can be given arithmetical meaning. We proceed to define a number of geometrical terms.

Definition (One-Dimensional Euclidean Space):
We speak of the real-number system as one-dimensional space, and of course we visualize it as a line. We denote it as $\mathbb{R}$. In one-dimensional space we use the word *point* to mean "number". A point set is a collection of points.

Definition (Two-Dimensional Euclidean Space):
Just as one-dimensional space is the collection of all real numbers, the two-dimensional space is the Cartesian product of two one-dimensional space, $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$, i.e. the ordered pairs (x, y) of real numbers, which is also known as the 2 dimensional Euclidean space. These ordered pairs are called points in the two-dimensional space.

Definition (n-Dimensional Euclidean Space):
The Cartesian product $\times_{i=1}^{n}\mathbb{R}$ is denoted by $\mathbb{R}^n$, which is known as the n-dimensional Euclidean space.

1.5 Relations and Functions

Let $A \times B$ be the Cartesian product of two sets, A and B.

Definition (Relations):
A relations ρ from A to B is a subset of $A \times B$, that is ρ consists of ordered pairs (a, b) such that $a \in A$ and $b \in B$. In particular, if $A = B$, Then ρ is said to be a relation in A. Whenever ρ is a relation and $(x, y) \in \rho$, then x and y are said to be

ρ-related. This is denoted by writing $x \, \rho \, y$.

Example:
If $A = \{7, 8, 9\}$ and $B = \{7, 8, 9, 10\}$, the $\rho = \{(a, b) | a < b, a \in A, b \in B\}$ is a relation from A to B that consists of the six ordered pairs $(7, 8), (7, 9), (7, 10), (8, 9), (8, 10), (9, 10)$.

Definition: (Functions):
Let ρ be a relation from A to B. Suppose that ρ has the property that for all x in A, if $x \, \rho \, y$ and $x \, \rho \, z$, where y and z are elements in B, then $y = z$. Such a relation is called a function.

Thus a function is a relation ρ such that any two elements in B that are ρ-related to the same x in A must be identical. In other words, to *each element* x in A, there corresponds *only one element* y in B. We call y the value of the function at x and denote it by writing $y = f(x)$. The set A is called the *domain* of the function f, and the set of all values of $f(x)$ for x in A is called the *range* of f, or the *image* of A under f, and is denoted by $f(A)$. In this case, we say that f is a function, or a *mapping*, from A to B. We express this fact by writing

$$f : A \mapsto B.$$

Note that $f(A)$ is a subset of B ($f(A) \subset B$). In particular, if $B = f(A)$, then f is said to be a function from A *onto* B. In this case, every element b in B has a corresponding element a in A such that $b = f(a)$.

Definition (one-to-one function):
A function f defined on a set A is said to be a one-to-one function if whenever $f(x_1) = f(x_2)$ for x_1, x_2 in A, one has $x_1 = x_2$. Equivalently, f is a one-to-one function if whenever $x_1 \neq x_2$, one has $f(x_1) \neq f(x_2)$.[4] Thus a function $f : A \to B$ is one-to-one if to each y in $f(A)$, there corresponds only one element x in A such that $y = f(x)$.

Definition (one-to-one correspondence):
If f is a one-to-one and onto function, then it is said to provide a one-to-one corre-

[4] The function $y = f(x) = 3$ is not a one-to-one function since for $x_1 = 2, x_2 = 4$, $f(2) = f(4) = 3$.

spondence between A and B. In this case, the set A and B are said to be equivalent, denoted by $A \mapsto B$.

Definition: (Inverse Function):
Whenever $A \mapsto B$, there is a function $g: B \to A$ such that if $y = f(x)$, then $x = g(y)$. The function g is called the inverse function of f and is denoted by f^{-1}.

Definition (Composite Function):
Let $f: A \to B$ and $h: B \to C$ be one-to-one and onto functions. Then, the composite function $h \circ f = h[f(x)]$, defines a one-to-one correspondence between A and C.

Example:
The relation $a \, \rho \, b$, where a and b are real numbers such that $a = b^2$, is not a function. This is true because both pairs (a, b) and $(a, -b)$ belong to ρ.[5]

Example:
The relation $a \, \rho \, b$, where a and b are real numbers such that $a^2 = b$, is a function. Since for each a, there is only one b that is ρ-related to a.[6] However, this is not a one-to-one function since there are two elements in A, (i.e. a and $-a$) that are ρ-related to a given b.

1.6 Finite, Countable and Uncountable Set

Let $J_n = \{1, 2, ..., n\}$ be s set consisting of the first n positive integers, and let J^+ denote the set of all positive integers.

Definition (Finite, Countable and Uncountable Set):
A set A is said to be:
(a). Finite if $A \mapsto J_n$ for some positive integer n.

[5] Think of $y = \pm\sqrt{x}$.
[6] Think of $y = x^2$.

(b). Countable if $A \mapsto J^+$. In this case, the set J^+ can be used as an index set for A, that is, the elements of A are assigned distinct indices (subscripts) that belong to J^+. Hence, A can be represented as $A = \{a_1, a_2, ..., a_n, ...\}$.
(c). Uncountable if A is neither finite nor countable. In this case, the elements of A cannot be indexed by J_n for any n, or by J^+.

Example:
Let $A = \{1, 4, 9, ..., n^2, ...\}$. The set is countable, since the function $f: \ J^+ \rightarrow A$ defined by $f(n) = n^2$ is one-to-one and onto. Hence, $A \mapsto J^+$.

Example:
Let $A = J$ be the set of all integers. Then A is countable. To show this, consider the function $f : J^+ \rightarrow A$ defined by

$$f(n) = \begin{cases} (n+1)/2, & n \ \ odd, \\ (2-n)/2, & n \ \ even. \end{cases}$$

It can be verified that f is one-to-one and onto. Hence, $A \mapsto J^+$.

Example:
Let $A = \{x | 0 \leq x \leq 1\}$. The set is uncountable.

This result implies that any subset of $\mathbb{R}$, the set of real numbers, must be uncountable. In particular, $\mathbb{R}$ is uncountable.

Theorem:
The set Q of all rational number is countable.

1.7 Bounded Set

Let us consider the set $\mathbb{R}$ of real numbers.

Definition (Bounded Set):
A set $A \subset \mathbb{R}$ is said to be:
(a). Bounded from above if there exists a number q such that $x \leq q$, $\forall x \in A$. This number is called an upper bound of A.

(b). Bounded from below if there exists a number p such that $x \geq q$, $\forall x \in A$. This number is called an lower bound of A.
(c). Bounded if A has an upper bounded q and a lower bounded p. In this case, there exist a nonnegative number r such that $-r \leq x \leq r$, $\forall x \in A$. This number is equal to $\max(|p|, |q|)$.

Definition (Supremum):
Let $A \subset \mathbb{R}$ be a set bounded from above. If there exists a number l that is an upper bounded of A and is less than or equal to any other upper bounded of A, then l is called the *least upper bound* of A and is denoted by $lub(A)$. Another name for $lub(A)$ is the supremum of A and is denoted by $\sup_A x$.[7]

Definition (Infimum):
Let $A \subset \mathbb{R}$ be a set bounded from below. If there exists a number g that is an lower bounded of A and is greater than or equal to any other lower bounded of A, then g is called the *greatest lower bound* of A and is denoted by $glb(A)$. Another name for $glb(A)$ is the infimum of A and is denoted by $\inf_A x$.

Theorem:
Let $A \subset \mathbb{R}$ be a non-empty set.
(a). If A is bounded from above, then $\sup_A x$ exists.
(b). If A is bounded from below, then $\inf_A x$ exists.

1.8 The Topology of the Real Line

The purpose of this section is to treat rigorously the idea of 'nearness', as it applies to points of the line. The key intergradient of the theory is the distance between a pair of points $x, y \in \mathbb{R}$.

Definition (Euclidean Distance):

[7] The $\sup_A x$ if it exist, is unique, but it may or may not belong to A. For example, let $A = \{x | x < 0\}$, then $\sup_A x = 0$, which does not belong to A.

The Euclidean distance of $\mathbb{R}$ is $|x-y|$, for $x, y \in \mathbb{R}$.

Definition (ϵ-Neighborhood):
An ϵ-neighborhood of a point $x \in \mathbb{R}$ is a set $N_\epsilon(x) = \{y : |x-y| < \epsilon\}$, for some $\epsilon > 0$.

Definition (Deleted ϵ-Neighborhood):
An deleted ϵ-neighborhood of a point $x \in \mathbb{R}$ is a set $N_\epsilon^d(x) = \{y : |x-y| < \epsilon\}$, for some $\epsilon > 0$, but $y \neq x$.

Definition (Open Set):
An open set is a set $A \subseteq \mathbb{R}$ such that for each $x \in A$, there exists for some $\epsilon > 0$ an ϵ-neighborhood which is a subset of A.

Definition (Closed Set):
The complement of an open set in $\mathbb{R}$ is a closed set.

Definition (Limit Point):
A limit point of a set A, denoted by $Lim(A)$ is a point $p \in \mathbb{R}$ such that, for every $\epsilon > 0$, the set $A \cap N_\epsilon^d(p)$ is not empty. The limit points of A are not necessarily elements of A, open set being a case in point.

Limit points can be used to describe closed sets, as can be seen from the following theorem.

Theorem:
A set B is closed if and only if every limit point of B belongs to B.

Example:
$A = \{x|0 < x < 1\}$ is an open subset of $\mathbb{R}$, but is not closed, since both 0 and 1 are limit points of B, but do not belong to it.

Example:
$A = \{x|0 \leq x \leq 1\}$ is closed, but is not open, since any neighborhood of 0 or 1 is not contained in B.

Example:
$A = \{x|0 < x \leq 1\}$ is not open, because any neighborhood of 1 is not contained in B. It is also not closed, because 0 is a limit point that does not belong to B.

Definition (Interior):
A point x_0 in $\mathbb{R}$ is an interior of a set $A \subset \mathbb{R}$ if there exists an $r > 0$ such that $N_r(x_0) \subset A$ and is denoted by $Int(A)$. Thus A is open if it contains entirely of interior points.

Definition (Boundary Point):
A point $p \in \mathbb{R}$ is a boundary point of a set $A \subset \mathbb{R}$ if every neighborhood of p contains points of A as well as points of A^c, the complement of A with respect to $\mathbb{R}$. The set of all boundary points of A is called its boundary and is denoted by $Br(A)$. Thus it is easy to see that $Br(A) = Lim(A) - Int(A)$.

Example:
For an open interval (a, b). Every points of (a, b) is a limit point, and a and b are also limit points not belong to (a, b). They are boundary points of both (a, b) and $[a, b]$.

Definition (Covering):
A collection of set $\{B_\alpha\}$ is said to be a covering of a set A if the union $\cup_\alpha B_\alpha$ contains A. If each B_α is an open set, then $\{B_\alpha\}$ is called an open covering.

Definition (Compact):

A set A is compact if each open covering $\{B_\alpha\}$ of A has a finite sub-covering, that is, there is a finite sub-collection $B_{\alpha_1}, B_{\alpha_2}, ..., B_{\alpha_n}$ of $\{B_\alpha\}$ such that $A \subset \cup_{i=1}^{n} B_{\alpha_i}$.

The concept of compactness is motivated by the classical Heine-Borel theorem, which characterizes compact sets in $\mathbb{R}$ as closed and bounded sets.

Theorem (Heine-Borel):
A set $B \subset \mathbb{R}$ is compact if and only if it is closed and bounded.

Thus, according to the Heine-Borel theorem, every closed and bounded interval $[a, b]$ is compact.

Example:
Let us examine the open interval $(0, 1)$. Consider the collection $B_\alpha = \{\left(\frac{1}{i}, 1\right), i = 1, 2, ...\}$ and observe that

$$(0, 1) = \left(\frac{1}{2}, 1\right) \cup \left(\frac{1}{3}, 1\right) \cup \cdots,$$

that is, B_α us an open cover of $(0, 1)$. Does B_α have a finite subset that covers $(0, 1)$? No, because the greatest lower bound of any finite subset of B_α is bounded away from 0, so no such subset can possibly cover $(0, 1)$ entirely. There, we conclude that $(0, 1)$ is not a compact subset of $\mathbb{R}$.

2 Measure

A measure is a set function, a mapping which associates a real number with a set. Commonplace examples of measure include the lengths, areas, and volumes of geometrical figures, but wholly abstract sets can be 'measured' in an analogous way. Formally, we have the following definition.

Definition:
Given a class $\mathcal{F}$ of subsets of a set Ω, a *measure*

$$\mu : \mathcal{F} \mapsto \mathbb{R}$$

is a function having the following properties:
(a). $\mu(A) \geq 0, \ \forall A \in \mathcal{F}$.
(b). $\mu(\varnothing) = 0$.
(c). For a countable collection $\{A_j \in \mathcal{F}, j \in \mathbb{N}\}$ with $A_i \cap A_j = \varnothing$ for $i \neq j$ and $\bigcup_j A_j \in \mathcal{F}$,

$$\mu\left(\bigcup_j A_j\right) = \sum_j \mu(A_j). \quad (\textit{Countable Additivity})$$

The particular cases at issue in this course are of course the probabilities of random events is a sample space Ω.

Definition (Measurable Space):
A measurable space is a pair $(\Omega, \mathcal{F})$ where Ω is any collection of objects, and $\mathcal{F}$ is a σ-field of subset of Ω.

Definition (Measure Space):
When $(\Omega, \mathcal{F})$ is a measurable space, the triple $(\Omega, \mathcal{F}, \mu)$ is called a measure space. More than one measure can be associated with the measurable space $(\Omega, \mathcal{F})$, hence the distinction between measure space and measurable space is important.

Example:

The case closest to everyday intuition if Lebesgue measure, m, on the measurable space $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel field on $\mathbb{R}$. Generalizing the notion of length in geometry, Lebesgue measure assigns $m((a,b]) = b-a$ to an interval $(a,b]$. Additivity is an intuitively plausible property if we think of measuring the total length of a collection of disjoint intervals.

Some additional properties may be deduced from the definition.

Theorem:

For arbitrary $\mathcal{F}$-set A, B, and $\{A_j, j \in \mathbb{N}\}$,
(a). If $A \subseteq B$ then $\mu(A) \leq \mu(B)$ (monotonicity).
(b). $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$.
(c). $\mu(\cup_j A_j) \leq \sum_j \mu(A_j)$. (countable subadditivity).

3 Metric Space

Central to the properties of $\mathbb{R}$ was the concept of distance. For any real number x and y, the Euclidean distance between them is the number $d_E(x,y) = |x-y| \in \diagdown^+$. Generalizing this idea, a set (otherwise arbitrary) having a distance measure, or *metric*, defined for each pair of elements is called a *metric space*. Let $\mathbb{S}$ denote such a set.

Degfinition (Metric and Metric Space):
A metric is a mapping $d : \mathbb{S} \times \mathbb{S} \mapsto \mathbb{R}^+$ having the properties
(a). $d(x,y) = d(y,x)$,
(b). $d(x,y) = 0$ iff $x = y$,
(c). $d(x,y) + d(y,z) \geq d(x,z)$ (triangle inequality),
A metric space $(\mathbb{S}, d)$ is a set $\mathbb{S}$ paired with metric d, such that conditions (a)-(c) hold for each pair of elements of $\mathbb{S}$.

While the Euclidean metric on $\mathbb{R}$ is the familiar case, and the proof that d_E satisfies (a)-(c) is elementary, d_E is not the only possible metric on $\mathbb{R}$.

In the space $\mathbb{R}^2$ a larger variety of metric is found.

Example:
The Euclidean distance on $\mathbb{R}^2$ is

$$d_E(x,y) = \|x-y\| = [(x_1-y_1)^2 + (x_2-y_2)^2]^{1/2},$$

and $(\mathbb{R}^2, d_E)$ is the Euclidean plane. An alternative is the 'taxicab' (honest) metric,

$$d_T = |x_1 - y_1| + |x_2 - y_2|.$$

d_E is the shortest distance between two address in a city, but d_T is the shortest distance by taxi.

In metric space theory, the properties of $\mathbb{R}$ are revealed as a special case.

Definition (ϵ-Neighborhood):
The concept of an open neighborhood in a metric space $(\mathbb{S}, d)$ is the set $N_\epsilon(x,d) =$

$\{y : y \in \mathbb{S}, d(x, y) < \epsilon\}$, where $x \in \mathbb{S}$ and $\epsilon > 0$.

Definition (Open Set):
An open set of $(\mathbb{S}, d)$ is a set $A \subseteq \mathbb{S}$ such that for each $x \in A$, $\exists \delta > 0$ such that $N_\delta(x, d)$ is a subset of A.

4 Limit and Continuity of Real Functions

The notation of limits and continuity of functions lie at the kernel of calculus. In this section, we review the concepts of limits and continuity of real-valued functions, and study some of their properties. The domain of the functions will be subsets of $\mathbb{R}$. A

4.1 Limits of a Functions

4.1.1 What is mean $x \to a$

Before defining the notation of a limit of a function, let us understand what is meant by the natation $x \to a$, where a and x are elements in $\mathbb{R}$.

Definition ($x \to a$ when a is finite):
If a is finite, then $x \to a$ means that x can have value that belong to a neighborhood $N_r(a)$ of a for any $r > 0$, but $x \neq a$, that is, $0 < |x - a| < r$. Such a neighborhood is called a deleted neighborhood of a, that is, a neighborhood from which the point a has been removed.

Definition ($x \to a$ when a is infinite):
If a is infinite ($+\infty$ or $-\infty$), then $x \to a$ indicates that $|x|$ can get larger and larger without any constraint on the extent of its increase.

Let us now study the behavior of a function $f(x)$ as $x \to a$.

Definition ($f(x)$ is finite and a is finite):
Suppose that the function $f(x)$ is defined in a deleted neighborhood of a point $a \in \mathbb{R}$. Then $f(x)$ is said to have a limit L as $x \to a$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - L| < \epsilon$$

for all x for which

$$0 < |x - a| < \delta.$$

In this case, we write $f(x) \to L$ as $x \to a$, which is equivalent to saying that $\lim_{x\to a} f(x) = L$. Less formally, we say that $f(x) \to L$ as $x \to a$ if, however small the positive number ϵ might be, $f(x)$ differs from L by less than ϵ for value of x sufficiently close to a.[8]

Example:

Use the $\epsilon - \delta$ definition of limit to prove that

$$\lim_{x\to 2}(3x-2) = 4.$$

Solution:

You must show that for each $\epsilon > 0$, there exists a $\delta > 0$ such that $|(3x-2)-4| < \epsilon$ whenever $0 < |x-2| < \delta$. Because your choice of δ depends on ϵ, you need to establish a connection between the absolute value $|(3x-2)-4|$ and $|x-2|$.

$$|(3x-2)-4| = |3x-6| = 3|x-2|.$$

So, for a given $\epsilon > 0$ you can choose $\delta = \epsilon/3$. This choice works because

$$0 < |x-2| < \delta = \frac{\epsilon}{3}$$

implies that

$$|(3x-2)-4| = 3|x-2| < 3\left(\frac{\epsilon}{3}\right) = \epsilon.$$

[8]If $f(x)$ has a limit L as $x \to a$, then L must be unique. To show this, suppose that L_1 and L_2 are two limits of $f(x)$ as $x \to a$. Then, for any $\epsilon > 0$ there exist $\delta_1 > 0, \delta_2 > 0$ such that

$$|f(x) - L_1| < \frac{\epsilon}{2},\ \ if\ 0 < |x-a| < \delta_1,$$
$$|f(x) - L_2| < \frac{\epsilon}{2},\ \ if\ 0 < |x-a| < \delta_2.$$

Hence, if $\delta = \min(\delta_1, \delta_2)$, then

$$\begin{aligned} |L_1 - L_2| &= |L_1 - f(x) + f(x) - L_2| \\ &\le |f(x) - L_1| + |f(x) - L_2| \\ &< \epsilon \end{aligned}$$

for all x for which $0 < |x-a| < \delta$. Since $|L_1 - L_2|$ is smaller than ϵ, which is an arbitrary positive number, we must have $L_1 = L_2$.

