MAL100 (Calculus) Notes

Dr. Avijit Pal
IIT Bhilai

Contents

1 Set Theory, Relation and Mapping

In this section we study set theory, cartesian product product of sets, relation and mapping.
Definition 1 (Set)**.** A set is a well defined collection of distinct objects (which are also called *elements* or *points*).

Example 1.1. 1. $\mathbb{N} = \{1, 2, 3, \dots\}$, the set of all natural numbers.

2. $\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$, the set of all integers.

3. $\mathbb{Q} = \{\frac{p}{q} : p, q$ are integers $q \neq 0\}$, the set of all rational numbers.

4. $\mathbb{I}$ the set of all irrational numbers; $(\sqrt{2}, \sqrt{3}, e, \pi, etc.)$.

5. $\mathbb{R} = \mathbb{Q} \cup \mathbb{I}$, the set of all real numbers, where $A \cup B$ is the union of two sets.

Our main aim is to study $\mathbb{R}$ in details.

Recall Some Facts: Let A and B be two sets. If $x \in A$ implies $x \in B$ then we say that A is a subset of B. We denote it as $A \subseteq B$. For example, it is easy to verify that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

Definition 2. The set which contains no element is called *null set* or *empty set* or *void set* and it is denoted by ϕ. And $\phi \subset A$ for any set A.

Definition 3. A set S is said to be *finite set* if either it is empty or it contains a finite number of elements, otherwise it is said to be *infinite set.*

1.1 Algebraic Operations on Sets

We discuss union, intersection, complementation, difference, symmetric difference etc.

(a) **Union:** The union of two sets A and B is defined by

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

For example, if $A = \{1, 2, \dots, 10\}$ and $B = \{6, 7, \dots, 15\}$ then $A \cup B = \{1, 2, \dots, 15\}$.

(b) **Intersection:** The intersection of two sets A and B is defined by

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

A example is, if $A = \{1, 2, \dots, 10\}$ and $B = \{6, 7, \dots, 15\}$ then $A \cup B = \{6, 7, \dots, 10\}$.

(c) **Disjoint Sets:** Two subsets A and B are said to be *disjoint* if $A \cap B = \phi$. An example of such sets are $\mathbb{Q}$ and $\mathbb{I}$.

(d) **Complementation:** The complementation of a subset A of the universal set U, is denoted by A^c and defined by

$$A^c = \{x \in U : x \neq A\}.$$

For $U = \{1, 2, \dots, 10\}$ and $A = \{1, 3, 5, 7, 9\}$ we have $A^c = \{2, 4, 6, 8, 10\}$.

Excercise 1.1. Let U be the universal set and A is a subset of U. Prove that

1. $A \cup A^c = U$ and $A \cap A^c = \phi$.
2. $(A^c)^c = A$.

Theorem 1.1 (De Morgan's Law)**.** For any two subsets A, B of the universal set U the following hold:

(i) $(A \cup B)^c = A^c \cap B^c$,

(ii) $(A \cap B)^c = A^c \cup B^c$.

Proof. Let $x \in (A \cup B)^c$. This implies

$$\begin{aligned} x \notin A \cup B &\Rightarrow x \notin A \text{ and } x \notin B \\ &\Rightarrow x \in A^c \text{ and } x \in B^c \\ &\Rightarrow x \in A^c \cap B^c. \end{aligned}$$

Thus we have that $(A \cup B)^c \subseteq A^c \cap B^c$.

Conversely, let $x \in A^c \cap B^c$. Thus

$$\begin{aligned} x \notin A \text{ and } x \notin B &\Rightarrow x \notin A \text{ and } x \notin B \\ &\Rightarrow x \notin A \cup B \\ &\Rightarrow x \in (A \cup B)^c. \end{aligned}$$

Hence $A^c \cap B^c \subseteq (A \cup B)^c$. And therefore (i) follows.

Proof of (ii) is similar, so we left it as an exercise. ☐

(e) **Difference of Sets:** The difference of two subsets A, B of U is denoted by

$$A \setminus B = \{x \in A : x \notin B\}.$$

For example, let $A = \{1, 2, \ldots, 6\}$ and $B = \{2, 4, 6, 8, 10\}$ be subsets of $\mathbb{N}$. Then $A \setminus B = \{1, 3, 5\}$ and $B \setminus A = \{8, 10\}$.

Excercise 1.2. Prove that $A \setminus B = A \cap B^c$ and $A \setminus B = A \setminus (A \cap B)$.

(f) **Symmetric Difference:** The symmetric difference of two subsets A and B is a subset of U, denoted by $A \Delta B$ and defined by

$$A \Delta B = (A \setminus B) \cup (B \setminus A).$$

Let us see an example. Take $A = \{1, 2, 3, 4\}, B = \{2, 4, 6, 8\}$ as a subset of $\mathbb{N}$. Then

$$A \Delta B = (A \setminus B) \cup (B \setminus A) = \{1, 3, 6, 8\}.$$

(g) **Family of Sets:** We have defined a set as a collection of its elements. If the elements of a set be the subsets of a universal set, then we call it family of sets.

Let X be a non-empty set. The collection of all subsets of X is a family of sets. This family of sets is called the *power set of* X and is denoted by $\mathcal{P}(X)$. If $X = \{1, 2, 3\}$ then $\mathcal{P}(X) = \{\phi, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}$. If X contains n elements then the cardinality of $\mathcal{P}(X)$ is 2^n.

Let I be the finite set $\{1, 2, \ldots, n\}$ and $\mathcal{F} = \{A_1, \ldots, A_n\}$, the family of n sets. Then $\mathcal{F}$ is expressed as $\{A_i : i \in I\}$, and I is called the *index set*. The elements of $\mathcal{F}$ are said to be *indexed by* the index set I.

1.2 Cartesian Product of Sets

Let A and B be non-empty sets. The *Cartesian product* of A and B, denoted by $A \times B$ and defined by

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

$A \times B$ is the set of all ordered pairs (a, b), the first coordinate of the pair being an element of A and the second one being an element of B. The Cartesian product of the collection of sets $\{A_1, \ldots, A_n\}$ is denoted by $A_1 \times \cdots \times A_n$ and is defined by

$$A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n) : a_i \in A_i, 1 \leqslant i \leqslant n\}.$$

Example 1.2. Let $A = \{1, 2, 3\}$ and $B = \{2, 4\}$. Then

$$A \times B = \{(1, 2), (1, 4), (2, 2), (2, 4), (3, 2), (3, 4)\}.$$

Now find $B \times A$ (Exercise!).

1.2.1 Relation on a Set

Let A and B be two non-empty sets. A relation ρ between A and B is a subset of $A \times B$. If the ordered pair $(a, b) \in \rho$ then the element a of the set A is said to be related to the element b of B by the relation ρ. We denote "a is related to the element b" by $a\rho b$ or $(a, b) \in \rho$ or a b.

Example 1.3. Let $A = \{2, 3, 4, 5\}$ and $B = \{4, 6, 8, 9\}$. A relation ρ between A and B is defined by specifying that an element a of A is related to an element b of B if a is a divisor of b. Then

$$\rho = \{(2, 4), (2, 6), (2, 8), (3, 6), (3, 9), (2, 4), (4, 4), (4, 8)\}.$$

Note that $(2, 5) \in A \times B$ but $(2, 5) \notin \rho$.

Definition 4. Let $A \neq \phi$. A *binary relation* ρ on A is a subset of $A \times A$.

1.2.2 Ordered Relation on a Set

Let $X \neq \phi$. A relation ρ on X is said to be

(1) reflexive if $a\,\rho\,a$ holds for all $a \in X$.

(2) symmetric if $a\,\rho\,b \Rightarrow b\,\rho\,a$.

(3) anti symmetric if $a\,\rho\,b$ and $b\,\rho\,a \Rightarrow a = b$.

(4) transitive if $a\,\rho\,b$ and $b\,\rho\,c \Rightarrow a\,\rho\,c$.

Example 1.4. Let a relation ρ is defined on $\mathbb{Z}$ by $a\,\rho\,b \iff a - b$ is even. Then ρ is reflexive, symmetric and transitive but not anti symmetric.

Example 1.5. Let a relation ρ be defined on $\mathbb{Z}$ by $a\,\rho\,b \iff a$ is a divisor of b for $a, b \in \mathbb{Z}$. Then ρ is reflexive, anti symmetric and transitive but not symmetric.

1.2.3 Partial Order Relation

Definition 5. Let X be a non-empty set. A relation ρ on X is said to be *partial order relation* if ρ is reflexive, anti symmetric and transitive.

A relation of partial order is often denoted by $\leq$ even it is not "*less than*".