Note:

To see the last results, suppose that $L_1 > L_2$, then $L_1 - L_2 > 0$. Let $L_1 - L_2 = 2\epsilon > 0$. However this is not the case that $|L_1 - L_2| < \epsilon$. This is a contradiction.

Definition ($f(x)$ is infinite and a is finite):
For every positive number M there exists a $\delta > 0$ such that

$$|f(x)| > M$$

for all x for which

$$0 < |x - a| < \delta.$$

Example:
Discuss the limit of $\lim_{x\to 0} \frac{1}{x^2}$.

Solition:
Let $f(x) = 1/x^2$, we can see that as x approach 0 from either the right or the left, $f(x)$ increase without bound. For instance,

$$0 < |x - 0| < \frac{1}{10} \Longrightarrow f(x) = \frac{1}{x^2} > 100.$$

Similarly, you can force $f(x)$ to be larger than 1000000 as follows.

$$0 < |x - 0| < \frac{1}{1000} \Longrightarrow f(x) = \frac{1}{x^2} > 100000.$$

Definition ($f(x)$ is finite and a is infinite):[9]
If a is infinite and L is finite, then $f(x) \to L$ as $x \to a$ if for any $\epsilon > 0$ there exists a positive number N such that

$$|f(x) - L| < \epsilon$$

for all x for which

$$|x| > N.$$

[9]This is the case for limit theorem in statistics

Definition ($f(x)$ is infinite and a is infinite):
If both a and L are infinite, then $f(x) \to L$ as $x \to a$ if for any $B > 0$ there exists a positive number A such that

$$|f(x)| > B$$

if

$$|x| > A.$$

4.1.2 One-Sided Limit

The limit of $f(x)$ as described above is actually called a two-sided limit. This is because x can approach a from either side. There are, however, cases where $f(x)$ can have a limit only when x approach a from one side. Such a limit is called a one-sided limit.

Definition (Left-sided limit):
If $f(x)$ has a limit as x approach a from the left, symbolically written as $x \to a^-$, then $f(x)$ has a left-sided limit, which we denote by L^-. In this case we write

$$\lim_{x \to a^-} f(x) = L^-.$$

It follows that $f(x)$ has a left-sided limit L^- as $x \to a^-$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - L^-| < \epsilon$$

for all x for which $-\delta < x - a < 0$.

Definition (Right-sided limit):
If $f(x)$ has a limit as x approach a from the right, symbolically written as $x \to a^+$, then $f(x)$ has a right-sided limit, which we denote by L^+. In this case we write

$$\lim_{x \to a^+} f(x) = L^+.$$

It follows that $f(x)$ has a right-sided limit L^+ as $x \to a^+$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - L^+| < \epsilon$$

for all x for which $0 < x - a < \delta$.

Theorem (The Existence of a Limit):
A necessary and sufficient condition that $\lim_{x\to a} f(x)$ exist is that both $\lim_{x\to a^-} f(x)$ and $\lim_{x\to a^+} f(x)$ exist and be equal.[10]

Example:
Find the limit of the smallest integer function $f(x) = [x]$ as x approach 0 from the left and from the right.

Solution:
The limit as x approach 0 from the left is given by

$$\lim_{x\to 0^-} [x] = -1$$

[10]**Proof**:
We first prove the sufficiency of the condition. We assume that $\lim_{x\to a^+} f(x) = \lim_{x\to a^-} f(x) = L$, and show $\lim_{x\to a} f(x) = L$.

For a given ϵ, there exist a δ_1 for which

$$|f(x) - L| < \epsilon \quad if\ a < x < a + \delta_1.$$

Similarly, there exist a δ_2 for which

$$|f(x) - L| < \epsilon \quad if\ a - \delta_2 < x < a.$$

If we let $\delta = \min(\delta_1, \delta_2)$, then

$$|f(x) - L| < \epsilon \quad if\ a - \delta < x < a + \delta,$$

which means that $0 < |x - a| < \delta$.

To prove the necessity of the condition we show that if $\lim_{x\to a} f(x) = L$, then $\lim_{x\to a^+} f(x) = L$. A similar proof will verify the condition for left-hand limits.

By definition, given $\epsilon > 0$ we must find a $\delta > 0$ for which

$$|f(x) - L| < \epsilon \quad if\ a < x < a + \delta.$$

For the given ϵ, there is a δ_1 for which

$$|f(x) - L| < \epsilon \quad if\ a - \delta_1 < x < a + \delta_1.$$

Hence, merely let $\delta = \delta_1$.

and the limit as x approach 0 from the right is given by

$$\lim_{x\to 0^+}[x] = 0.$$

The smallest integer function $\lim_{x\to 0}[x]$ does not exist. By similar reasoning, you can see that the smallest integer function do not have a limit at any integer n.

4.2 Some Properties Associated with Limits of Functions

The following theorems give some fundamental properties associated with function limits.

Theorem:
Let $f(x)$ and $g(x)$ be real-valued functions defined on $D \subset \mathbb{R}$. Suppose that $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. Then
(a). $\lim_{x\to a}[f(x) + g(x)] = L + M$,
(b). $\lim_{x\to a}[f(x)g(x)] = LM$,
(c). $\lim_{x\to a}[1/g(x)] = 1/M$ if $M \neq 0$,
(d). $\lim_{x\to a}[f(x)/g(x)] = L/M$ if $M \neq 0$.[11]

Theorem:
If $f(x) \leq g(x)$, $\forall x \in D \subset \mathbb{R}$, then $\lim_{x\to a} f(x) \leq \lim_{x\to a} g(x)$.

4.3 The o and O Notation

These symbols provide a convenient way to describe the limiting behavior of a function $f(x)$ as x tends to a certain limit.

[11]**Proof** (for part (c)):
Let $\epsilon > 0$ be given. If $M \neq 0$, then there exists a $\lambda_1 > 0$ such that $|g(x)| > |M|/2$ if $0 < |x - a| < \lambda_1$.[12] Also, there exists a λ_2 such that $g(x) - M| < \epsilon M^2/2$ if $0 < |x - a| < \lambda_2$. Then,

$$\begin{aligned} \left|\frac{1}{g(x)} - \frac{1}{M}\right| &= \frac{|g(x) - M|}{|g(x)||M|} \\ &< \frac{2|g(x) - M|}{|M|^2} \\ &< \epsilon, \end{aligned}$$

if $0 < |x - a| < \lambda$, where $\lambda = \min(\lambda_1, \lambda_2)$.

Definition (Big O):
Let $f(x)$ and $g(x)$ be two functions defined on $D \subset \mathbb{R}$. The function $g(x)$ is positive and usually has a simple form such as 1, x or $1/x$. Suppose there exist a positive number K such that

$$\frac{|f(x)|}{g(x)} \leq K$$

for all $x \in E$, where $E \subset D$. Then, $f(x)$ is said to be of an order of magnitude not exceeding that of $g(x)$. This fact is denoted by writing

$$f(x) = O(g(x))$$

for all $x \in E$. In particular, if $g(x) = 1$, then $f(x)$ is necessarily a bounded function on E.

Example:

$$\begin{aligned} \cos(x) &= O(1) \quad for\ all\ x, \\ x &= O(x^2) \quad for\ large\ values\ of\ x, \\ x^2 + x &= O(x^2) \quad for\ all\ x. \end{aligned}$$

Definition (Small o, as $x \to a$ (finite)):
Suppose that the relationship between $f(x)$ and $g(x)$ is such that

$$\lim_{x \to a} \frac{f(x)}{g(x)} = 0.$$

Then we say that $f(x)$ is of smaller order of magnitude than $g(x)$ in a deleted neighborhood of a. This fact is denoted by writing

$$f(x) = o(g(x)) \quad as\ x \to a,$$

which is equivalent to saying that $f(x)$ tends to zero more rapidly than $g(x)$ as $x \to a$.

Definition (Small o, as $x \to \infty$):
The o symbol can also be used when x tends to infinity. In this case we write

$$f(x) = o(g(x)) \quad for\ x > A,$$

where A is some positive number.

Example:

$$\begin{aligned} x^2 &= o(x) \quad as\ x \to 0, \\ \sqrt{x} &= o(x) \quad as\ x \to \infty. \end{aligned}$$

Definition (Asymptotically Equal):
If $f(x)$ and $g(x)$ be any two functions such that[13]

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

then $f(x)$ and $g(x)$ are said to be asymptotically equal, written symbolically $f(x) \sim g(x)$, as $x \to a$.

Example:

$$\begin{aligned} x^2 &\sim x^2 + 3x + 1 \ \ as\ x \to \infty, \\ \sin x &\sim x \quad as\ x \to 0. \end{aligned}$$

One the basis of the above definitions, the following properties can be deduced:
(a). $O(f(x) + g(x)) = O(f(x)) + O(g(x))$.[14]

[13] Or write as $\frac{f(x)}{g(x)} \to 1$ as $x \to a$.

[14] **Proof**:
Let $h_1(x) = O(f(x))$, then $|h_1(x)| \leq Kf(x)$. Similarly $|h_2(x)| \leq Kg(x)$. Therefore

$$|h_1(x) + h_2(x)| \leq |h_1(x)| + |h_2(x)| \leq K(f(x) + g(x)),$$

i.e. $h_1(x) + h_2(x)$ is $O((f(x) + g(x))$.

(b). $O(f(x)g(x)) = O(f(x))O(g(x))$.[15]
(c). $o(f(x)g(x)) = O(f(x))o(g(x))$.
(d). If $f(x) \sim g(x)$ as $x \to a$, then $f(x) = g(x) + o(g(x))$ as $x \to a$.

4.4 Continuous Functions

A function $f(x)$ may has a limit L as $x \to a$. This limit, **may** or **may not be** equal to the value of the function at $x = a$. In fact, the function may not even be defined at this point. If $f(x)$ is defined at $x = a$ and $L = f(a)$, then $f(x)$ is said to be continuous at $x = a$.

Definition (Continuity):
Let $f : D \to \mathbb{R}$, where $D \subset \mathbb{R}$, and let $a \in D$.[16] Then $f(x)$ is continuous at $x = a$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon$$

for all $x \in D$ for which $|x - a| < \delta$.

Note:
To show the continuity of $f(x)$ at $x = a$, the following conditions must be verified:
(a). $f(x)$ is defined at all points inside a neighborhood of the point a.
(b). $f(x)$ has a limit from the left and a limit from the right as $x \to a$, and that these two limits are equal to L.
(c). The value of $f(x)$ at $x = a$ is equal to L.

Definition (Discontinuity of the First Kind):
A function $f : D \to \mathbb{R}$ has a discontinuity of the first kind at $x = a$ if $f(a^+)$ and $f(a^-)$ exist, but at least one of them is different from $f(a)$.

[15]**Proof**:
Let $h_1(x) = O(f(x))$, then $|h_1(x)| \leq Kf(x)$. Similarly $|h_2(x)| \leq Kg(x)$. Therefore

$$|h_1(x) \cdot h_2(x)| \leq |h_1(x)| \cdot |h_2(x)| \leq K^2(f(x)g(x)),$$

i.e. $h_1(x)h_2(x)$ is $O((f(x)g(x))$.
[16]So, the point of interest in defined in the domain of f.

Definition (Discontinuity of the Second Kind):
A function $f : D \to \mathbb{R}$ has a discontinuity of the second kind at $x = a$ if at least one of $f(a^+)$ and $f(a^-)$ does not exist.

Definition: (Continuity on a Point Set):
A function $f : D \to \mathbb{R}$ is continuous on $E \subset D$ if it is continuous at every point of E.

Definition (One-sided continuity):
A function $f : D \to \mathbb{R}$ is left-continuous at $x = a$ if $\lim_{x \to a^-} f(x) = f(a)$. It is right-continuous at $x = a$ if $\lim_{x \to a^+} f(x) = f(a)$.[17]

Definition (Continuity on a Closed Interval):
A function f is continuous on the closed interval $[a, b]$ if it is continuous on the open interval (a, b) and

$$\lim_{x \to a^+} f(x) = f(a) \quad \textit{and} \quad \lim_{x \to b^-} f(x) = f(b).$$

The function is continuous from the right at a and continuous from the left at b.

Definition (Uniformly Continuous):
The function $f : D \to \mathbb{R}$ is uniformly continuous on $E \subset D$ if for every $\epsilon > 0$ there exists a $\delta > 0$ such that

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

for all $x_1, x_2 \in E$ for which $|x_1 - x_2| < \delta$.

[17] For example, the function

$$f(x) = \begin{cases} x - 1, & x \leq 0, \\ 1, & x > 0 \end{cases}$$

is left-continuous at $x = 0$, since $f(0^-) = -1 = f(0)$. If $f(x)$ were defined so that $f(x) = x - 1$ for $x < 0$ and $f(x) = 1$ for $x \geq 0$, then it would be right-continuous at $x = 0$.

Example:

Show that $f(x) = x^2$ is continuous at x_0. We Choose any h we please, say $h = 1$. We consider f, then, in the interval $I : \{x_0 - 1 < x < x_0 + 1\}$. Here, $|x - x_0| < 1$. From this we see that

$$|x| < |x_0| + 1.$$

Now for x in I we consider

$$\begin{aligned} |f(x) - f(x_0)| &= |x^2 - x_0^2| = |x - x_0||x + x_0| \\ &\leq |x - x_0|(|x| + |x_0|) \leq |x - x_0|(2|x_0| + 1). \end{aligned}$$

Consequently, we see that choosing $|x - x_0| < \epsilon/(2|x_0| + 1)$ gives us

$$|x^2 - x_0^2| < \epsilon.$$

To achieve this equality, we have imposed two conditions on x:

$$\begin{aligned} |x - x_0| < 1 \quad that\ is,\ x \in I, \\ |x - x_0| < \epsilon/(2|x_0| + 1). \end{aligned}$$

Thus, if we choose $\delta = \min[1, \epsilon/(2|x_0|+1)]$, our definition is satisfied. Note that the dependence of δ on x_0 is quite explicit.

Example:

Show that the function defined by $f(x) = 1/x$ is continuous(uniformly) in the set $D : \{|x| \geq 1/2\}$. We have

$$|f(x) - f(x_0)| = \left|\frac{1}{x} - \frac{1}{x_0}\right| = \frac{|x - x_0|}{|xx_0|}.$$

Hence, if we take x and x_0 in D, and choose $\delta = |x_0|\epsilon/2$,[18] we get

$$|f(x) - f(x_0)| \leq \frac{2|x - x_0|}{|x_0|} < \frac{2\epsilon|x_0|}{2|x_0|} = \epsilon \quad if \quad |x - x_0| < \delta.$$

Now, as in any limit, if δ satisfies the conditions of the definition, so will any smaller number, say $\delta' \leq \delta$. In this example we have $\delta = |x_0|\epsilon/2$, where $x_0 \geq 1/2$. Hence $\delta' = \epsilon/4$ is no larger than δ, and therefore

$$|f(x) - f(x_0)| < \epsilon \quad if \quad |x - x_0| < \delta' = \epsilon/4$$

[18] Here, it look like that δ still depends on x_0.

for all x, x_0 in D.

The significance of our new choice, δ', is that it will work for any arbitrary choice of x_0 in D; that is, δ' is independent of x_0 in D. It depends only upon the set D itself, but not on the individual point x_0 in D. Such a δ' is called uniform, and we say that $f(x)$ is uniformly continuous in D.

4.4.1 Some Properties of Continuous Functions

Theorem:

Let $f(x)$ and $g(x)$ be two continuous functions defined on a set $D \subset \mathbb{R}$. Then:
(a). $f(x) + g(x)$ and $f(x)g(x)$ are continuous on D.
(b). $\alpha f(x)$ is continuous on D, where α is a constant.
(c). $f(x)/g(x)$ is continuous on D provided that $g(x) \neq 0$ on D.

Theorem:

Suppose that $f : D \to \mathbb{R}$ is continuous on D, and $g : f(D) \to \mathbb{R}$ is continuous on $f(D)$, the image of D under f. Then the composite function $h : D \to \mathbb{R}$ defined as $h(x) = g[f(x)]$ is continuous on D.

Proof:

Let $\epsilon >$ be given, and let $a \in D$. Since g is continuous at $f(a)$, there exist a δ' such that $|g[f(x)] - g[f(a)]| < \epsilon$ if $|f(x) - f(a)| < \delta'$. Since $f(x)$ is continuous at $x = a$, there exists a $\delta > 0$ such that $|f(x) - f(a)| < \delta'$ if $|x - a| < \delta$. It follows that by takeing $|x - a| < \delta$ we must have $|h(x) - h(a)| < \epsilon$.

Theorem:

If $f(x)$ is continuous at $x = a$ and $f(a) > 0$, then these exists a neighborhood $N_\delta(a)$ in which $f(x) > 0$.

Proof:

Since $f(x)$ is continuous at $x = a$, there exists a $\delta > 0$ such that

$$|f(x) - f(a)| < \frac{1}{2}f(a),$$

if $|x - a| < \delta$. This implies that

$$f(x) > \frac{1}{2}f(a) > 0$$

$\forall x \in N_\delta(a)$.

Theorem (The Intermediate-Value Theorem):
Let $f : D \to \mathbb{R}$, and let $[a, b]$ be a closed interval contained in D. Suppose that $f(a) > f(b)$. If λ is a number such that $f(a) > \lambda > f(b)$, then there exists a point c, where $a < c < b$, such that $\lambda = f(c)$.[19]

In the following we want to establish some properties of continuous functions, defined on closed bounded intervals, that is intervals that not only are bounded above and below, but also include both end points.

Theorem:
Suppose that $f : D \to \mathbb{R}$ is continuous and that D is bounded and closed. Then $f(x)$ is bounded in D.

Theorem:
If $f : D \to \mathbb{R}$ is continuous, where D is closed and bounded, then $f(x)$ achieves its infimum and supremum at least once in D, that is, there exists $\xi, \eta \in D$ such that

$$f(\xi) \leq f(x) \ \ \forall x \in D,$$
$$f(\eta) \geq f(x) \ \ \forall x \in D.$$

Equivalently,

$$f(\xi) = \inf_{x \in D} f(x),$$
$$f(\eta) = \sup_{x \in D} f(x).$$

[19] The direct implication of the intermediate-value theorem is that a continuous function possesses the properties of assuming at least once every value between any two distinct value taken inside its domain.

Theorem:
Let $f : D \mapsto \mathbb{R}$ be continuous on D. If D is closed and bounded, then f is uniformly continuous on D.

4.4.2 Lipschitz Continuous Functions

Lipschitz continuity is a specialized form of uniform continuity.