Definition 6 (Poset)**.** A non-empty set X together with a relation of partial order $\leq$ on X is called a *poset* (partially ordered set) and is denoted by $(X, \leq)$.

Example 1.6. $(\mathbb{R}, \leq)$ is a poset where $x \leq y$ means "x is less than or equal to y" for $x, y \in \mathbb{R}$.

Example 1.7. Let $X \neq \phi$ and $\mathcal{P}(X)$ be the power set of X. $(\mathcal{P}(X), \leq)$ is a poset where $A \leq B$ means "A is a subset of B" for $A, B \in \mathcal{P}(X)$.

1.3 Mapping

Let A and B be two non-empty sets. A *mapping* f from A to B is a rule that assigns to each element x of A to a definite element y in B.

A is said to be the domain of f and B is said to be the co-domain of f. A mapping is also called a *function* or a *transformation.*

Let $f : A \to B$ be a mapping and $x \in A$. Then the unique element y of B that corresponds to x by the mapping f is called the f-image of x and is denoted by $f(x)$. By $\mathrm{Im} f = \{f(x) : x \in A\} = R(f)$ we denote the image or range of f and by $D(f)$ we denote the domain of f.

Example 1.8 (Non example)**.** Let $S = \{1, 2, 3, 4\}$ and $T = \{a, b, c, d\}$. Define $f : S \to T$ by $f(1) = a = b, f(2) = c, f(3) = d$. Then f is not a function.

Example 1.9. Let $f : S \to T$ by $f(1) = a, f(2) = b, f(3) = c, f(4) = d$. Then f is a mapping.

1.3.1 Different Types of Mappings

Definition 7. A mapping $f : A \to B$ is said to be

(1) *injective* if for each pair of different elements of A, their f-images are distinct, i.e., if $x_1 \neq x_2$ in A then $f(x_1) \neq f(x_2)$.

(2) *surjective* if $f(A) = B$. In other words, if f is surjective then each element in B has at least one pre-image.

(3) *bijective* if f is both injective and surjective.

Example 1.10. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = 3x + 1$.

(1) f is injective because $f(x_1) = f(x_2)$ implies that $x_1 = x_2$.

(2) f is surjective because, for any $y \in \mathbb{R}$ with $f(x) = y$ implies that $x = \frac{y-1}{3}$. This shows that each elements in $\mathbb{R}$ has a pre-image.

(1) and (2) together imply that f is bijective.

Example 1.11. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x^2$. Then f is not injective as $f(2) = f(-2) = 4$ and not surjective as $-1 \in \mathbb{R}$ does not have any pre-image.

Definition 8. A mapping $f : A \to B$ is said to be *constant mapping* if f maps each elements if A to one and the same element in B.

Example 1.12. The map $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2$ for all $x \in \mathbb{R}$ is a constant map.

Definition 9. A mapping $f : A \to A$ is said to be *identity mapping* if $f(x) = x$ for all $x \in A$. The "*identity mapping on* A" is denoted by i_A.

1.3.2 Equality of Mapping

Two mappings $f : A \to B$ and $g : B \to C$ are said to be equal if $f(x) = g(x)$ for all $x \in A$.

Note 1.1. For the equality of two mappings f and g the following conditions must hold:

(1) f and g must have same domain D,

(2) for all $x \in D, f(x) = g(x)$.

Example 1.13. Let $S = \{x \in \mathbb{R} : x > 0\} = \mathbb{R}_{>0}$. Let $f : S \to \mathbb{R}$ defined by $f(x) = \frac{|x|}{x}, x \in S$ and $g : S \to \mathbb{R}$ defined by $g(x) = 1, x \in S$. Then it can be checked that $f = g$.

1.3.3 Restriction of a Mapping

Let $f : A \to B$ be a mapping and $\phi \neq D \subseteq A$. Then the mapping $g : D \to B$ defined by $g(x) = f(x)$ for all $x \in D$ is said to be the *restriction of* f *to* D and is denoted by $f|_D$.