Definition:
The function $f : D \mapsto \mathbb{R}$ is said to satisfy the Lipschitz condition on a set $E \subset D$ if there exist constants, K and α, where $K > 0$ and $0 < \alpha \leq 1$ such that

$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|^\alpha, \ \forall x_1, x_2 \in E.$$

Notationally, whenever $f(x)$ satisfies the Lipschitz condition with constant K and α on a set E, we say that it is $Lip(K, \alpha)$.[20]

4.5 Convex Functions

Convex functions are frequently used in operations research. They also happen to be continuous. The natural domains for such functions are convex sets.

[20] As an example of Lipschitz continuous function, consider $f(x) = \sqrt{x}, x \geq 0$. We claim that $\sqrt{x}$ is $Lip(1, 1/2)$ on its domain. To see this, we first write

$$|\sqrt{x_1} - \sqrt{x_2}| \leq \sqrt{x_1} + \sqrt{x_2}.$$

Hence,

$$|\sqrt{x_1} - \sqrt{x_2}|^2 \leq |x_1 - x_2|.$$

Thus,

$$|\sqrt{x_1} - \sqrt{x_2}| \leq |x_1 - x_2|^{1/2}.$$

Definition (Convex Set):
A set $D \subset \mathbb{R}^1$ is convex if $\lambda x_1 + (1-\lambda)x_2 \in D$ whenever x_1, x_2 belong to D and $0 \leq \lambda \leq 1$. Geometrically, a convex set contains the line segment connecting any two of its points. The same definition actually applies to convex sets in $\mathbb{R}^n$.

Definition:
A function $f: D \to \mathbb{R}$ is convex if

$$f[\lambda x_1 + (1-\lambda)x_2] \leq \lambda f(x_1) + (1-\lambda)f(x_2), \ \forall x_1, x_2 \in D,$$

and any λ such that $0 \leq \lambda \leq 1$.

Geometrically, a convex function in $\mathbb{R}$ means that if P and Q are any two points on the graph of $y = f(x)$, then the portion of the graph between P and Q lies between the chord PQ. Example of convex functions include $f(x) = x^2$ on $\mathbb{R}$, $f(x) = e^x$ on $\mathbb{R}$, to name just a few.

Definition:
A function $f: D \to \mathbb{R}$ is concave if $-f$ is convex.

Lemma:
If $f : [a,b] \to \mathbb{R}$ is convex and the value of f at a and b are finite, then $f(x)$ is bounded from above on $[a,b]$ by $M = \max\{f(a), f(b)\}$, and $f(x)$ is also bounded from below.

Proof:
(a). Because $x \in [a,b]$, then $x = \lambda a + (1-\lambda)b$ for some $\lambda \in [0,1]$, since $[a,b]$ is a convex set. Hence,

$$\begin{aligned} f(x) &\leq \lambda f(a) + (1-\lambda)f(b) \\ &\leq \lambda M + (1-\lambda)M = M. \end{aligned}$$

(b). We First note that any $x \in [a,b]$ can be written

$$x = \frac{a+b}{2} + t,$$

where

$$a - \frac{a+b}{2} \leq t \leq b - \frac{a+b}{2}.$$

Now, if $(a+b)/2 + t = x$ belong to $[a, b]$, so does $(a+b)/2 - t$, then since

$$\frac{1}{2}\left[\frac{a+b}{2} + t\right] + \frac{1}{2}\left[\frac{a+b}{2} - t\right] = \frac{a+b}{2},$$

so

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{2} f\left(\frac{a+b}{2} + t\right) + \frac{1}{2} f\left(\frac{a+b}{2} - t\right),$$

or

$$f\left(\frac{a+b}{2} + t\right) \geq 2f\left(\frac{a+b}{2}\right) - f\left(\frac{a+b}{2} - t\right).$$

Since

$$f\left(\frac{a+b}{2} - t\right) \leq M,$$

then

$$f\left(\frac{a+b}{2} + t\right) \geq 2f\left(\frac{a+b}{2}\right) - M,$$

That is, $f(x) \geq m \ \forall x \in [a, b]$, where $m = f\left(\frac{a+b}{2}\right) - M$.

Theorem:

Let $f : D \to \mathbb{R}$ be a convex function, where D is an open interval. Then f is Lip$(K, 1)$ on any closed interval $[a, b]$ contained in D, that is,

$$|f(x_1) - f(x_2| \leq K|x_1 - x_2|, \ \ \forall x_1, x_2 \in [a, b].$$

Proof:

Consider the closed interval $[a - \epsilon, b + \epsilon]$, where $\epsilon > 0$ is chosen that this interval is contained in D. Let m' and M' be, respectively, the lower and upper bounds of f on $[a - \epsilon, b + \epsilon]$. Let x_1, x_2 be any two distinct points in $[a, b]$. Define z_1 and λ as

$$\begin{aligned} z_1 &= x_2 + \frac{\epsilon(x_2 - x_1)}{|x_1 - x_2|}, \\ \lambda &= \frac{|x_1 - x_2|}{\epsilon + |x_1 - x_2|}. \end{aligned}$$

Then $z_1 \in [a-\epsilon, b+\epsilon]$. This is true because $(x_2-x_1)/|x_1-x_2|$ is either equal to 1 or to -1. Since $x_2 \in [a,b]$, then

$$a-\epsilon \leq x_2-\epsilon \leq x_2+\frac{\epsilon(x_2-x_1)}{|x_1-x_2|} \leq x_2+\epsilon \leq b+\epsilon.$$

Furthermore, it can be verified that

$$x_2 = \lambda z_1 + (1-\lambda)x_1.$$

We then have

$$f(x_2) \leq \lambda f(z_1) + (1-\lambda)f(x_1) = \lambda[f(z_1)-f(x_1)] + f(x_1).$$

Thus,

$$\begin{aligned} f(x_2)-f(x_1) &\leq \lambda[f(z_1)-f(x_1)] \\ &\leq \lambda[M'-m'] \\ &\leq \frac{|x_1-x_2|}{\epsilon}(M'-m') = K|x_1-x_2|. \end{aligned}$$

Corollary:
Let $f:\ D \to \mathbb{R}$ be a convex function, where D is an open interval. If $[a,b]$ is any closed interval contained in D, then $f(x)$ is uniformly continuous on $[a,b]$ and is therefore continuous on D.

5 Differentiation

Differentiation originated in connection with the problem of drawing tangents to curves and of finding maxima and minima of functions.

5.1 The Derivative of a Function

The notation of differentiation was motivated by the need to find the tangent to a curve at a given point. Fetmat's approach to this problem was inspired by a geometric reasoning. His method uses the idea of a tangent as the limiting position of a secant when two of its points of intersection with the curve tend to coincide.

Definition (Derivative) :
Let $f(x)$ be a function defined in a neighborhood $N_r(x_0)$ of a point x_0. Consider the ration

$$\phi(h) = \frac{f(x_0 + h) - f(x_0)}{h},$$

where h is a nonzero increment of x_0 such that $-r < h < r$. If $\phi(h)$ has a limit as $h \to 0$, then the limit is called the derivative of $f(x)$ at x_0 and is denoted by $f'(x_0)$. It is also common to sue the notation

$$\left.\frac{df(x)}{dx}\right|_{x=x_0} = f'(x_0).$$

We thus have

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}. \tag{1}$$

By putting $x = x_0 + h$, (1) can be written as

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

If $f'(x_0)$ exists, then $f(x)$ is said to be differentiable at $x = x_0$. Geometrically, $f'(x_0)$ is the slope of the tangent to the graph of the function $y = f(x)$ at the point

(x_0, y_0), where $y_0 = f(x_0)$.[21]

Definition (Differentiable on a set):
If $f(x)$ has a derivative at every point of a set D, then $f(x)$ is said to be differentiable on D.

Definition (Second and Higher Derivative):
If $f(x)$ is differentiable on a set D, then $f'(x)$ is a function itself defined on D. In the event $f'(x)$ itself is differentiable on D, then its derivative is called the second derivative of $f(x)$ and is denoted by $f''(x)$. It is also common to use the notation

$$\frac{d\left(\frac{df(x)}{dx}\right)}{dx} = \frac{d^2 f(x)}{dx^2} = f''(x).$$

By the same token, we can define the nth ($n \geq 2$) derivative of $f(x)$ as the derivative of the $(n-1)$st derivative of $f(x)$. We denote this derivative by

$$\frac{d^n f(x)}{dx^n} = f^{(n)}(x), \quad n = 2, 3, ...$$

Theorem (Differentiability implies Continuity):
Let $f(x)$ be defined at in a neighborhood of a point x_0. If $f(x)$ has derivative at x_0, then it must be continuous at x_0.

Proof:
From definition above we can write

$$f(x_0 + h) - f(x_0) = h\phi(h).$$

If the derivative of $f(x)$ exists at x_0, then $\phi(h) \to f'(x_0)$ as $h \to 0$. It follows that

$$f(x_0 + h) - f(x_0) \to 0 \cdot f'(x_0) = 0$$

[21] It is important to note that in order for $f'(x_0)$ to exist, the left-sided and right-sided limits of $\phi(h)$ must exist and be equal as $h \to 0$, or as x approaches x_0 from either side. It is possible to consider only one-sided derivatives at $x = x_0$. These occur when $\phi(h)$ has just a one-sided limit as $h \to 0$. We shall not, however, concern ourselves with such derivatives in this courses.

as $h \to 0$. Thus for a given $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x_0 + h) - f(x_0)| < \epsilon$$

if $|h| < \delta$, i.e.

$$|f(x) - f(x_0)| < \epsilon$$

if $|x - x_0| < \delta$. This indicates that $f(x)$ is continuous at x_0.

5.1.1 Rules Pertaining to Differentiation

Theorem (Power Rule):
If n is a rational number, then the function $f(x) = x^n$ is differentiable and

$$\frac{df(x)}{dx} = \frac{dx^n}{dx} = nx^{n-1}.$$

Theorem:
Let $f(x)$ and $g(x)$ be defined and differentiable on a set D. Then
(a). $[\alpha f(x) + \beta g(x)]' = \alpha f'(x) + \beta g'(x)$.
(b). $[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$.
(c). $[f(x)/g(x)]' = [f'(x)g(x) - f(x)g'(x)]/g^2(x)$ if $g(x) \neq 0$.

Proof:
To prove (b) we write

$$\begin{aligned}
&\lim_{h\to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\
&= \lim_{h\to 0} \frac{[f(x+h) - f(x)]g(x+h) + f(x)[g(x+h) - g(x)]}{h} \\
&= \lim_{h\to 0} \frac{f(x+h) - f(x)]g(x+h)}{h} + \lim_{h\to 0} \frac{f(x)[g(x+h) - g(x)]}{h} \\
&= \lim_{h\to 0} g(x+h) \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} + f(x) \lim_{h\to 0} \frac{g(x+h) - g(x)}{h}.
\end{aligned}$$

However, $lim_{h\to 0} g(x+h) = g(x)$, since $g(x)$ is continuous (because it is differentiable). Hence,

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

Now, to prove (c) we write

$$\begin{aligned}
&\lim h \to 0 \frac{f(x+h)/g(x+h) - f(x)/g(x)}{h} \\
&= \lim_{h\to 0} \frac{g(x)f(x+h) - f(x)g(x+h)}{hg(x)g(x+h)} \\
&= \lim_{h\to 0} \frac{g(x)[f(x+h) - f(x)] - f(x)[g(x+h) - g(x)]}{hg(x)g(x+h)} \\
&= \frac{\lim_{h\to 0}\{g(x)[f(x+h) - f(x)]/h - f(x)[g(x+h) - g(x)]/h\}}{g(x)\lim_{h\to 0} g(x+h)} \\
&= \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)}.
\end{aligned}$$

Theorem (The Chain Rule):
Let $f : D_1 \to \mathbb{R}$ and $g : D_2 \to \mathbb{R}$ be two functions. Suppose that $f(D_1) \subset D_2$. If $f(x)$ is differentiable on D_1 and $g(x)$ is differentiable on D_2, then the composite function $h(x) = g[f(x)]$ is differentiable on D_1 and

$$\frac{dg[f(x)]}{dx} = \frac{dg[f(x)]}{df(x)}\frac{df(x)}{dx}.$$

Proof:
Let $z = f(x)$ and $t = f(x+h)$. By the fact that $g(z)$ is differentiable we can write

$$\begin{aligned}
g[f(x+h)] - g[f(x)] &= g(t) - g(z) \\
&= (t-z)g'(z) + o(t-z).
\end{aligned}$$

We then have

$$\frac{g[f(x+h)] - g[f(x)]}{h} = \frac{t-z}{h}g'(z) + \frac{o(t-z)}{t-z}\frac{t-z}{h}.$$

Since f is differentiable, then it is continuous, so as $h \to 0$, $f(x+h) \to f(x)$, i.e. $t \to z$. Hence,

$$\lim_{h\to 0}\frac{t-z}{h} = \lim_{h\to 0}\frac{f(x+h) - f(x)}{h} = \frac{df(x)}{dx}.$$

Noting that

$$\lim h \to 0 \frac{o(t-z)}{t-z} = \lim_{t\to z}\frac{o(t-z)}{t-z} = 0,$$

and we conclude that

$$\frac{dg[f(x)]}{dx} = \frac{dg[f(x)]}{df(x)}\frac{df(x)}{dx}.$$

Theorem:
Let $f : D \mapsto \mathbb{R}$, where D is an open set. Suppose that $f'(x)$ is positive at a point $x_0 \in D$. Then there is a neighborhood $N_\delta(x_0) \in D$ such that for each x in this neighborhood, $f(x) > f(x_0)$ if $x > x_0$, and $f(x) < f(x_0)$ if $x < x_0$.[22]

5.2 The Mean Value Theorem

This is one of the most important theorems in differential calculus. It is also known as the theorem of the mean.

Theorem (Rolle's Theorem):
Let $f(x)$ be continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). If $f(a) = f(b)$, then there exists a point c, $a < c < b$, such that $f'(c) = 0$.

Theorem (Mean Value Theorem):
If $f(x)$ is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there exists a point c, $a < c < b$, such that

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

Proof:
Consider the function

$$F(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

The function $F(x)$ is continuous on $[a, b]$ and is differentiable on (a, b), since $F'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$. Furthermore, $F(a) = F(b) = 0$ it follows from Rolle's theorem that

[22] That is, the sign of $f'(x)$ provides information about the behavior of $f(x)$ in a neighborhood of x.

there exists a point c, $a < c < b$, such that $F'(c) = 0$. Thus

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

Theorem (Cauchy's Mean Value Theorem):
If $f(x)$ and $g(x)$ are continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there exists a point c, $a < c < b$, such that

$$[f(b) - f(a)]g'(c) = [g(b) - g(a)]f'(c).$$

Proof:
See Fulks, p.112 for a good interpretation. When the mean values theorem is a special case of Cauchy's mean values theorem when $g(x) = x$.

An immediate application of the Cauchy's mean theorem is a very popular method in calculating the limit of certain ratios of functions. This method is known as l'Hospital's rule. It deal with the limit of the ratio $f(x)/g(x)$ as $x \to a$ when both the numerator and the denominator tend simultaneously to zero or to infinity as $x \to a$. In either case, we have what is called an indeterminate ratio caused by having $0/0$ or ∞/∞ as $x \to a$.

Theorem (l'Hospital's Rule):
Let $f(x)$ and $g(x)$ be continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). Suppose that we have the following:
(a). $g(x)$ and $g'(x)$ are not zero at any point inside (a, b).
(b). $\lim_{x \to a^+} f'(x)/g'(x)$ exists.
(c). $f(x) \to 0$ and $g(x) \to 0$ as $x \to a^+$, or $f(x) \to \infty$ and $g(x) \to \infty$ as $x \to a^+$.
Then,

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

Example:
$\lim_{x \to \infty} x \log \left(\frac{x+1}{x-1}\right)$.
This is of the form $\infty \times 0$ as $x \to \infty$, which is indeterminate. But

$$\lim_{x \to \infty} x \log \left(\frac{x+1}{x-1}\right) = \frac{\log \left(\frac{x+1}{x-1}\right)}{1/x}$$

is of the form $0/0$ as $x \to \infty$. Hence,

$$\begin{aligned}\lim_{x\to\infty} x\log\left(\frac{x+1}{x-1}\right) &= \lim_{x\to\infty}\frac{\frac{-2}{(x+1)(x-1)}}{-1/x^2}\\ &= \lim_{x\to\infty}\frac{2}{(1-1/x)(1+1/x)}\\ &= 2.\end{aligned}$$

5.3 Taylor's Theorem

This theorem is also known as the general mean value theorem, since it is considered as an extension of the mean value theorem. It was used to expand functions into infinite series.

Theorem (Taylor's Theorem):
If the $(n-1)$st $(n \geq 1)$ derivative of $f(x)$, namely $f^{(n-1)}(x)$, is continuous on the closed interval $[a,b]$ and the nth derivative $f^{(n)}$ exists on the open interval (a,b), then for each $x \in [a,b]$ we have[23]

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!}f''(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!}f^{(n-1)}(a) + \frac{(x-a)^n}{n!}f^{(n)}(\xi),$$

where $a < \xi < x$.

This is known as Taylor's formula. It can also be expressed as

$$f(a+h) = f(a) + hf'(a) + \frac{h^2}{2!}f''(a) + \cdots + \frac{h^{n-1}}{(n-1)!}f^{(n-1)}(a) + \frac{h^n}{n!}(a+\theta_n h),$$

where $h = x - a$ and $0 < \theta_n < 1$.[24]

5.4 Maxima and Minima of a Function

[23] Set $x = b$ and $n = 1$, we have $f(b) = f(a) + (b-a)f'(c)$, the mean value theorem.
[24] To make $a < \xi < x$, we need $a < \xi = a + \theta_n h < a + 1 \cdot h = x$.

Definition:
A function $f: D \to \mathbb{R}$ has a local maximum at a point $x_0 \in D$ if there exists a $\delta > 0$ such that $f(x) \leq f(x_0)$ $\forall x \in N_\delta(x_0) \cap D$. The function f has a local minimum at a point $x_0 \in D$ if there exists a $\delta > 0$ such that $f(x) \geq f(x_0)$ $\forall x \in N_\delta(x_0) \cap D$.

Definition:
A function $f: D \to \mathbb{R}$ has an absolute maximum (minimum) over D if there exist a point $x^* \in D$ such that $f(x) \leq f(x^*)$ $(f(x) \geq f(x^*))$ $\forall x \in D$.

Theorem:
Let $f(x)$ be differentiable on the open interval (a, b). If $f(x)$ has a local maximum, or a local minimum, at a point x_0 in (a, b), then $f'(x_0) = 0$.

Proof:
Suppose that $f(x)$ has a local maximum at x_0. Then $f(x) \leq f(x_0)$, $\forall x \in N_\delta(x_0) \subset (a, b)$. It follows that

$$\frac{f(x) - f(x_0)}{x - x_0} \begin{cases} \leq 0 & if\ x > x_0, \\ \geq 0 & if\ x < x_0, \end{cases} \tag{2}$$

$\forall x \in N_\delta(x_0)$. As $x \to x_0^+$, the ratio in (2) have a non-positive limit, and if $x \to x_0^-$, the ration will have a nonnegative limit. Since $f'(x_0)$ exists, these two limits must be equal and equal to $f'(x_0)$ as $x \to x_0$. We therefore conclude that $f'(x_0) = 0$.

Note:
It is important to note that $f'(x_0) = 0$ is a necessary condition for a differentiable function to have a local optimum at x_0. It is not, however, a sufficient condition. That is, if $f'(x_0) = 0$, then it is not necessary true that x_0 is a point of local optimum. We shall in the next subsection make use of Taylor's expansion to come up with a condition for $f(x)$ to have local optimum at $x = x_0$.

Note:
We recall from last section that if $f(x)$ is continuous on $[a, b]$, then it must achieve its absolute optima at some points inside $[a, b]$. These points can be interior points, that is, points that belong to the open interval (a, b), or they can be end (boundary) points. In particular, if $f'(x)$ exists on (a, b), to determine the locations of the

absolute optima we must solve the equation $f'(x) = 0$ and then compare the value of $f(x)$ at the roots of this equation with $f(a)$ and $f(b)$. The largest of these values is the absolute maximum. In the event $f'(x) \neq 0$ on (a, b), then $f(x)$ must achieve its absolute optimum at an end point.

5.4.1 A Sufficient Condition for a Local Optimum

Suppose that $f(x)$ has n derivatives in a neighborhood $N_\delta(x_0)$ such that $f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0$, but $f^{(n)}(x_0) \neq 0$. Then by Taylor's theorem we have

$$f(x) = f(x_0) + \frac{h^n}{n!} f^{(n)}(x_0 + \theta_n h)$$

for any x in $N_\delta(x_0)$, where $h = x - x_0$ and $0 < \theta_n < 1$. Furthermore, if we assume that $f^{(n)}(x)$ is continuous at x_0, then

$$f^{(n)}(x_0 + \theta_n h) = f^{(n)}(x_0) + o(1),$$

where $o(1) \to 0$ as $h \to 0$. We can therefore write [25]

$$f(x) - f(x_0) = \frac{h^n}{n!} f^{(n)}(x_0) + o(h^n). \tag{3}$$

In order for $f(x)$ to have a local optimum at x_0, $f(x) - f(x_0)$ must have the same sign (positive or negative) for small values of h inside a neighborhood of 0. But from (3), the sign of $f(x) - f(x_0)$ is determined by the sign of $h^n f^{(n)}(x_0)$. We can then conclude that:

(a). If n is even, then a local optimum is achieved at x_0. In this case, a local maximum occurs at x_0 if $f^{(n)} < 0$, whereas $f^{(n)} > 0$ indicates that x_0 is a point of local minimum.

(b). If n is odd, then x_0 is not a point of local optimum, since $f(x) - f(x_0)$ changes sign around x_0.

In particular, if $f'(x_0) = 0$ and $f''(x_0) \neq 0$, then x_0 is a point of local optimum. When $f''(x_0) < 0$, $f(x)$ has a local maximum at x_0, and when $f''(x_0) > 0$, $f(x)$ has a local minimum at x_0.

[25] Think of $o(1) \cdot h^n = o(h^n)$ since by definition $o(h^n) = \frac{h^n o(1)}{h^n} = o(1)$ as $h \to 0$.

6 Infinite Sequence and Series

The study of the theory of infinite sequence and series is an integral part of advanced calculus. All limiting processes, such as differentiation and integration, can be investigated on the basis of this theory. In this chapter we shall study the theory of infinite sequences and series, and investigate their convergence. Unless otherwise stated, the terms of all sequences and series considered in this chapter are real-valued.

6.1 Infinite Sequences

Definition (Infinite Sequences):
An infinite sequence is a particular function $f: J^+ \to \mathbb{R}$ defined on the set of all positive integers. For a given $n \in J^+$, the value of this function , namely $f(n)$, is called the nth term of the infinite sequence and is denoted by a_n. The sequence itself is denoted by the symbol $\{a_n\}_{n=1}^{\infty}$.

6.1.1 Bound, Convergence and Divergence

Since a sequence is a function, then in particular, the sequence $\{a_n\}_{n=1}^{\infty}$ can have the following properties:
(a). It is bounded if there exists a constant $K > 0$ such that $|a_n| \leq K$, $\forall\, n$.
(b). It is monotone increasing if $a_n \leq a_{n+1}$ $\forall\, n$, and is monotone decreasing if $a_n \geq a_{n+1}$ $\forall\, n$.
(c). It converges to a finite number c if $\lim_{n\to\infty} a_n = c$, that is, for a given $\epsilon > 0$ there exists an integer N such that

$$|a_n - c| < \epsilon \quad if \;\; n > N.$$

In this case, c is called the limit of the sequence and this fact is denoted by writing

$$a_n \to c \quad as \;\; n \to \infty.$$

If the sequence does not converge to a finite limit, then it is said to be divergent.
(d). It is said to oscillate if it does not converge to a finite limit, nor to $+\infty$ or $-\infty$

as $n \to \infty$.

Example:
let $a_n = (n^2 + 2n)/(2n^2 + 3)$. Then $a_n \to \frac{1}{2}$ as $n \to \infty$, since

$$\lim_{n\to\infty} = \lim_{n\to\infty} \frac{1 + 2/n}{2 + 3/n^2} = \frac{1}{2}.$$

Theorem (Convergence implies Boundedness):
Every convergent sequence is bounded.

Proof:
Let $c = \lim a_n$. Then there is an N such that

$$|a_n - c| < 1 \quad if \;\; n > N.$$

Then

$$|a_n| - |c| \leq |a_n - c| < 1.$$

Thus

$$|a_n| < |c| + 1 \equiv K_1 \quad if \;\; n > N. \tag{4}$$

Now we look among the number $|a_1, |a_2|, ..., |a_N|$, and choose the largest, calling it K_2. Then

$$|a_n| \leq K_2 \quad if \;\; n \leq N. \tag{5}$$

It is clear that if we take K to be the larger of K_1 and K_2, then by (4) and (5),

$$|a_n| < K \quad \forall\, n.$$

The converse of this theorem is not necessarily true. That is, if a sequence is bounded, then it does not have to be convergent.[26] To guarantee converge of a bounded sequence we obviously need an additional condition.

[26] As a example, consider the sequence $a_n = (-1)^n$. The sequence is bounded, but is not convergent.

Theorem:
Every bounded monotone sequence converges.

Theorem:
(a). If $\{a_n\}_{n=1}^{\infty}$ is bounded from above and is monotone increasing, then $\{a_n\}_{n=1}^{\infty}$ converges to $c = \sup_{n\geq 1} a_n$.
(b). If $\{a_n\}_{n=1}^{\infty}$ is bounded from below and is monotone decreasing, then $\{a_n\}_{n=1}^{\infty}$ converges to $d = \inf_{n\geq 1} a_n$

Definition (Subsequence):
A sequence $\{a_n\}$ is a function on the integers; that is

$$a_n = f(n) \quad for\ all \ \ n \geq n_0.$$

Now suppose that we consider the function restricted to a subset of the integers. Let us choose an integer greater than or equal to n_0 and denote it by n_1, another greater than n_1 and denote it by n_2, another greater than n_2 and denote it by n_3, and so forth. Then the new sequence, defined by

$$b_k = a_{n_k} = f(n_k), \quad k = 1, 2, ...,$$

we call a subsequence of $\{a_n\}$. It is clear that there are many subsequences of a given sequence. If we assume the $n_0 \geq 1$, then $n_k \geq k$. The theorem we want to prove is the following.

Theorem:
Suppose that $\{a_n\}$ converges; then any subsequence $\{a_{n_k}\}$ also converges and has the same limit.

Proof:
Let A be the limit of the sequence $\{a_n\}$. We know that for each $\varepsilon > 0$ there is an N for which

$$|a_{n_k} - A| < \varepsilon \quad if \ \ n_k > N.$$

But $n_k \geq k$; hence

$$n_k > N \quad if \;\; k > N.$$

Thus

$$|b_k - A| = |a_{n_k} - A| < \varepsilon \quad if \;\; k > N.$$

Note:
It should be noted that if a sequence diverges, then it does not necessarily follows that every one of its subsequences must diverge. A sequence may fail to converge, yet several of its subsequence converges. We have noted that a bounded sequence may not converge. It is possible, however, that one of its subsequences is convergent.

Theorem:
Every bounded sequence has a convergent subsequence.

Definition (Upper and Lower Limit of a Sequence):
Let $\{a_n\}_{n=1}^{\infty}$ be a bounded sequence,[27] and let E be the set of all its subsequential limits. Then the least upper bound of E is called the upper limit of $\{a_n\}_{n=1}^{\infty}$ and is denoted by $\limsup_{n\to\infty} a_n$. Similarly, the greatest lower bound of E is called the lower limit of $\{a_n\}_{n=1}^{\infty}$ and is denoted by $\liminf_{n\to\infty} a_n$.

Example:
The sequence $\{a_n\}_{n=1}^{\infty}$, where $a_n = (-1)^n[1 + (1/n)]$, has two subsequential limits, namely -1 and $+1$. Thus $E = \{-1, 1\}$, and $\limsup_{n\to\infty} = 1$ and $\liminf_{n\to\infty} = -1$.

Theorem:
The sequence $\{a_n\}_{n=1}^{\infty}$ converges to c if and only if

$$\lim \sup_{n\to\infty} = \lim \inf_{n\to\infty} = c.$$

[27] Therefore it has a convergent subsequence.

6.1.2 The Cauchy Criterion

We have seen earlier that the definition of convergence of a sequence $\{a_n\}_{n=1}^{\infty}$ requires finding the limit of a_n as $n \to \infty$. In some cases, such a limit may be difficult to figure out. [28] Fortunately, however, there is another convergence criterion for sequence, known as the Cauchy Criterion after Augustin-Louis Cauchy.

Theorem (The Cauchy Criterion):
The sequence $\{a_n\}_{n=1}^{\infty}$ converges if and only if it satisfies the following condition, known as the ϵ-condition: For each $\epsilon > 0$ there is an integer N such that

$$|a_m - a_n| < \epsilon \quad for\ all\ m > N,\ n > N.$$

Definition:
A sequence $\{a_n\}_{n=1}^{\infty}$ that satisfies the ϵ-condition of the Cauchy criterion is said to be a Cauchy sequence.

6.2 Infinite Series

Let $\{a_n\}_{n=1}^{\infty}$ be a given sequence. Consider the symbolic expression

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + \cdots + a_n + \cdots .$$

By definition, this expression is called an infinite series, or just a series for simplicity, and a_n is referred as the nth term of the series. The finite sum

$$s_n = \sum_{i=1}^{n} a_i, \quad n = 1, 2, ...,$$

[28] Consider the sequence whose nth term is $a_n = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots + \frac{(-1)^{n-1}}{2n-1},\ \ n = 1, 2, ...$ It is not easy to calculate the limit of a_n in order to find out if the sequence converges.

is called the nth partial sum of the series.

Definition (Convergence of an Infinite Series):
Consider the series $\sum_{n=1}^{\infty} a_n$. Let s_n be its nth partial sum ($n = 1, 2, ...$).
(a). The series is said to be convergent if the sequence $\{s_n\}_{n=1}^{\infty}$ converges. In this case, if $\lim_{n\to\infty} s_n = s$, where s is finite, then we say that the series converges to s, or that s is **the sum** of the series. This expressed by writing

$$s = \sum_{n=1}^{\infty} a_n.$$

(b). If s_n does not tend to a finite limit, then the series is said to be divergent.

Theorem:
The series $\sum_{n=1}^{\infty} a_n$, converges if and only if for a given $\epsilon > 0$ there is an integer N such that

$$\left| \sum_{i=m+1}^{n} a_i \right| < \epsilon \quad \textit{for all } n > m > N. \tag{6}$$

Eq.(6) follows from applying Cauchy Criterion to the sequence $\{s_n\}_{n=1}^{\infty}$ of partial sums of the series and noting that

$$|s_n - s_m| = \left| \sum_{i=m+1}^{n} a_i \right| \quad \textit{for} \;\; n > m.$$

In particular, if $n = m + 1$, the (6) becomes

$$|a_{m+1}| < \epsilon$$

for all $m > N$. This implies that $\lim_{m\to\infty} a_{m+1} = 0$ and hence $\lim_{n\to\infty} a_n = 0$. We therefore conclude the following result:

Result:
If $\sum_{n=1}^{\infty} a_n$ is a convergent series, then $\lim_{n\to\infty} a_n = 0$.

It is important here to note that the convergence of the nth term of series to zero as $n \to \infty$ is a necessary condition for the convergence of the series. It is not,

however, a sufficient condition, that is, if $\lim_{n\to\infty} a_n = 0$, then it **do not** follow that $\sum_{n=1}^{\infty} a_n$ converges. It is true, however, that if $\lim_{n\to\infty} a_n \neq 0$, then $\sum_{n=1}^{\infty} a_n$ is divergent. This follows from applying the law of contraposition to the necessary condition of convergence. We conclude the following:
(a). If $a_n \to 0$ as $n \to \infty$, then no conclusion can be reached regarding the convergence or divergence of $\sum_{n=1}^{\infty} a_n$.
(b). If $a_n \nrightarrow 0$ as $n \to \infty$, then $\sum_{n=1}^{\infty} a_n$ is divergent.

Example:
One of the simplest series is the geometric series, $\sum_{n=1}^{\infty} a^n$. This series is divergent if $|a| \geq 1$, since $\lim_{n\to\infty} a^n \neq 0$. It is convergent if $|a| < 1$ by the Cauchy criterion: Let $n > M$. Then

$$s_n - s_m = a^{m+1} + a^{m+2} + \cdots + a^n. \tag{7}$$

By multiplying the two sides of (7) by a, we get

$$a(s_n - s_m) = a^{m+2} + a^{m+3} + \cdots + a^{n+1}. \tag{8}$$

If we now subtract (8) from (7), we obtain

$$s_n - s_m = \frac{a^{m+1} - a^{n+1}}{1-a}. \tag{9}$$

Since $|a| < 1$, we can find an integer N such that for $m > N, n > N$,

$$|a|^{m+1} < \frac{\epsilon(1-a)}{2},$$
$$|a|^{n+1} < \frac{\epsilon(1-a)}{2}.$$

Hence, for a given $\epsilon > 0$,

$$|s_n - s_m| < \epsilon \quad if\ n > m > N.$$

Eq. (9) can actually be used to find the sum of the geometric series when $|a| < 1$. Let $m = 1$. By taking the limits of both sides of (9) as $n \to \infty$ we get

$$\begin{aligned}
\lim_{n\to\infty} s_n &= s_1 + \frac{a^2}{1-a} \quad since\ \lim_{n\to\infty} a^{n+1} = 0, \\
&= a + \frac{a^2}{1-a} \\
&= \frac{a}{1-a}.
\end{aligned}$$

Theorem:
If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty}$ are two convergent series, and if c is a constant, then the following series are also convergent:
(a). $\sum_{n=1}^{\infty}(ca_n) = c\sum_{n=1}^{\infty} a_n$.
(b). $\sum_{n=1}^{\infty}(a_n + b_n) = \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n$.

Definition (Absolutely Convergent):
The series $\sum_{n=1}^{\infty} a_n$ is absolutely convergent if $\sum_{n=1}^{\infty} |a_n|$ is convergent.

Theorem:
Every absolutely convergent series is convergent.

Deginition (Conditional Convergent):
In the case that $\sum_{n=1}^{\infty} a_n$ is convergent while $\sum_{n=1}^{\infty} |a_n|$ is divergent, then the series $\sum_{n=1}^{\infty} a_n$ is said to be conditionally convergent.

6.2.1 Multiplication of Series

There are several ways to define the product of two series. We shall consider the so-called Cauchy's product.

Definition (Cauchy's Product):
Let $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty}$ be two series in which the summation index start at zero instead of one. Cauchy's product of these two series is the series $\sum_{n=0}^{\infty} c_n$, where

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}, \quad n = 0, 1, 2, ...,$$

that is,

$$\sum_{n=0}^{\infty} c_n = a_0 b_0 + (a_0 b_1 + a_1 b_0) + (a_0 b_2 + a_1 b_1 + a_2 b_0) + \cdots .$$

Theorem:
Let $\sum_{n=0}^{\infty} c_n$ be the Cauchy's product of $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$. Suppose that these two series are convergent and have sums equal to s and t, respectively.
(a). If at least one of $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$ converges absolutely, then $\sum_{n=0}^{\infty} c_n$ converges and its sum equal to st.
(b). If both series are absolutely convergent, then $\sum_{n=0}^{\infty} c_n$ converges absolutely to the product st.

6.3 Sequences and Series of Functions

All the sequences and series considered thus far in this chapter had constant terms. We now extend our study to sequence and series whose terms are function of x.

Definition:
Let $\{f_n(x)\}_{n=1}^{\infty}$ be a sequence of functions defined on a set $D \subset \mathbb{R}$.
(a). If there exists a function $f(x)$ defined on D such that for every $x\ inD$,

$$\lim_{n\to\infty} f_n(x) = f(x),$$

then the sequence $\{f_n(x)\}_{n=1}^{\infty}$ is said to converge to $f(x)$ on D. Thus for a given $\epsilon > 0$ there exists an inter N such that $|f_n(x) - f(x)| < \epsilon$ if $n > N$. In general, N depends on ϵ as well on x. In particular, if N depends on ϵ but not on $x \in D$, then the sequence $\{f_n(x)\}_{n=1}^{\infty}$ is said to converge uniformly to $f(x)$ on D.
(b). If $\sum_{n=1}^{\infty} f_n(x)$ converges for every x in D to $s(x)$, then $s(x)$ is said to be the sum of the series. In this case, for a given $\epsilon > 0$ there exists an inter N such that

$$|s_n(x) - s(x)| < \epsilon \quad if\ n > N,$$

where $s_n(x)$ is the nth partial sum of the series $\sum_{n=1}^{\infty} f_n(x)$. The integer N depends on ϵ and in general, on x also. In particular, if N depends on ϵ but not on $x \in D$, then the series $\sum_{n=1}^{\infty} f_n(x)$ is said to converge uniformly to $s(x)$ on D.

Theorem:
Let $\{f_n(x)\}_{n=1}^{\infty}$ be a sequence of functions depends on $D \subset \mathbb{R}$ and converging to $f(x)$. Define the number λ_n as

$$\lambda_n = \sup_{x \in D} |f_n(x) - f(x)|.$$

Then the sequence converges uniformly to $f(x)$ on D iff $\lambda_n \to 0$ as $n \to \infty$.

Proof:

Sufficiency: Suppose that $\lambda_n \to 0$ as $n \to \infty$. Then there exists an inter $N(\epsilon)$ such that

$$|\lambda_n - 0| < \epsilon \quad if \ n > N(\epsilon),$$

i.e., for $n > N$, $\lambda_n < \epsilon$. Hence for such value of n,

$$|f_n(x) - f(x)| \leq \lambda_n < \epsilon$$

for all $x \in D$. Since N depends only on ϵ, the sequence $\{f_n(x)\}_{n=1}^{\infty}$ converges uniformly to $f(x)$ on D.

Necessity: Suppose that $f_n(x) \to f(x)$ uniformly on D. Then for each $\epsilon > 0$ there is an $N(\epsilon)$, independent of x, for which

$$|f_n(x) - f(x)| < \epsilon \quad if \ n > N.$$

This ϵ is an upper bound for the number $|f_n(x) - f(x)|$. Hence the least upper bound λ_n is also less equal to ϵ. That is,

$$0 \leq \lambda_n \leq \epsilon \quad if \ \ n > N.$$

Thus $\lambda_n \to 0$.