Example 1.14. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \sin x$. If we reduce the co-domain to $T = \{x \in \mathbb{R} : -1 \leq x \leq 1\}$ and the domain $S = \{x \in \mathbb{R} : -\pi/2 \leq x \leq \pi/2\}$ then the mapping $g : S \to T$ defined by $g(x) = \sin x, x \in S$ is a bijection and $g = f|_S$

1.3.4 Composition of Mappings

Let $f : A \to B$ and $g : C \to D$ be two mappings such that $f(A) \subseteq C$. The *composition mapping* of f and g is defined by $(g \circ f)(x) = g(f(x))$. Thus $g \circ f : A \to D$ is defined if and only if $f(A) \subseteq C$, the the domain of g. Similarly we can define $f \circ g$.

Example 1.15. Let $f : \mathbb{Z} \to \mathbb{Q}$ and $g : \mathbb{Q} \to \mathbb{Q}$ be defined by $f(x) = \frac{1}{2}x$ for all $x \in \mathbb{Z}$ and $g(x) = x^2$ for all $x \in \mathbb{Q}$. Then $g \circ f : \mathbb{Z} \to \mathbb{Q}$ is defined by $(g \circ f)(x) = g(\frac{1}{2}x) = \frac{1}{4}x^2$ for all $x \in \mathbb{Z}$. Here $f \circ g$ is not defined since the range of g is not a subset of the domain of f.

Example 1.16. Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x + 1$ for all $x \in \mathbb{R}$ and $g(x) = 3x$ for all $x \in \mathbb{R}$. Here $f \circ g$ and $g \circ f$ both are defined and $(f \circ g)(x) = 3x + 1$ and $(g \circ f)(x) = 3x + 3$ for all $x \in \mathbb{R}$. This shows that $f \circ g \neq g \circ f$.

Theorem 1.2. Let $f : A \to B, g : B \to C, h : C \to D$ be three mappings. Then

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

Theorem 1.3. If $f : A \to B$ and $g : B \to C$ be both injective mappings then the composition mapping $g \circ f : A \to C$ is injective.

Proof. Let x_1, x_2 be two distinct elements in A. Let $f(x_1) = y_1$ and $f(x_2) = y_2$. Since f is injective then y_1 and y_2 are also distinct elements in B. Let $g(y_1) = z_1$ and $g(y_2) = z_2$. Since g is injective then z_1 and z_2 are also distinct elements in C. Thus $(g \circ f)x_1 = z_1 \neq z_2 = (g \circ f)(x_1)$. This shows that $x_1 \neq x_2 \Rightarrow (g \circ f)(x_1) \neq (g \circ f)(x_2)$ $\square$

Theorem 1.4. If $f : A \to B$ and $g : B \to C$ are both surjective, then the composition $g \circ f : A \to C$ is surjective.

Proof. Let z be an element of C. Since g is surjective, there is at least one preimage $y \in B$ such that $g(y) = z$. Since f is surjective and $y \in B$, there is at least one preimage of y in A, say $x \in A$, such that $f(x) = y$. Note that

$$(g \circ f)(x) = g(f(x)) = g(y) = z.$$

This implies that z has a preimage in A under the mapping $g \circ f$. Thus, $g \circ f$ is surjective. $\square$

1.3.5 Inverse Mapping

Let $f : A \to B$ be a mapping. If there exists a mapping $g : B \to A$ such that $g \circ f = I_A$, then g is said to be a left inverse of f. If there exists a mapping $h : B \to A$ such that $f \circ h = I_B$, then h is said to be a right inverse of f. f is said to be invertible if there exists a $g : B \to A$ such that $g \circ f = I_A$ and $f \circ g = I_B$. We denote it by $f^{-1} = g$.

Example 1.17. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 3x$ and let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = \frac{x}{3}$. Then,

$$(g \circ f)(x) = g(f(x)) = g(3x) = \frac{3x}{3} = x, \quad \forall x \in \mathbb{R},$$

and

$$(f \circ g)(x) = f(g(x)) = f\left(\frac{x}{3}\right) = 3 \cdot \frac{x}{3} = x, \quad \forall x \in \mathbb{R}.$$

Hence, $g \circ f = f \circ g = I_{\mathbb{R}}$. Therefore, g is the inverse of f.

2 Sequence of Real Numbers

A mapping $f : \mathbb{N} \to \mathbb{R}$ is said to be a sequence in $\mathbb{R}$. The f images are real numbers

$$f(1), f(2), f(3), \ldots.$$

A sequence of f is generally denoted by $\{f(n)\}$. We will write the symbol $\{x_n\}$.