6.3.1 Properties of Uniformly Convergent Sequences and Series

Sequence and series of functions that are uniformly convergent have several interesting properties. We shall study some

Theorem:
Let $\{f_n(x)\}_{n=1}^{\infty}$ be uniformly convergent to $f(x)$ on a set D. If for each n, $f_n(x)$ has a limit τ_n as $x \to x_0$, where x_0 is a limit point of D, then the sequence $\{\tau_n\}_{n=1}^{\infty}$ converges to $\tau_0 = \lim_{x\to x_0} f(x)$. This is equivalent to stating that

$$\lim_{n\to\infty}\left[\lim_{x\to x_0} f_n(x)\right] = \lim_{x\to x_0}\left[\lim_{n\to\infty} f_n(x)\right].$$

Corollary:
Let $\{f_n(x)\}_{n=1}^{\infty}$ be a sequence of continuous functions that converges uniformly to $f(x)$ on a set D. Then $f(x)$ is continuous on D.

Proof:
We must show that, for each $\epsilon > 0$, there is a $\delta(\epsilon)$ for which

$$|f(x) - f(x_0)| < \epsilon \quad if\ |x - x_0| < \delta.$$

For any n,

$$|f(x) - f(x_0)| = |f(x) - f_n(x) + f_n(x) - f_n(x_0) + f_n(x_0) - f(x_0)|$$

so

$$|f(x) - f(x_0)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)|. \tag{10}$$

By uniform convergence, there is an $N(\epsilon)$, independent of x, for which

$$|f_n(x) - f(x)| < \epsilon/3, \quad if\ n > N,\ \forall x \in D. \tag{11}$$

From (10) and (11), we then have

$$|f(x) - f(x_0)| \le \epsilon/3 + |f_n(x) - f_n(x_0)| + \epsilon/3. \tag{12}$$

But $f_n(x)$ is a continuous function. Hence there is a $\delta(\epsilon)$ for which

$$|f_n(x) - f_n(x_0)| < \epsilon/3 \quad if\ |x - x_0| < \delta.$$

Putting all this together, we have

$$|f(x) - f(x_0)| < \epsilon \quad if\ |x - x_0| < \delta.$$

7 Integration

The origin of integral calculus can be tracked back to the ancient Greeks. They are motivated by the need to measure the length of a curve, the area of a surface, or the volume of a solid. In the present chapter we shall study integration of real-valued functions of a single variable x according to the concepts put forth by the German mathematician Georg Friedrich Riemann (1826-1866). He was the first to establish a rigorous analytical foundation for integration, based on the older geometric approach.

7.1 Some Basic Definitions

Let $f(x)$ be a function defined and bounded on a finite interval $[a, b]$.Suppose that this interval is partitioned into a finite number of subintervals by a set of points $P = \{x_0, x_1, ..., x_n\}$ such that $a = x_0 < x_1 < x_2 < \cdots < x_n = b$. This set is called a partition of $[a, b]$. Let $\triangle x_i = x_i - x_{i-1}$ $(i = 1, 2, ..., n)$, and $\triangle_p$ be the largest of $\triangle x_1, \triangle x_2, ..., \triangle x_n$. This value is called the norm of P and is denoted by $\|P\|$. Consider the sum

$$S(P, f) = \sum_{i=1}^{n} f(t_i)\triangle x_i,$$

where t_i is a point in the subinterval $[x_{i-1}, x_i]$, $i = 1, 2, ..., n$.

Definition: (Riemann Integrable):
The function $f(x)$ is said to be Riemann integrable on $[a, b]$ if a number A exists with the following property: For any given $\epsilon > 0$ there exists a number $\delta > 0$ such that

$$|S(P, f) - A| < \epsilon$$

for any partition P of $[a, b]$ with a norm $\|P\| < \delta$, and for any choice of the points t_i in $[x_{i-1}, x_i]$, $i = 1, 2, ..., n$. This is expressed as

$$\lim_{\|P\|\to 0} S(P, f) = A.$$

The number A is called the Riemann integral of $f(x)$ on $[a, b]$ and is denoted by $\int_a^b f(x)dx$.

In order to investigate the existence of the Riemann integral, we shall need the following theorem:

Theorem (The Existence of the Riemann Integral):
Let $f(x)$ be a bounded function on a finite interval $[a, b]$. For every partition $P = \{x_0, x_1, ..., x_n\}$ of $[a, b]$, Let m_i and M_i be, respectively, the infimum and supremum of $f(x)$ on $[x_{i-1}, x_i]$, $i = 1, 2, ..., n$.
(a). If, for a given $\epsilon > 0$ there exist a $\delta > 0$ such that[29]

$$US_P(f) - LS_P(f) < \epsilon$$

whenever $\|P\| < \delta$ and

$$LS_P(f) = \sum_{i=1}^{n} m_i \triangle x_i,$$
$$US_P(f) = \sum_{i=1}^{n} M_i \triangle x_i,$$

then $f(x)$ is Riemann integrable on $[a, b]$. Conversely
(b). If $f(x)$ is Riemann integrable, then

$$US_P(f) - LS_P(f) < \epsilon$$

for any partition P such that $\|P\| < \delta$.

Example:
Let $f(x): [0, 1] \to \mathbb{R}$ be the function $f(x) = x^2$. Then, $f(x)$ is Riemann integrable on $[0, 1]$. To show this, let $P = \{x_0, x_1, ..., x_n\}$ be any partition of $[0, 1]$, where $x_0 = 0,\ x_n = 1$. Then

$$\begin{aligned} LS_P(f) &= \sum_{i=1}^{n} x_{i-1}^2 \triangle x_i, \\ US_P(f) &= \sum_{i=1}^{n} x_i^2 \triangle x_i. \end{aligned}$$

Hence,

$$\begin{aligned} US_P(f) - LS_P(f) &= \sum_{i=1}^{n} (x_i^2 - x_{i-1}^2) \triangle x_i \\ &\leq \|P\| \sum_{i=1}^{n} (x_i^2 - x_{i-1}^2). \end{aligned}$$

[29] That is. the lower sum LS_P and the upper sum US_P is equal.

But

$$\sum_{i=1}^{n}(x_i^2 - x_{i-1}^2) = x_n^2 - x_0^2 = 1.$$

Thus,

$$US_P(f) - LS_P(f) \leq \|P\|.$$

It follows that for a given $\epsilon > 0$ we can choose $\delta = \epsilon$ such that for any partition P whose norm $\|P\| < \delta$,

$$US_P(f) - LS_P(f) < \epsilon.$$

Thus, $f(x) = x^2$ is Riemann integrable on $[0, 1]$.

Example:

Consider the function $f(x) : [0, 1] \mapsto \mathbb{R}$ such that $f(x) = 0$ if x is a rational number and $f(x) = 1$ if x is irrational. Since every (countable) subinterval of $[0, 1]$ contains both rational and irrational numbers, then for any partition $P = \{x_0, x_1, \cdots, x_n\}$ of $[0, 1]$ we have

$$US_P(f) = \sum_{i=1}^{n} M_i \triangle x_i = 1,$$
$$LS_P(f) = \sum_{i=1}^{n} m_i \triangle x_i = 0.$$

It follows that

$$\inf_P US_P(f) = 1 \quad and \quad \sup_P LS_P(f) = 0.$$

Therefore $f(x)$ is not Riemann integrable on $[a, b]$.

7.2 Some Classes of Functions That Are Riemann Integrable

There are certain classes of functions that are Riemann integrable. Identifying a given function as a member of such a class can facilitate the determination of its Riemann integrability. Some of there classes of functions include: (a) continuous functions; (b) mon0tone function; (c) functions of bounded variation.

Theorem:
If $f(x)$ is continuous on $[a,b]$, then it is Riemann integrable there.

Proof:
Since $f(x)$ is continuous on a closed and bounded interval, then it must be uniformly continuous on $[a,b]$. Consequently, for a given $\epsilon > 0$ there exists a $\delta > 0$ that depends only on ϵ such that for any x_1, x_2 in $[a,b]$ we have

$$|f(x_1) - f(x_2)| < \frac{\epsilon}{b-a}$$

if $|x_1 - x_2| < \delta$. Let $P = \{x_0, x_1, ..., x_n\}$ be a partition of P with a norm $\|P\| < \delta$. Then

$$US_P(f) - LS_P(f) = \sum_{i=1}^{n}(M_i - m_i)\triangle x_i,$$

where m_i and M_i are, respectively, the infimum and supremum of $f(x)$ on $[x_{i-1}, x_i], i = 1, 2, ..., n$. By Corollary 3.4. there exist points ξ_i, η_i in $[x_{i-1}, x_i]$ such that $m_i = f(\xi_i), M_i = f(\eta_i), i = 1, 2, ..., n$. Since $|\eta_i - \xi_i| \leq \|P\| < \delta$ for $i = 1, 2, ..., n$, then

$$\begin{aligned} US_P(f) - LS_P(f) &= \sum_{i=1}^{n}[f(\eta_i) - f(\xi_i)]\triangle x_i \\ &< \frac{\epsilon}{b-a}\sum_{i=1}^{n}\triangle x_i = \epsilon. \end{aligned}$$

By Theorem of the existence of Riemann Integral, we conclude that $f(x)$ is Riemann integrable on $[a,b]$.

7.3 Properties of the Riemann Integral

The Riemann integral has several properties that are useful at both the theoretical and practical levels. The definition of the Riemann integral of f on the interval $[a,b]$ specifies that $a < b$. Now, however, it is convenient to extend the definition to cover cases in which $a = b$ or $a > b$. Geometrically, the following two definitions seem reasonable. For instance, it makes sense to define the area of a region of zero width and finite height to be 0.

Definition:
(a). If f is defined at $x = a$, then $\int_a^a f(x)dx = 0$.
(b). If f is integrable on $[a, b]$, then $\int_b^a f(x)dx = -\int_a^b f(x)dx$.

Theorem:
If $f(x)$ and $g(x)$ are Riemann integrable on $[a, b]$ and if c_1 and c_2 are constants, then $c_1 f(x) + c_2 g(x)$ is Riemann integrable on $[a, b]$, and

$$\int_a^b [c_1 f(x) + c_2 g(x)]dx = c_1 \int_a^b f(x)dx + c_2 \int_a^b g(x)dx.$$

Theorem:
If $f(x)$ is Riemann integrable on $[a, b]$ and if $a < c < b$, then

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Threoem:
If $f(x)$ and $g(x)$ are Riemann integrable on $[a, b]$, then so is their product $f(x)g(x)$.

Theorem:
If $f(x)$ is Riemann integrable on $[a, b]$, and $m \le f(x) \le M$ for all x in $[a, b]$, then

$$m(b-a) \le \int_a^b f(x)dx \le M(b-a).$$

Theorem (The Mean Value Theorem for Integrals):
If $f(x)$ is continuous on $[a, b]$, then there exist a point $c \in [a, b]$ such that

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

Proof:
By Theorem above we have

$$m \leq \frac{1}{b-a}\int_a^b f(x)dx \leq M,$$

where m and M are, respectively, the infimum and supremum of $f(x)$ on $[a,b]$. By the intermediate- value theorem, there is a point $c \in [a,b]$ such that

$$f(c) = \frac{1}{b-a}\int_a^b f(x)dx.$$

Definition (Indefinite Integral):
Let $f(x)$ be Riemann integrable on $[a,b]$. The function

$$F(x) = \int_a^x f(t)dt, \quad a \leq x \leq b,$$

is called an indefinite integral of $f(x)$.

Theorem:
If $f(x)$ is Riemann integrable on $[a,b]$, then $F(x) = \int_a^x f(t)dt$ is uniformly continuous on $[a,b]$.

Proof:
Let x_1, x_2 be in $[a,b]$, $x_1 < x_2$. Then

$$\begin{aligned}
|F(x_2) - F(x_1)| &= \left|\int_a^{x_2} f(t)dt - \int_a^{x_1} f(t)dt\right| \\
&= \left|\int_{x_1}^{x_2} f(t)dt\right| \\
&\leq \int_{x_1}^{x_2} |f(t)dt \\
&\leq M'(x_2 - x_1),
\end{aligned}$$

where M' is the supremum of $|f(x)|$ on $[a,b]$. Thus if $\epsilon > 0$ is given, then $|F(x_2) - F(x_1)| < \epsilon$ provided that $|x_2 - x_1| < \epsilon/M'$. This provides uniform continuity of $F(x)$ on $[a,b]$.

Theorem (Fundamental Theorem of Calculus):
Suppose that $f(x)$ is continuous on $[a,b]$. Let $F(x) = \int_a^x f(t)dt$. Then we have the following:
(a). $\frac{dF(x)}{dx} = f(x),\ a \leq x \leq b.$
(b). $\int_a^b f(x)dx = G(b) - G(a) = F(b) - F(a)$, where $G(x) = F(x) + c$, and c is an arbitrary constant.

Proof:
We have

$$\begin{aligned} \frac{dF(x)}{dx} &= \frac{d}{dx}\int_a^x f(t)dt = \lim_{h\to 0}\frac{1}{h}\left[\int_a^{x+h} f(t)dt - \int_a^x f(t)dt\right] \\ &= \lim_{h\to 0}\frac{1}{h}\int_x^{x+h} f(t)dt \\ &= \lim_{h\to 0} f(x+\theta h), \quad (by\ mean\ value\ theorem\ for\ integral) \end{aligned}$$

where $0 \leq \theta \leq 1$. Hence,

$$\frac{dF(x)}{dx} = \lim_{h\to 0} f(x+\theta h) = f(x)$$

by the continuity of $f(x)$.

To prove the second part of the theorem, let $G(x)$ be defined on $[a,b]$ as

$$G(x) = F(x) + c = \int_a^x f(t)dt + c,$$

that is, $G(x)$ is an indefinite integral of $f(x)$. If $x = a$, then $G(a) = c$ since $F(a) = 0$. Also, if $x = b$, then $G(b) = F(b) + c = \int_a^b f(t)dt + G(a)$. It follows that

$$\int_a^b f(t)dt = G(b) - G(a) = F(b) + c - F(a) - c = F(b) - F(a).$$

Example:
Evaluate $\int_0^2 (2x^2 - 3x + 2)dx$.

Solution:

$$\begin{aligned} \int_0^2 (2x^2 - 3x + 2)dx &= \left[\frac{2x^3}{3} - \frac{3x^2}{2} + 2x\right]_0^2 \\ &= \left(\frac{16}{3} - 6 + 4\right) - 0 \\ &= \frac{10}{3}. \end{aligned}$$

Example:
Evaluate $\frac{d}{dx}\left[\int_0^x \sqrt{t^2+1}dt\right]$.

Solution:
Note that $f(t) = \sqrt{t^2+1}$ is continuous on the entire real line. So using the fundamental theorem of calculus, we can write

$$\frac{d}{dx}\left[\int_0^x \sqrt{t^2+1}dt\right] = \sqrt{x^2+1}.$$

7.4 Change of Variables in Riemann Integration

There are situations in which the variable x in a Riemann integral is a function of some other variable, sat u. In this case, it may be of interest to determine how the integral can be expressed and evaluated under the given transformation.

Theorem (Integration by Substitution):
Let $f(x)$ be continuous on $[\alpha, \beta]$, and let $x = g(u)$ be a function whose derivative $g'(u)$ exists and is continuous on $[c, d]$. Suppose that the range of g is contained inside $[\alpha, \beta]$. If a, b are points in $[\alpha, \beta]$ such that $a = g(c)$ and $b = g(d)$, then

$$\int_a^b f(x)dx = \int_c^d f[g(u)]g'(u)du.$$

Proof:

Let $F(x) = \int_a^x f(t)dt$, then $F'(x) = f(x)$. According to the chain rule

$$\begin{aligned} \frac{dF(g(u))}{du} &= \frac{dF(g(u))}{dg(u)} \frac{dg(u)}{du} \\ &= f(g(u))g'(u). \end{aligned}$$

By the fundamental theorem,

$$\begin{aligned} \int_c^d f(g(u))g'(u)du &= F(g(u))|_c^d \\ &= F(g(d)) - F(g(c)) \\ &= F(b) - F(a) \\ &= \int_a^b f(x)dx. \end{aligned}$$

Example:
Evaluate $\int_0^1 u(u^2+1)^3 du$.

Solution:
To evaluating this integral, let $x = g(u) = (u^2+1)$, $g'(u) = 2u$ and $f(x) = x^3$. So $g(0) = 1$ and $g(1) = 2$. Then

$$\begin{aligned} \int_0^1 u(u^2+1)^3 du &= \frac{1}{2}\int_0^1 g'(u)f(g(u))du \\ &= \frac{1}{2}\int_1^2 x^3 dx \\ &= \frac{15}{8}. \end{aligned}$$

Theorem (Integration by Parts):
If f and g are differentiable on $[a, b]$, and if f' and g' are integrable there, then

$$\int_a^b f(x)g'(x)dx = f(x)g(x)|_a^b - \int_a^b g(x)f'(x)dx.$$

Proof:

We have the familiar formula

$$(fg)' = fg' + f'g.$$

Integrating this formula

$$f(x)g(x)|_a^b = \int_a^b [f(x)g(x)]'dx = \int_a^b f(x)g'(x)dx + \int_a^b f'(x)g(x)dx.$$

Example:

Evaluate $\int x^2 \ln x dx$.

Solution:

In this case, x^2 is more easily integrated than $\ln x$, and the derivative of $\ln x$ is simpler than $\ln x$. There we set $g(x) = \ln x$, $g'(x) = \frac{1}{x}$ and $f'(x) = x^2$, $f(x) = 1/3x^3$. So

$$\begin{aligned}
\int x^2 \ln x dx = \int f'(x)g(x)dx &= f(x)g(x) - \int f(x)g'(x)dx \\
&= \frac{1}{3}x^3 \ln x - \frac{1}{3}\int x^3 \frac{1}{x} dx \\
&= \frac{1}{3}x^3 \ln x - \frac{1}{3}\int x^2 dx \\
&= \frac{1}{3}x^3 \ln x - \frac{x^3}{9} + C.
\end{aligned}$$

7.5 Improper Riemann Integrals

In our study of the Riemann integral we have only considered integrals of functions that are bounded (talking about $f(x)$) on a finite interval $[a, b]$. We now extend the scope of Riemann integration to include situations where the integrand can became unbounded at one or more points inside the range of integration, which can also be infinite. In such situations, the Riemann integral is called an improper integral.

There are two kinds of improper integrals. If $f(x)$ is Riemann integrable on $[a, b]$ for any $b > a$, then $\int_a^\infty f(x)dx$ is called an improper integral of the first kind, where the range of integration if infinite. If, however, $f(x)$ becomes infinite at a finite number of points inside the range of integration, then the integral $\int_a^b f(x)dx$ is said to be improper of the second kind.

7.5.1 Improper Riemann Integrals of the First Kind

Definition:
Let $F(z) = \int_a^z f(x)dx$. Suppose that $F(z)$ exists for any value of z greater than a. If $F(z)$ has a finite limit L as $z \to \infty$, then the improper integral $\int_a^\infty f(x)dx$ is said to converge to L. In this case, L represent the Riemann integral of $f(x)$ on $[a, \infty]$ and we write

$$L = \int_a^\infty f(x)dx.$$

On the other hand, if $L = \pm\infty$, then the improper integral $\int_a^\infty f(x)dx$ is said to diverge. By the same token, we can define the integral $\int_{-\infty}^a f(x)dx$ as the limit, if it exists, of $\int_{-z}^a f(x)dx$ as $z \to \infty$. Also, $\int_{-\infty}^\infty f(x)dx$ is defined as

$$\int_{-\infty}^\infty f(x)dx = \lim_{u\to\infty} \int_{-u}^a f(x)dx + \lim_{z\to\infty} \int_a^z f(x)dx,$$

where a is any finite number, provided that both limit exist.

Example (An improper integral that diverges):
Evaluate $\int_1^\infty \frac{1}{x}dx$.

Solution:

$$\begin{aligned}
\int_1^\infty \frac{1}{x}dx &= \lim_{z\to\infty} \int_1^z \frac{1}{x}dx \\
&= \lim_{z\to\infty} [\ln x]_1^z \\
&= \lim_{z\to\infty} (\ln z - 0) \\
&= \infty.
\end{aligned}$$

Example (An improper integral that converges):
Evaluate $\int_0^\infty e^{-x}dx$.

Solution:

$$\begin{aligned}\int_0^\infty e^{-x}dx &= \lim_{z\to\infty}\int_0^z e^{-x}dx\\ &= \lim_{z\to\infty}\left[-e^{-x}\right]_0^z\\ &= \lim_{z\to\infty}(-e^{-z}+1)\\ &= 1.\end{aligned}$$

Exercise:
Evaluate $\int_1^\infty (1-x)e^{-x}dx$.

7.5.2 Improper Riemann Integrals of the Second Kind

Let us now consider integrals of the form $\int_a^b f(x)dx$ where $[a,b]$ is a finite interval and the integrand become infinite at a finite number points **inside** $[a,b]$. Such integrals are called improper integrals of the second kind.

Definition (End point):
Suppose that $f(x)\to\infty$ as $x\to a^+$, then $\int_a^b f(x)dx$ is said to be converge if the limit

$$\lim_{\epsilon\to 0^+}\int_{a+\epsilon}^b f(x)dx$$

exists and is finite.

Definition (End point):
Suppose that $f(x)\to\infty$ as $x\to b^-$, then $\int_a^b f(x)dx$ is said to be converge if the limit

$$\lim_{\epsilon\to 0^+}\int_a^{b-\epsilon} f(x)dx$$

exists and is finite.

Definition (Interior point):
Suppose that $f(x) \to \infty$ as $x \to c$, where $a < c < b$, then $\int_a^b f(x)dx$ is the sum of $\int_a^c f(x)dx$ and $\int_c^b f(x)dx$ provided that both integrals converges.

Definition:
If $f(x) \to \infty$ as $x \to x_0$, where $x_0 \in [a, b]$, then x_0 is said to be a **singularity** of $f(x)$.

Example:
Evaluate $\int_0^1 \frac{1}{\sqrt[3]{x}}dx$.

Solution:
The integral is improper of the second since $\frac{1}{\sqrt[3]{x}} \to \infty$ as $x \to 0^+$. Thus, $x = 0$ is a singularity of the integrand. We can evaluate this integral as follows.

$$\begin{aligned} \int_0^1 x^{-1/3}dx &= \lim_{a\to 0^+} \left[\frac{x^{2/3}}{2/3}\right]_a^1 \\ &= \lim_{a\to 0^+} \frac{3}{2}(1 - a^{2/3}) \\ &= \frac{3}{2}. \end{aligned}$$

Example:
Evaluate $\int_0^2 \frac{1}{x^3}dx$.

Solution:
The integral is improper of the second since $\frac{1}{x^3} \to \infty$ as $x \to 0^+$. Thus, $x = 0$ is a singularity of the integrand. We can evaluate this integral as follows.

$$\begin{aligned} \int_0^2 \frac{1}{x^3}dx &= \lim_{a\to 0^+} \left[-\frac{1}{2x^2}\right]_a^2 \\ &= \lim_{a\to 0^+} \left(\frac{1}{8} + \frac{1}{2a^2}\right) \\ &= \infty. \end{aligned}$$

Example:
Evaluate $\int_{-1}^{2} \frac{1}{x^3} dx$.

Solution:
The integral is improper of the second since $\frac{1}{x^3} \to \infty$ as $x \to 0^+$. Thus, $x = 0$ is a singularity of the integrand. We can evaluate this integral as follows.

$$\int_{-1}^{2} \frac{1}{x^3} dx = \int_{-1}^{0} \frac{1}{x^3} dx + \int_{0}^{2} \frac{1}{x^3} dx.$$

From the example above we know that the second integral diverge. Therefore, the original improper integral also diverges.

Note:
Remember to check for singularity point at interior points as well as endpoints when determining whether an integral is improper. For instance, if you do not recognized that the integral in the above example $\int_{-1}^{2} \frac{1}{x^3} dx$ was improper, you would have obtained the **incorrect** results

$$\int_{-1}^{2} \frac{1}{x^3} dx = \left[-\frac{1}{2x^2} \right]_{-1}^{2} = -\frac{1}{8} + \frac{1}{2} = \frac{3}{8}. \quad (\textit{incorrect evaluation})$$

Example:
Consider the integral $\int_0^2 (x^2 - 3x + 1)/[x(x-1)^2] dx$. Here, the integrand has two singularities, namely $x = 0$ and $x = 1$, inside $[0, 2]$. We can therefore write

$$\begin{aligned}\int_0^2 \frac{x^2 - 3x + 1}{x(x-1)^2} dx &= \lim_{t \to 0^+} \int_t^{1/2} \frac{x^2 - 3x + 1}{x(x-1)^2} dx \\ &\quad + \lim_{u \to 1^-} \int_{1/2}^{u} \frac{x^2 - 3x + 1}{x(x-1)^2} dx \\ &\quad + \lim_{v \to 1^+} \int_{v}^{2} \frac{x^2 - 3x + 1}{x(x-1)^2} dx.\end{aligned}$$

We note that

$$\frac{x^2 - 3x + 1}{x(x-1)^2} = \frac{1}{x} - \frac{1}{(x-1)^2}.$$

Hence,

$$\begin{aligned}\int_0^2 \frac{x^2-3x+1}{x(x-1)^2}dx &= \lim_{t\to 0^+}\left[\log x + \frac{1}{x-1}\right]_t^{1/2} \\ &+ \lim_{u\to 1^-}\left[\log x + \frac{1}{x-1}\right]_{1/2}^{u} \\ &+ \lim_{v\to 1^+}\left[\log x + \frac{1}{x-1}\right]_v^{2}.\end{aligned}$$

None of the above limits exists as a finite number. The integral is therefore divergent.

7.6 Convergence of a Sequence of Riemann Integrals

In this section we confine our attention to the limiting behavior of the integrals of a sequence of functions $\{f_n(x)\}_{n=1}^{\infty}$.

Theorem:

Suppose that $f_n(x)$ is Riemann integrable on $[a, b]$ for $n \geq 1$. If $f_n(x)$ converges uniformly to $f(x)$ on $[a, b]$ as $n \to \infty$, then $f(x)$ is Riemann integrable on $[a, b]$ and

$$\lim_{n\to\infty}\int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

Proof:

Let us now show the part of

$$\lim_{n\to\infty}\int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

Let $\epsilon > 0$ be given. Since $f_n(x)$ converges uniformly to $f(x)$, then there exists an integer n_0 that depends only on ϵ such that

$$|f_n(x) - f(x)| < \frac{\epsilon}{b-a}.$$

From the basic properties of Riemann integral, we have

$$\left|\int_a^b [f_n(x) - f(x)]dx\right| = \left|\int_a^b f_n(x)dx - \int_a^b f(x)dx\right|$$
$$\leq \int_a^b |f_n(x) - f(x)|\, dx$$
$$\leq \int_a^b \frac{\epsilon}{b-a} = \epsilon,$$

and the result follows, since ϵ is an arbitrary positive number.

7.7 Riemann-Stieltjes Integral

The Riemann-Stieltjes integral involves two functions $f(x)$ and $g(x)$, both defined on the interval $[a, b]$, and is denoted by $\int_a^b f(x)dg(x)$. In particular, if $g(x) = x$ we obtain the Riemann integral $int_a^b f(x)dx$. Thus the Riemann integral is the special case of the Riemann-Stieltjes integral

Definition (The Riemann-Stieltjes integral):
If $f(x)$ is bounded on $[a, b]$, if $g(x)$ is monotone increasing on $[a, b]$, and if $P = \{x_0, x_1, ..., x_n\}$ is a partition of $[a, b]$, we define the sums

$$LS_P(f, g) = \sum_{i=1}^{n} m_i \triangle g_i,$$
$$US_P(f, g) = \sum_{i=1}^{n} M_i \triangle g_i$$

where m_i and M_i are, respectively, the infimum and supremum of $f(x)$ on $[x_{i-1}, x_i]$, $\triangle g_i = g(x_i) - g(x_{i-1}),\ i = 1, 2, ..., n$. If for a given $\epsilon > 0$ there exists a $\delta > 0$ such that

$$US_P(f, g) - LS_P(f, g) < \epsilon$$

whenever $\triangle_P < \delta$, where $\triangle_P$ is the norm of P, then $f(x)$ is said to be Riemann-Stieltjes integrable with respect to $g(x)$ on $[a, b]$. In this case,

$$\int_a^b f(x)dg(x) = \inf_P US_P(f, g) = \sup_P LS_P(f, g).$$

Equivalently, suppose that for a given partition $P = \{x_0, x_1, \cdots, x_n\}$ we define the sum

$$S(P, f, g) = \sum_{i=1}^{n} f(t_i) \triangle g_i,$$

where t_i is a point in the interval $[x_{i-1}, x_i]$, $i = 1, 2, .., n$. Then $f(x)$ is Riemann-Stieltjes integrable with respect to $g(x)$ on $[a, b]$ if for any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\left| S(P, f, g) - \int_a^b f(x) dg(x) \right| < \epsilon$$

for any partition P on $[a, b]$ with a norm $\triangle_p < \delta$, and for any choice of the point t_i in $[x_{i-1}, x_i]$, $i = 1, 2, .., n$.

The next theorem shows that under certain conditions, the Riemann-Stieltjes integral reduces to the Riemann integral.

Theorem:

Suppose that $f(x)$ is Riemann-Stieltjes integrable with respect to $g(x)$ on $[a, b]$, where $g(x)$ has continuous derivative $g'(x)$ on $[a, b]$. Then

$$\int_a^b f(x) dg(x) = \int_a^b f(x) g'(x) dx.$$

Example:

Under this theorem, it is easy to see that if $f(x) = 1$ and $g(x) = x^2$,[30] then $\int_a^b f(x) dg(x) = \int_a^b f(x) g'(x) dx = \int_a^b 2x dx = b^2 - a^2$.

It is possible, however, for the Riemann-Stieltjes integral to exist even if $g(x)$ is a discontinuous function.

Theorem:

Let $g(x)$ be a step function defined on $[a, b]$ with jump discontinuities at $x =$

[30] It is noted here to check whether $g(x)$ is monotone increasing on $[a, b]$.

$c_1, c_2, ..., c_n$ and $a, c_1 < c_2 < ... < c_n = b$, such that

$$g(x) = \begin{cases} \lambda_1, & a \leq x < c_1 \\ \lambda_2, & c_1 \leq x < c_2 \\ . & . \\ . & . \\ , & . \\ \lambda_n, & c_{n-1} \leq x < c_n \\ \lambda_{n+1}, & x = c_n \end{cases}$$

If $f(x)$ is bounded on $[a, b]$ and continuous at $x = c_1, c_2, ..., c_n$, then

$$\int_a^b f(x)dg(x) = \sum_{i=1}^{n} (\lambda_{i+1} - \lambda_i) f(c_i).$$

Example:

The mathematical expectation of a random variable X is defined by

$$E(X) = \int x dF(x),$$

where $F(x)$ is the *c.d.f* pf X and the integral is the Riemann-Stieltjes integrals. When X has a probability density function $f(x)$, we see that

$$E(x) = \int x dF(x) = \int x f(x) dx$$

reduce to a Riemann integral. For a discrete probability distribution with $F(x)$ as a step function,

$$E(X) = \int x dF(x) = \sum p_r x_r,$$

where x_r is a discontinuity point of F and p_r is the saltus at x_r. It is interesting to note that in all the three cases the same symbol $\int x dF(x)$ could be employed.

8 Multivariate Calculus

In this section we extend the notions of limits, continuity, differentiation, and integration to multivariate functions, that us functions of several variables. These functions can be real-valued or possibly vector- valued. More specifically, if $\mathbb{R}^n$ denotes the n-dimensional Euclidean space, $n \geq 1$, then we shall in general consider functions defined on a set $D \subset \mathbb{R}^n$ and have values in $\mathbb{R}^m$, $m \geq 1$. Such functions are represented symbolically as $\mathbf{f} : D \to \mathbb{R}^m$, where for $\mathbf{x} = (x_1, x_2, ..., x_n)' \in D$,

$$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_m(\mathbf{x})]'$$

and $f_i(\mathbf{x})$ is a real-valued function of $x_1, x_2, ..., x_n$ $(i = 1, 2, ..., m)$.

8.1 Some Basic Definition

In this section we extend these concept of one-dimensional Euclidean space to higher dimensional Euclidean spaces.

Definition (Euclidean Norm):
Any point $\mathbf{x}$ in $\mathbb{R}^n$ can be represented as a column vector of the form $\mathbf{x} = (x_1, x_2, ..., x_n)'$, where x_i is the ith element of $\mathbf{x}$ $(i = 1, 2, ..., n)$. The Euclidean norm of $\mathbf{x}$ was defined as $\|\mathbf{x}\| = (\sum_{i=1}^n x_i^2)^{1/2}$.

Definition (Neighborhood of a point $\mathbf{x}_0$ in $\mathbb{R}^n$):
Let $\mathbf{x}_0 \in \mathbb{R}^n$. A neighborhood $N_r(\mathbf{x}_0)$ of $\mathbf{x}_0$ is a set of points in $\mathbb{R}^n$ that lie within some distance, say r, from $\mathbf{x}_0$, that is

$$N_r(\mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x} - \mathbf{x}_0\| < r\}.$$

Thus a neighborhood of $\mathbf{x}_0$ is inside a circle centered at $\mathbf{x}_0$.[31] If $\mathbf{x}_0$ is deleted from $N_r(\mathbf{x}_0)$, we obtain the so-called deleted neighborhood of $\mathbf{x}_0$, which we denote by $N_r^d(\mathbf{x}_0)$.

[31] See the picture on page 250 of Fulks.

Definition (Limit point of a set):
A point $\mathbf{x}_0$ in $\mathbb{R}^n$ is a limit point of a set $A \subset \mathbb{R}^n$ if every neighborhood of $\mathbf{x}_0$ contains an element $\mathbf{x}$ of A such that $\mathbf{x} \neq \mathbf{x}_0$, that is, every deleted neighborhood of $\mathbf{x}_0$ contains points of A.

Definition (Closed Set):
A set $A \subset \mathbb{R}^n$ is closed if every limit point of A belongs to A.

Definition (Interior):
A point $\mathbf{x}_0$ is $\mathbb{R}^n$ is an interior of a set $A \subset \mathbb{R}^n$ if there exists an $r > 0$ such that $N_r(\mathbf{x}_0) \subset A$.

Definition (Open Set):
A set $A \subset \mathbb{R}^n$ is open if for every point $\mathbf{x}$ in A there exists a neighborhood $N_r(\mathbf{x})$ that is contained in A. Thus, A is open if it contains entirely of interior points.

Definition (Boundary Point):
A point $p \in \mathbb{R}^n$ is a boundary point of a set $A \subset \mathbb{R}^n$ if every neighborhood of p contains points of A as well as points of A^c, the complement of A with respect to $\mathbb{R}^n$. The set of all boundary points of A is called its boundary and is denoted by $Br(A)$.

Definition (Bounded):
A set $A \subset \mathbb{R}^n$ is bounded if there exists an $r > 0$ such that $\|\mathbf{x}\| \leq r\ \forall \mathbf{x} \in A$.

Definition:
Let $\{\mathbf{a}_i\}_{i=1}^{\infty}$ represent a sequence of point in $\mathbb{R}^n$. Then $\{\mathbf{a}_i\}_{i=1}^{\infty}$ converges to a point $\mathbf{c} \in \mathbb{R}^n$ if for a given $\epsilon > 0$ there exists an integer N such that $\|\mathbf{a}_i - \mathbf{c}\| < \epsilon$ whenever $i > N$. This is written as $\lim_{i\to\infty} \mathbf{a}_i = \mathbf{c}$, or $\mathbf{a}_i \to \mathbf{c}$ as $i \to \infty$.

Definition:
A sequence $\{\mathbf{a}_i\}_{i=1}^{\infty}$ is bounded if there exist a number $K > 0$ such that $\|\mathbf{a}_i\| \leq K$ $\forall i$.

8.2 Limits of a Multivariate Function

For a function of several variables, say $x_1, x_2, ..., x_n$, its limit at a point $\mathbf{a} = (a_1, a_2, ..., a_n)'$ is considered when $\mathbf{x} = (x_1, x_2, ..., x_n)'$ approach $\mathbf{a}$ in any possible way. Thus when $n > 1$ there are infinitely many ways in which $\mathbf{x}$ can approach $\mathbf{a}$.

Definition:
Let $\mathbf{f}: D \to \mathbb{R}^n$. The $\mathbf{f}(\mathbf{x})$ is said to have a limit $\mathbf{L} = (L_1, L_2, ..., L_m)'$ as $\mathbf{x}$ approach $\mathbf{a}$, written symbolically as $\mathbf{x} \to \mathbf{a}$, where $\mathbf{a}$ is a limit point of D, if for a given $\epsilon > 0$, there exists a $\delta > 0$ such that $\|\mathbf{f}(\mathbf{x}) - \mathbf{L}\| < \epsilon$ for all $\mathbf{x}$ in $D \cap N_\delta^d(\mathbf{a})$, where $N_\delta^d(\mathbf{a})$ is a deleted neighborhood of $\mathbf{a}$ of radius δ. If it exist, this limit is written symbolically as $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$.

Note that whether a limit $\mathbf{f}(\mathbf{x})$ exists, its value must be the same no matter how $\mathbf{x}$ approach $\mathbf{a}$. It is important here to understand the meaning of "$\mathbf{x}$ approach $\mathbf{a}$". By this we do not necessarily mean that $\mathbf{x}$ moves along a straight line leading to $\mathbf{a}$. Rather we mean that $\mathbf{x}$ moves closer and closer to $\mathbf{a}$ along any curve that goes through $\mathbf{a}$.

Example:
Let $f(\mathbf{x}) = x^2y^2/(x^2 + y^2)$. The domain of this function is the whole plane except the origin. We take $\mathbf{a} = (0, 0)'$ and investigate the limit of $f(\mathbf{x})$ as $\mathbf{x} \to \mathbf{a}$.

Let $\epsilon > 0$ be given. Then we want to show that there is a δ such that

$$|f(\mathbf{x}) - 0| < \epsilon \quad \textit{whenever} \quad \|\mathbf{x} - \mathbf{a}\| = \sqrt{(x-0)^2 + (y-0)^2} < \delta. \tag{13}$$

Now clearly

$$\begin{aligned} x^2 &\leq x^2 + y^2 \;\; \textit{and} \\ y^2 &\leq x^2 + y^2, \end{aligned}$$

so that

$$|f(\mathbf{x})| = \frac{x^2y^2}{x^2+y^2} \leq \frac{(x^2+y^2)^2}{x^2+y^2} = (x^2+y^2) < \varepsilon$$

if

$$\|\mathbf{x}-\mathbf{a}\| = \sqrt{(x-0)^2+(y-0)^2} < \sqrt{\epsilon}.$$

Thus we have satisfied (13) with $\delta = \sqrt{\epsilon}$, and we see that

$$\lim_{\mathbf{x}\to\mathbf{0}} f(\mathbf{x}) = \mathbf{0}.$$

Example:

Consider the behavior of function

$$f(x_1, x_2) = \frac{x_1x_2}{x_1^2+x_2^2}$$

as $\mathbf{x} = (x_1, x_2)' \to \mathbf{0}$, where $\mathbf{0} = (0,0)'$. This function is defined everywhere in $\mathbb{R}^2$ except at $\mathbf{0}$. It is convenient here to represent the point $\mathbf{x}$ using polar coordinates, r and θ, such that $x_1 = r\cos\theta$ and $x_2 = r\sin\theta$, $r > 0,\ 0 \leq \theta \leq 2\pi$. We then have

$$f(x_1, x_2) = \cos\theta\sin\theta,$$

which depends on θ, but not on r. Since θ can have infinitely many values, $f(x_1, x_2)$ cannot be made close to any one constant L no matter how small r is. Thus the limit of this function does not exist as $\mathbf{x} \to \mathbf{0}$.

8.3 Continuity of a Multivariate Function

The notation of continuity for a function of several variables is much the same as that for a function of a single variable.

Definition:

Let $\mathbf{f} : D \to \mathbb{R}^m$, where $D \subset \mathbb{R}^n$, and let $\mathbf{a} \in D$. Then $\mathbf{f}(\mathbf{x})$ is continuous at $\mathbf{a}$ if

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}),$$

where $\mathbf{x}$ remains in D as it approaches $\mathbf{a}$. This is equivalent to stating that for a given $\epsilon > 0$ there exist a $\delta > 0$ such that

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})\| < \epsilon$$

for all $\mathbf{x} \in D \cap N_\delta(\mathbf{a})$. If $\mathbf{f}(\mathbf{x})$ is continuous at every point $\mathbf{x}$ in D, then it is said to be continuous in D. In particular, if $\mathbf{f}(\mathbf{x})$ is continuous in D and if δ depends only on ϵ, then $\mathbf{f}(\mathbf{x})$ is said to be uniformly continuous in D.

Lemma:

Suppose that $f, g : D \mapsto \mathbb{R}$ are real-valued continuous functions, where $D \subset \mathbb{R}^n$. Then we have the following:

(a). $f + g$, $f - g$ and fg are continuous in D.

(b). $|f|$ is continuous in D.

(c). $1/f$ is continuous in D provided that $f(\mathbf{x}) \neq 0 \ \forall \mathbf{x} \in D$.

Lemma:

Suppose that $\mathbf{f} : D \to \mathbb{R}^m$ is continuous, where $D \subset \mathbb{R}^n$, and that $\mathbf{g} : G \to \mathbb{R}^v$ is also continuous, where $G \subset \mathbb{R}^m$ is the image of D under $\mathbf{f}$. Then the composite function $\mathbf{g} \circ \mathbf{f}(\mathbf{x}) : D \to \mathbb{R}^v$, defined as $\mathbf{g} \circ \mathbf{f}(\mathbf{x}) = \mathbf{g}[\mathbf{f}(\mathbf{x})]$, is also continuous in D.

Theorem:

Suppose that $f : D \mapsto \mathbb{R}$ be a real-valued continuous function defined on a closed and bounded set $D \subset \mathbb{R}^n$. Then there exist points $\mathbf{p}$ and $\mathbf{q}$ in D for which

$$f(\mathbf{p}) = \sup_{\mathbf{x} \in D} f(\mathbf{x}),$$
$$f(\mathbf{q}) = \inf_{\mathbf{x} \in D} f(\mathbf{x}).$$

Thus, $f(\mathbf{x})$ attains each of its infimum and supremum at least once in D.

Theorem:

Suppose that D is a closed and bounded set (or called compact set) in $\mathbb{R}^n$. If $\mathbf{f} : D \to \mathbb{R}^m$ is continuous, then it is uniformly continuous in D.

8.4 Derivatives of A Multivariate Function

In this section we generalize the concept of differentiation given in section 4 to a multivariate function $\mathbf{f} : D \mapsto \mathbb{R}^m$, where $D \subset \mathbb{R}^n$.

8.4.1 Partial Derivatives

Definition (Partial Derivative):
Let $\mathbf{a} = (a_1, a_2, ..., a_n)'$ be an interior point of D. Suppose that the limit

$$\lim_{h_i \to 0} \frac{\mathbf{f}(a_1, a_2, ..., a_i + h_i, ..., a_n) - \mathbf{f}(a_1, a_2, ..., a_i, ..., a_n)}{h_i}$$

exists; then $\mathbf{f}$ is said to have partial derivative with respect to x_i at $\mathbf{a}$. This derivative is denoted by

$$\frac{\partial \mathbf{f}(\mathbf{a})}{\partial x_i}, \quad or \quad \mathbf{f}_{x_i}(\mathbf{a}),\ i = 1, 2, ..., n.$$

Hence, partial differentiation with respect to x_i is done in the usual way while treating all the remaining variables as constant.

Higher-order partial derivative of $\mathbf{f}$ are defined similarly. For example, the second-order partial derivative of $\mathbf{f}$ with respect to x_i at $\mathbf{a}$ is defined as

$$\lim_{h_i \to 0} \frac{\mathbf{f}_{x_i}(a_1, a_2, ..., a_i + h_i, ..., a_n) - \mathbf{f}_{x_i}(a_1, a_2, ..., a_i, ..., a_n)}{h_i}$$

and is denoted by

$$\frac{\partial^2 \mathbf{f}(\mathbf{a})}{\partial x_i^2}, \quad or \quad \mathbf{f}_{x_i x_i}(\mathbf{a}),\ i = 1, 2, ..., n.$$

Also, the second-order partial derivative of $\mathbf{f}$ with respect to x_i and x_i, $i \neq j$ at $\mathbf{a}$ is defined as

$$\lim_{h_j \to 0} \frac{\mathbf{f}_{x_i}(a_1, a_2, ..., a_j + h_j, ..., a_n) - \mathbf{f}_{x_i}(a_1, a_2, ..., a_j, ..., a_n)}{h_j}$$

and is denoted by

$$\frac{\partial^2 \mathbf{f}(\mathbf{a})}{\partial x_i \partial x_j}, \quad or \quad \mathbf{f}_{x_i x_j}(\mathbf{a}),\ \forall i \neq j.$$

Example:
Suppose that there is a function f that satisfies the equation

$$f(x,y) = x^3 + 5xy - y^2.$$

The first partial derivatives of this function are

$$f_x = 3x^2 + 5y \quad and \quad f_y = 5x - 2y.$$

Therefore, upon further differentiation, we get

$$f_{xx} = 6x, \quad f_{yx} = 5, \quad f_{xy} = 5, \quad f_{yy} = -2.$$

Definition (Jacobian Matrix):
In general if f_j is the jth element of $\mathbf{f}$ $(j = 1, 2, .., m)$, then the term $\frac{\partial f_j(\mathbf{x})}{\partial x_i}$ for $i = 1, 2, ..., n$; $j = 1, 2, ..., m$, constitute an $m \times n$ matrix called the Jacobian matrix of $\mathbf{f}$ at $\mathbf{x}$ and is denoted by $\mathbf{J_f}(\mathbf{x})$. If $m = n$, the determinant of $\mathbf{J_f}(\mathbf{x})$ is called the Jacobian determinant.; it is some times represented as

$$\det[\mathbf{J_f}(\mathbf{x})] = \frac{\partial(f_1, f_2, ..., f_n)}{\partial(x_1, x_2, ..., x_n)}.$$

Theorem:
Let f, f_x, f_y, and f_{xy} exist and be continuous on a neighborhood of a point $\mathbf{a} = (x_0, y_0)$. Then $f_{yx}(\mathbf{a})$ exists and $f_{xy}(\mathbf{a}) = f_{yx}(\mathbf{a})$.

Proof:
Let $\phi(x) = f(x, y_0+k) - f(x, y_0)$, where k and y are held fixed. Then for x sufficiently near x_0 and k small, ϕ is a function of the single variable x near x_0. To this function we apply the mean value theorem for function of one variable between x_0 and x_0+h:

$$\phi(x_0 + h) - \phi(x_0) = h\phi'(x_0 + \theta_1 h),$$

where the prime $(')$ denote differentiation with respect to x and where $0 < \theta_1 < 1$ and $\phi'(x_0 + \theta_1 h) = f_x(x_0 + \theta_1 h, y_0 + k) - f_x(x_0 + \theta_1 h, y_0)$. Thus (for fixed y and k)

$$\phi(x_0 + h) - \phi(x_0) = h[f_x(x_0 + \theta_1 h, y_0 + k) - f_x(x_0 + \theta_1 h, y_0)].$$

Now for each (fixed) h we apply the mean value theorem for functions of one variable $f_x(x_0 + \theta_1 h, y_0 + k) - f_x(x_0 + \theta_1 h, y_0)$ between y_0 and $y_0 + k$ to obtain

$$f_x(x_0 + \theta_1 h, y_0 + k) - f_x(x_0 + \theta_1 h, y_0) = k \cdot f_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 k),$$

where $\theta_2 < 1$. Hence

$$\begin{aligned}\phi(x_0 + h) - \phi(x_0) &= h[f_x(x_0 + \theta_1 h, y_0 + k) - f_x(x_0 + \theta_1 h, y_0)] \\ &= hk[f_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 k)].\end{aligned}$$

Recalling the meaning of ϕ, we can rewrite this as

$$\begin{aligned}&[f(x_0 + h, y_0 + k) - f(x_0 + h, y_0)] - [f(x_0, y_0 + k) - f(x_0, y_0)] \\ &\qquad = hk[f_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 k)]. \qquad (14)\end{aligned}$$

Dividing (14) by k and letting $k \to 0$ we get

$$\frac{f(x_0 + h, y_0 + k) - f(x_0 + h, y_0)}{k} - \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k} = h[f_{xy}(x_0 + \theta_1 h, y_0)],$$

that is

$$f_y(x_0 + h, y_0) - f_y(x_0, y_0) = h[f_{xy}(x_0 + \theta_1 h, y_0)]. \qquad (15)$$

Dividing (15) by h and letting $h \to 0$ we get

$$\frac{f_y(x_0 + h, y_0) - f_y(x_0, y_0)}{h} = f_{xy}(x_0, y_0),$$

i.e.

$$f_{yx}(x_0, y_0) = f_{xy}(x_0, y_0),$$

the results desired.

8.4.2 The Total Derivatives

Let $f(\mathbf{x})$ be a real valued function defined on a set $D \subset \mathbb{R}^n$. Suppose that $x_1, x_2, ..., x_n$ are functions of a single variable t. Then f is a function of t. The ordinary derivative of $4f$ with respect to t, namely, df/dt, is called the total derivative of t.

Lemma:

Assume that for the value of t under consideration dx_i/dt exist for $i = 1, 2, ..., n$ and that $\partial f(\mathbf{x})/\partial x_i$ exists and is continuous in the interior of D for $i = 1, 2, ..., n$. Under theses considerations, the total derivatives of f is given by

$$\frac{df}{dt} = \sum_{i=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_i}\frac{dx_i}{dt}.$$

Proof:

Let $n = 2$ and $\triangle x_1$, $\triangle x_2$ be increment of x_1, x_2 that correspond to an increment of $\triangle t$ of t. In turn, f will have the increment $\triangle f$. We then have

$$\begin{aligned}\triangle f &= f(x_1 + \triangle x_1, x_2 + \triangle x_2) - f(x_1, x_2)\\ &= [f(x_1 + \triangle x_1, x_2 + \triangle x_2) - f(x_1, x_2 + \triangle x_2)] + [f(x_1, x_2 + \triangle x_2) - f(x_1, x_2)]\end{aligned}$$

By mean-value theorem,

$$\begin{aligned}&(x_1 + \triangle x_1, x_2 + \triangle x_2) - f(x_1, x_2 + \triangle x_2)\\ &= (x_1 + \triangle x_1 - x_1)f'(x_1 + \theta \cdot \triangle x_1, x_2 + \triangle x_2)\\ &= \triangle x_1 \frac{\partial f(x_1 + \theta \cdot \triangle x_1, x_2 + \triangle x_2)}{\partial x_1}, \quad \theta < 1.\end{aligned}$$

we then have

$$\begin{aligned}\triangle f &= \triangle x_1 \frac{\partial f(x_1 + \theta_1 \triangle x_1, x_2 + \triangle x_2)}{\partial x_1}\\ &\quad + \triangle x_2 \frac{\partial f(x_1, x_2 + \theta_2 \triangle x_2)}{\partial x_2}.\end{aligned}$$

Hence

$$\begin{aligned}\frac{\triangle f}{\triangle t} &= \frac{\triangle x_1}{\triangle t}\frac{\partial f(x_1 + \theta_1 \triangle x_1, x_2 + \triangle x_2)}{\partial x_1}\\ &\quad + \frac{\triangle x_2}{\triangle t}\frac{\partial f(x_1, x_2 + \theta_2 \triangle x_2)}{\partial x_2}.\end{aligned}$$

As $\triangle t \to 0$, $\theta_i \to 0$, $\triangle x_i/\triangle t \to dx_i/dt$, and by the continuity of $\partial f(\mathbf{x})/\partial x_i$ we also have

$$\begin{aligned}\frac{\partial f(x_1 + \theta_1 \triangle x_1, x_2 + \triangle x_2)}{\partial x_1} &= \frac{\partial f(x_1, x_2)}{\partial x_1}\\ \frac{\partial f(x_1, x_2 + \theta_2 \triangle x_2)}{\partial x_2} &= \frac{\partial f(x_1, x_2)}{\partial x_2}.\end{aligned}$$

The desired results thus obtained.

In general, the expression

$$df = \sum_{i=1}^{n} \frac{\partial f(\mathbf{x})}{\partial x_i} dx_i$$

is called the total differentials of f at $\mathbf{x}$.

8.4.3 Directional Derivatives

Let $\mathbf{f} : D \mapsto \mathbb{R}^m$, where $D \subset \mathbb{R}^n$, and let $\mathbf{v}$ be a unit vector in $\mathbb{R}^n$, which represents a certain direction in the n-dimensional Euclidean space. By definition, the directional derivative of $\mathbf{f}$ at a point $\mathbf{x}$ is the interior of D in the direction of $\mathbf{v}$ is given by the limit

$$\lim_{h \to 0} \frac{\mathbf{f}(\mathbf{x} + h\mathbf{v}) - \mathbf{f}(\mathbf{x})}{h},$$

if it exists. In particular, if $\mathbf{v} = \mathbf{e}_i$, the unit vector in the direction of the ith coordinate axis, then the directional derivative of $\mathbf{f}$ in the direction of $\mathbf{v}$ is just the partial derivatives of $\mathbf{f}$ with respect to x_i $(i = 1, 2, .., n)$.

Lemma:
Let $\mathbf{f} : D \mapsto \mathbb{R}^m$, where $D \subset \mathbb{R}^n$. If the partial derivative $\partial f_j / \partial x_i$ exist at a point $\mathbf{x} = (x_1, x_2, ..., x_n)'$ in the interior of D for $i = 1, 2, ..., n;\ j = 1, 2, .., m$, then the directional derivative of $\mathbf{f}$ at $\mathbf{x}$ in the direction of a unit vector $\mathbf{v}$ exists and is equal to $\mathbf{J_f}(\mathbf{x})\mathbf{v}$, where $\mathbf{J_f}(\mathbf{x})$ is the $m \times n$ Jacobian of $\mathbf{f}$ at $\mathbf{x}$.

Proof:
Let $m = 1, n = 2$. The increment of f is

$$\begin{aligned} \triangle f &= f(x_1 + hv_1, x_2 + hv_2) - f(x_1, x_2) \\ &= [f(x_1 + hv_1, x_2 + hv_2) - f(x_1, x_2 + hv_2)] + [f(x_1, x_2 + hv_2) - f(x_1, x_2)] \end{aligned}$$

By mean-value theorem,

$$\begin{aligned}
& f(x_1+hv_1, x_2+hv_2) - f(x_1, x_2+hv_2) \\
& = (x_1+hv_1-x_1)f'(x_1+\theta\cdot hv_1, x_2+hv_2) \\
& = hv_1 \frac{\partial f(x_1+\theta\cdot hv_1, x_2+hv_2)}{\partial x_1}, \quad \theta<1.
\end{aligned}$$

we then have

$$\begin{aligned}
\triangle f &= hv_1 \frac{\partial f(x_1+\theta_1 hv_1, x_2+hv_2)}{\partial x_1} \\
&\quad + hv_2 \frac{\partial f(x_1, x_2+\theta_2 hv_2)}{\partial x_2}.
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{\triangle f}{h} &= \frac{hv_1}{h}\frac{\partial f(x_1+\theta_1 hv_1, x_2+hv_2)}{\partial x_1} + \frac{hv_2}{h}\frac{\partial f(x_1, x_2+\theta_2 hv_2)}{\partial x_2} \\
&= v_1 \frac{\partial f(x_1+\theta_1 hv_1, x_2+hv_2)}{\partial x_1} + v_2 \frac{\partial f(x_1, x_2+\theta_2 hv_2)}{\partial x_2}.
\end{aligned}$$

As $h \to 0$, and by the continuity of $\partial f(\mathbf{x})/\partial x_i$ we also have

$$\begin{aligned}
\frac{\partial f(x_1+\theta_1 hv_1, x_2+hv_2)}{\partial x_1} &= \frac{\partial f(x_1, x_2)}{\partial x_1} \\
\frac{\partial f(x_1, x_2+\theta_2 hv_2)}{\partial x_2} &= \frac{\partial f(x_1, x_2)}{\partial x_2}.
\end{aligned}$$

The desired results thus obtained.

Example:

Let $\mathbf{f}: \mathbb{R}^3 \mapsto \mathbb{R}^2$ be defined as

$$\mathbf{f}(x_1, x_2, x_3) = \begin{bmatrix} x_1^2 + x_2^2 + x_3^2 \\ x_1^2 - x_1x_2 + x_3^2 \end{bmatrix}$$

The directional derivative of $\mathbf{f}$ at $\mathbf{a} = (1, 2, 1)'$ in the direction of $\mathbf{v} = (1/\sqrt{2}, -1/\sqrt{2}, 0)'$ is

$$\begin{aligned}
\mathbf{J_f}(\mathbf{x})\mathbf{v} &= \begin{bmatrix} 2x_1 & 2x_2 & 2x_3 \\ 2x_1 - x_2 & -x_1 & 2x_3 \end{bmatrix}_{(1,2,1)} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 4 & 2 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} = \begin{bmatrix} -\frac{2}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}
\end{aligned}$$

Definition (The Gradient of f):
Let $f : D \mapsto \mathbb{R}$, where $D \subset \mathbb{R}^n$. If the partial derivatives $\partial f/\partial x_i$ $i = 1, 2, ..., n$ exists at a point $\mathbf{x} = (x_1, x_2, ..., x_n)'$ in the interior of D, then the vector $(\partial f/\partial x_1, \partial f/\partial x_2, ..., \partial f/\partial x_n)'$ is called the gradient of f at $\mathbf{x}$ and is denoted by $\boldsymbol{\nabla} f(\mathbf{x})$.

Definition (The Hessian Matrix):
Let $f : D \mapsto \mathbb{R}$, where $D \subset \mathbb{R}^n$. Then $\boldsymbol{\nabla} f : D \mapsto \mathbb{R}^n$. The Jacobian matrix of $\boldsymbol{\nabla} f(\mathbf{x})$ is called the Hessian matrix of f and is denoted by $\mathbf{H}_f(\mathbf{x})$. Thus $\mathbf{H}_f(\mathbf{x}) = \mathbf{J}_{\boldsymbol{\nabla} f}(\mathbf{x})$, that is,

$$\mathbf{H}_f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & . \;\; . \;\; . & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} \\ . & . & . \;\; . \;\; . & . \\ . & . & . \;\; . \;\; . & . \\ . & . & . \;\; . \;\; . & . \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} & . \;\; . \;\; . & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix}.$$

The determinant of $\mathbf{H}_f(\mathbf{x})$ is called the Hessian determinant.

8.5 Taylor's Theorem for a Multivariate Function

Notation (Del Operator):
Let us first introduce the following notation: Let $\mathbf{x} = (x_1, x_2, ..., x_n)'$. Then $\mathbf{x}'\boldsymbol{\nabla}$ denotes a first-order differential operator of the form

$$\mathbf{x}'\boldsymbol{\nabla} = \sum_{i=1}^{n} x_i \frac{\partial}{\partial x_i}.$$

The symbol, $\boldsymbol{\nabla} = (\partial/\partial x_1, ..., \partial/\partial x_n)'$, is called the del operator. If m is a positive integer, then $(\mathbf{x}'\nabla)^m$ denote an mth order differential operator. For example $n = 2$,

$$\begin{aligned} (\mathbf{x}'\nabla)^2 &= \left(x_1 \frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_2} \right)^2 \\ &= x_1^2 \frac{\partial^2}{\partial x_1^2} + 2x_1 x_2 \frac{\partial^2}{\partial x_1 \partial x_2} + x_2^2 \frac{\partial^2}{\partial x_2^2}. \end{aligned}$$

Thus $(\mathbf{x}'\nabla)^2$ is obtained by squaring $x_1\frac{\partial}{\partial x_1} + x_2\frac{\partial}{\partial x_2}$ in the usual fashion, except that the squares of $\frac{\partial}{\partial x_1}$ and $\frac{\partial}{\partial x_2}$ are replaced by $\frac{\partial^2}{\partial x_1^2}$ and $\frac{\partial^2}{\partial x_2^2}$, and the product of $\frac{\partial}{\partial x_1}$ and $\frac{\partial}{\partial x_2}$ is replaced by $\frac{\partial^2}{\partial x_1 \partial x_2}$.

The notation $(\mathbf{x}'\nabla)^m f(\mathbf{x}_0)$ indicate that $(\mathbf{x}'\nabla)^m f(\mathbf{x})$ is evaluated at $\mathbf{x}_0$.

Theorem (Multivariate Taylor's Theorem):
Let $f : D \mapsto \mathbb{R}$, where $D \subset \mathbb{R}^n$, and let $N_\delta(\mathbf{x}_0)$ be a neighborhood of $\mathbf{x}_0 \in D$ such that $N_\delta(\mathbf{x}_0) \subset D$. If f and all its partial derivatives of order $\leq r$ exists and continuous in $N_\delta(\mathbf{x}_0)$. Then for $\mathbf{x}$ in $N_\delta(\mathbf{x}_0)$,

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{r-1} \frac{[(\mathbf{x}-\mathbf{x}_0)'\nabla]^i f(\mathbf{x}_0)}{i!} + \frac{[(\mathbf{x}-\mathbf{x}_0)'\nabla]^i f(\mathbf{z}_0)}{r!},$$

where $\mathbf{z}_0$ is a point on the line segment from $\mathbf{x}_0$ to $\mathbf{x}$.

Example:
In two dimensions, if we set $\mathbf{x}_0 = (a,b)'$ and $\mathbf{x} = (x,y)'$, we get

$$\begin{aligned} f(x,y) &= f(a,b) + \left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right] f(a,b) \\ &\quad + \frac{1}{2!}\left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right]^2 f(a,b) \\ &\quad + \frac{1}{3!}\left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right]^3 f(a,b) \\ &\quad + ... + \frac{1}{r!}\left[(x-a)\frac{\partial}{\partial x} + (y-b)\frac{\partial}{\partial y}\right]^r f(a_0,b_0). \end{aligned}$$

Expanding, we get

$$\begin{aligned} f(x,y) &= f(a,b) + (x-a)\frac{\partial f}{\partial x}(a,b) + (y-b)\frac{\partial f}{\partial y}(a,b) \\ &\quad + \frac{1}{2!}\left[(x-a)^2\frac{\partial^2 f(a,b)}{\partial x^2} + 2(x-a)(y-b)\frac{\partial^2 f(a,b)}{\partial x \partial y} + (y-b)^2\frac{\partial^2 f(a,b)}{\partial y^2}\right] \\ &\quad + \frac{1}{3!}\left[(x-a)^3\frac{\partial^3 f(a,b)}{\partial x^3} + 3(x-a)^2(y-b)\frac{\partial^3 f(a,b)}{\partial^2 x \partial y}\right. \\ &\quad \left. + 3(x-a)(y-b)^2\frac{\partial^3 f(a,b)}{\partial x \partial^2 y} + (y-b)^3\frac{\partial^3 f(a,b)}{\partial y^3}\right] \\ &\quad + \frac{1}{4!}[\cdots] + \cdots . \end{aligned}$$

In the case of one dimensional, it reduce to

$$\begin{aligned} f(x) &= f(a) + (x-a)\frac{\partial f}{\partial x}(a) \\ &+\frac{1}{2!}\left[(x-a)^2\frac{\partial^2 f(a)}{\partial x^2}\right] \\ &+\frac{1}{3!}\left[(x-a)^3\frac{\partial^3 f(a)}{\partial x^3}\right] \\ &+\frac{1}{4!}[\cdots]+\cdots. \end{aligned}$$

8.6 Optimum of A Multivariate Function

In general, any point at which $\partial f/\partial x_i = 0$ for $i = 1, 2, .., n$ is called a stationary point. The following theorem gives the conditions needed to have a local optimum at s stationary point.

Theorem:
Let $f: D \to \mathbb{R}$, where $D \subset \mathbb{R}^n$. Suppose that f has continuous second-order partial derivatives in D. If $\mathbf{x}_0$ is a stationary point of f, then at $\mathbf{x}_0$ has the following:
(a). A local minimum if $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) > 0$ for all $\mathbf{h} = (h_1, h_2, ..., h_n)'$ in a neighborhood of $\mathbf{0}$, where the element of $\mathbf{h}$ are not all equal to zero.
(b). A local maximum if $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) < 0$ for all $\mathbf{h} = (h_1, h_2, ..., h_n)'$ in a neighborhood of $\mathbf{0}$, where the element of $\mathbf{h}$ are not all equal to zero.
(c). A saddle point if $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$ changes sign for value of $\mathbf{h} = (h_1, h_2, ..., h_n)'$ in a neighborhood of $\mathbf{0}$.

Proof:
By applying Taylor's theorem to $f(\mathbf{x}_0)$ in a neighborhood of $\mathbf{x}_0$ we obtain

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + (\mathbf{h}'\nabla)f(\mathbf{x}_0) + \frac{1}{2!}(\mathbf{h}'\nabla)^2 f(\mathbf{z}_0),$$

where $\mathbf{h}$ is a nonzero vector in a neighborhood of $\mathbf{0}$ and $\mathbf{z}_0$ is a point on the line segment from $\mathbf{x}_0$ to $\mathbf{x}_0 + \mathbf{h}$. Since $\mathbf{x}_0$ is a stationary point, then $(\mathbf{h}'\nabla)f(\mathbf{x}_0) = 0$. Hence

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = \frac{1}{2!}(\mathbf{h}'\nabla)^2 f(\mathbf{z}_0).$$

Also, since the second-order partial derivative of f is continuous at $\mathbf{x}_0$, then we can write

$$f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)=\frac{1}{2!}(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)+o(\|\mathbf{h}\|),$$

where $\|\mathbf{h}\|=(\mathbf{h}'\mathbf{h})^{1/2}$ and $o(\|\mathbf{h}\|)\to 0$ as $\mathbf{h}\to\mathbf{0}$. We note that for small values of $\|\mathbf{h}\|$, the sign of $f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)$ depends on the value of $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$. It follows that if

(a). $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)>0$, then $f(\mathbf{x}_0+\mathbf{h})>f(\mathbf{x}_0)$ for all nonzero value of $\mathbf{h}$ in some neighborhood of $\mathbf{0}$. Thus $\mathbf{x}_0$ is a local minimum of f.

(b). $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)<0$, then $f(\mathbf{x}_0+\mathbf{h})<f(\mathbf{x}_0)$ for all nonzero value of $\mathbf{h}$ in some neighborhood of $\mathbf{0}$. In this case, $\mathbf{x}_0$ is a local maximum of f.

(c). $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$ changes sign inside a neighborhood of $\mathbf{0}$, then $\mathbf{x}_0$ is neither a point of local maximum nor a point of local minimum. In this case, $\mathbf{x}_0$ is a saddle point.

Note:

We note that $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$ can be written as a quadratic form of the form $\mathbf{h}'\mathbf{H}_f(\mathbf{x}_0)\mathbf{h}$, where $\mathbf{H}_f(\mathbf{x}_0)$ is the $n\times n$ Hessian matrix evaluated at $\mathbf{x}_0$, that is,

$$\mathbf{H}_f(\mathbf{x}_0)=\begin{bmatrix} f_{11} & f_{12} & . & . & . & f_{1n}\\ f_{21} & f_{22} & . & . & . & f_{2n}\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ f_{n1} & f_{n2} & . & . & . & f_{nn}\end{bmatrix}_{\mathbf{x}_0}.$$

Example:

For the case of $n=2$,

$$\begin{aligned}(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) &= h_1^2\frac{\partial^2 f(\mathbf{x}_0)}{\partial x_1^2}+2h_1h_2\frac{\partial^2 f(\mathbf{x}_0)}{\partial x_1\partial x_2}+h_2^2\frac{\partial^2 f(\mathbf{x}_0)}{\partial x_2^2}\\ &= h_1^2 f_{11}+2h_1h_2 f_{12}+h_2^2 f_{22}\\ &= [h_1\ h_2]\begin{bmatrix} f_{11} & f_{12}\\ f_{21} & f_{22}\end{bmatrix}_{\mathbf{x}_0}\begin{bmatrix} h_1\\ h_2\end{bmatrix}\\ &= \mathbf{h}'\mathbf{H}_f(\mathbf{x}_0)\mathbf{h},\quad \forall\, \mathbf{h}\neq\mathbf{0}.\end{aligned}$$

Corollary:
Let f be the same function as in the previous theorem, and let $\mathbf{H}_f(\mathbf{x}_0)$ be the Hessian matrix. If $\mathbf{x}_0$ is a stationary point of f, then at $\mathbf{x}_0$ f has the following:
(a). A local minimum if $\mathbf{H}_f(\mathbf{x}_0)$ is positive definite.
(b). A local maximum if $\mathbf{H}_f(\mathbf{x}_0)$ is negative definite.
(c). A saddle point if $\mathbf{H}_f(\mathbf{x}_0)$ is neither positive definite nor negative definite.

8.7 The Method of Lagrange Multipliers

The method, which is due to Joseph Louis de Lagrange (1936-1813), is used to optimize a real-valued function $f(x_1, x_2, ..., x_n)$, where $x_1, x_2, ..., x_n$ are subject to $m(< n)$ equality constrains of the form

$$\begin{aligned} g_1(x_1, x_2, ..., x_n) &= 0 \quad , \\ g_2(x_1, x_2, ..., x_n) &= 0 \quad , \\ &\ \cdot \\ &\ \cdot \\ &\ \cdot \\ g_m(x_1, x_2, ..., x_n) &= 0, \end{aligned} \tag{16}$$

where $g_1, g_2, ..., g_m$ are differentiable functions.

The determination of stationary points in this constrained optimization problem is done by first considering the function

$$F(\mathbf{x}) = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j g_j(\mathbf{x}), \tag{17}$$

where $\lambda_1, \lambda_2, ..., \lambda_m$ are scalars called Lagrange multipliers. By differentiating () with respect to $x_1, x_2, ..., x_n$ and equating the partial derivatives to zero we obtain

$$\frac{\partial F}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{m} \frac{\partial g_j}{\partial x_i} = 0, \quad i = 1, 2, ..., n. \tag{18}$$

Equations (16) and (18) consists of $m + n$ equations in $m + n$ unknowns, namely, $x_1, x_2, ..., x_n; \lambda_1, \lambda_2, ..., \lambda_m$. The solutions for $x_1, x_2, ..., x_n$ determine the locations of stationary points.

8.8 The Rinmann Integral of a Multivariate Function

In this section we extend the concept of Riemann integration to real-valued function of n variables, $x_1, x_2, ..., x_n$.

Definition (Cell):
The set of points in $\mathbb{R}^n$ whose coordinates satisfy the inequalities

$$a_i \leq x_i \leq b_i, \quad i = 1, 2, ..., n,$$

where $a_i < b_i$, $i = 1, 2, .., n$, form an n-dimensional cell denoted by $c_n(a, b)$. The content of this cell is $\prod_{i=1}^{n}(b_i - a_i)$ and is denoted by $\mu[c_n(a, b)]$.

Definition (Sub-Cell):
Suppose that P_i is a partition of the interval $[a_i, b_i]$, $i = 1, 2, ..., n$. The Cartesian product $P = \times_{i=1}^{n} P_i$ is a partition of $c_n(a, b)$ and consists of n-dimensional subcells of $c_n(a, b)$. We denote these subcells by $S_1, S_2, ..., S_v$. The content of S_i is denoted by $\mu(S_i)$, $i = 1, 2, ..., v$, where v is the number of subcells.

8.8.1 The Riemann Integral on Cells

We first define the Riemann integral of a real-valued function $f(\mathbf{x})$ on an n-dimensional cell.

Definition (Lower and Upper Sum):
Let $f : D \mapsto \mathbb{R}$, where $D \subset \mathbb{R}^n$. Suppose that $c_n(a, b)$ is an n-dimensional cell contained in D and that f is bounded on $c_n(a, b)$. Let P be a partition of $c_n(a, b)$ consisting of the subcells $S_1, S_2, ..., S_v$. Let m_i and M_i be, respectively the infimum and supremum of f on S_i, $i = 1, 2, ..., v$. Consider the sums

$$LS_P(f) = \sum_{i=1}^{v} m_i \mu(S_i),$$

$$US_P(f) = \sum_{i=1}^{v} M_i \mu(S_i).$$

We refer to $LS_P(f)$ and $US_P(f)$ as the lower and upper sums, respectively, of f respect to the partition P.

Theorem:

Let $f : D \mapsto \mathbb{R}$, where $D \subset \mathbb{R}^n$. Suppose that f is bounded on $c_n(a,b) \subset D$. Then f is Riemann integrable on $c_n(a,b)$ if and only if for every $\epsilon > 0$ there exists a partition P of $c_n(a,b)$ such that

$$US_P(f) - LS_P(f) < \epsilon.$$

Definition:

Let $f : c_n(a,b) \mapsto \mathbb{R}$ be a bounded function. Then f is Riemann integrable on $c_n(a,b)$ if and only if

$$\sup_P LS_P(f) = \inf_P US_P(f).$$

Their common value is called the Riemann integral of f on $c_n(a,b)$ and is denoted by

$$\int_{c_n(a,b)} f(\mathbf{x})d\mathbf{x} \equiv \int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n} f(x_1, x_2, ..., x_n)dx_1dx_2\cdots dx_n.$$

Theorem:

If f is continuous on an n-dimensional cell $c_n(a,b)$, then it is Riemann integrable there.

8.8.2 Iterated Riemann Integrals on Cells

The definition of the n-tuple integral does not provide a practicable way to evaluate it. We now show that the evaluation of the integral can be obtained by performing n Riemann integrals each of which is carried out with respect to one variable. Let us first consider a double integral.

Lemma:
Suppose that f is real-valued and continuous on $c_2(a,b)$. Define the function $g(x_2)$ as

$$g(x_2) = \int_{a_1}^{b_1} f(x_1, x_2) dx_1.$$

Then $g(x_2)$ is continuous on $[a_2, b_2]$. Therefore $g(x_2)$ is Riemann integrable on $[a_2, b_2]$, that is, $\int_{a_2}^{b_2} g(x_2) dx_2$ exists. We call the integral

$$\int_{a_2}^{b_2} g(x_2) dx_2 = \int_{a_2}^{b_2} \left[\int_{a_1}^{b_1} f(x_1, x_2) dx_1 \right] dx_2$$

an iterated integral of order 2.

Theorem:
If f is continuous on $c_2(a,b)$, then

$$\int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2 = \int_{a_2}^{b_2} \left[\int_{a_1}^{b_1} f(x_1, x_2) dx_1 \right] dx_2.$$

That is, the double integral is equal to the iterated integral of order 2.

Theorem (Fubini's Theorem):
If f is continuous on $c_2(a,b)$, then

$$\begin{aligned} \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 dx_2 &= \int_{a_2}^{b_2} \left[\int_{a_1}^{b_1} f(x_1, x_2) dx_1 \right] dx_2 \\ &= \int_{a_1}^{b_1} \left[\int_{a_2}^{b_2} f(x_1, x_2) dx_2 \right] dx_1. \end{aligned}$$

8.8.3 Integration Over General Set

We now consider n-tuple Riemann integration over regions in $\mathbb{R}^n$ that are not necessarily cell shaped.

Let $f : D \mapsto \mathbb{R}$ be a bounded and continuous function, where D is a bounded region in $\mathbb{R}^n$. There exists an n-dimensional cell $c_n(a,b)$ such that $D \subset c_n(a,b)$. Let

$g : c_n(a, b) \mapsto \mathbb{R}$ be defined as

$$g(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \mathbf{x} \in D, \\ 0, & \mathbf{x} \notin D. \end{cases}$$

Then

$$\int_{c_n(a,b)} g(\mathbf{x}) d\mathbf{x} = \int_D f(\mathbf{x}) d\mathbf{x}.$$

In practice, it is not always necessary to make reference to $c_n(a, b)$ that encloses D in order to evaluate the integral on D. Rather, we only need to recognize that the limits of integration in the iterated Riemann integral depend in general on variables that have not yet been integrated out as the following examples.

Example:

Let $f(x_1, x_2) = x_1 x_2$ and D be the region

$$D = \{(x_1, x_2) | x_1^2 + x_2^2 \leq 1,\ x_1 \geq 0, x_2 \geq 0\}.$$

It is easy to see that D is contained inside the two-dimensional cell

$$c_2(0, 1) = \{(x_1, x_2) | 0 \leq x_1 \leq 1, 0 \leq x_2 \geq 1\}.$$

Then

$$\begin{aligned} \int\int_D x_1 x_2 dx_1 dx_2 &= \int_0^1 \left[\int_0^{(1-x_1^2)^{1/2}} x_1 x_2 dx_2 \right] dx_1 \\ &= \int_0^1 x_1 \left[\int_0^{(1-x_1^2)^{1/2}} x_2 dx_2 \right] dx_1 \\ &= \frac{1}{2} \int_0^1 x_1 (1 - x_1^2) dx_1 \\ &= \frac{1}{8}. \end{aligned}$$

8.9 Differentiation Under the Integral Sign