Example 2.1. Let $f : \mathbb{N} \to \mathbb{R}$ be defined by $f(n) = n, n \in \mathbb{N}$. Then $f(1) = 1, f(2) = 2, f(3) = 3, \ldots.$ The sequence is denoted by $\{f(n)\}$. It is also denoted by $\{1, 2, 3, \ldots\}$.

Example 2.2. Let $f : \mathbb{N} \to \mathbb{R}$ be defined by $f(n) = 2n + 1$. The sequence is $\{2n + 1\}$ and is denoted by $\{3, 5, 7, \ldots\}$.

Example 2.3. Let $f : \mathbb{N} \to \mathbb{R}$ be defined by $f(n) = \frac{1}{n}$. The sequence is $\left\{\frac{1}{n}\right\}$ and is denoted by $\left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$.

Example 2.4. Let $f : \mathbb{N} \to \mathbb{R}$ be defined by $f(n) = \frac{n}{n+1}$. The sequence is $\left\{\frac{n}{n+1}\right\}$ and is denoted by $\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\right\}$.

Example 2.5. Let $f : \mathbb{N} \to \mathbb{R}$ be defined by $f(n) = (-1)^n$, $n \in \mathbb{N}$. The sequence is $\{(-1)^n\}$ and is denoted by $\{-1, 1, -1, 1, \ldots\}$.

Example 2.6. Let $f : \mathbb{N} \to \mathbb{R}$ be defined by $f(n) = \sin \frac{n\pi}{2}$, $n \in \mathbb{N}$. The sequence is $\{1, 0, -1, 0, 1, 0, \ldots\}$.

Example 2.7. Let $f : \mathbb{N} \to \mathbb{R}$ be defined by $f(n) = 2$ for all $n \in \mathbb{N}$. The sequence is $\{2, 2, 2, 2, \ldots\}$. It is called a *constant sequence*.

2.1 Bounded Sequence

A real sequence $\{f(n)\}$ is said to be *bounded above* if there exists a real number G such that $f(n) \leq G$ for all $n \in \mathbb{N}$ then G is said to be an upper bound of the sequence.

A real sequence $\{f(n)\}$ is said to be *bounded below* if there exists a real number g such that $f(n) \geq g$ for all $n \in \mathbb{N}$ then g is said to be a lower bound of the sequence.

A real sequence $\{f(n)\}$ is said to be a *bounded sequence* if there exist real numbers G, g such that $g \leq f(n) \leq G$ for all $n \in \mathbb{N}$.

Example 2.8. The sequence $\left\{\frac{1}{n}\right\}$ is a bounded sequence. Here 0 is the greatest lower bound and 1 is the least upper bound.

Example 2.9. The sequence $\{n^2\}$ is bounded below and unbounded above. Here,

$$\sup\{f(n)\} = \infty, \inf\{f(n)\} = 1.$$

Example 2.10. The sequence $\{-2n\}$ is bounded above and unbounded below. Here,

$$\sup\{f(n)\} = -2, \inf\{f(n)\} = -\infty.$$

Example 2.11. Let $f(n) = (-1)^n n$. The sequence $\{f(n)\}$ is unbounded above and unbounded below. Here,

$$\sup\{f(n)\} = \infty, \inf\{f(n)\} = -\infty.$$

2.2 Limit of a Sequence

Let $\{f(n)\}$ be a real sequence. A real number l is said to be a limit of the sequence $\{f(n)\}$ if corresponding to $\varepsilon > 0$, there exists a natural number N (depending on ε) such that

$$|f(n) - l| < \varepsilon \quad \text{for all } n \geq N.$$

In other words,

$$l - \varepsilon < f(n) < l + \varepsilon \quad \text{for all } n \geq N.$$

Mathematically, a real number l is said to be the limit of the sequence $\{f(n)\}$ if for every $\varepsilon > 0$, there exists a natural number N such that all elements of the sequence, except the first $N-1$ elements, lie in the interval $(l - \varepsilon, l + \varepsilon)$.

Theorem 2.1. The limit of a sequence $\{f(n)\}$ is unique.

Proof. Let l_1 and l_2 be two limits of $\{f(n)\}$. By definition, for every $\varepsilon > 0$, there exist N_1 and $N_2 \in \mathbb{N}$ such that:

$$|f(n) - l_1| < \varepsilon/2 \quad \text{for all } n \geq N_1,$$
$$|f(n) - l_2| < \varepsilon/2 \quad \text{for all } n \geq N_2.$$

Let $N = \max\{N_1, N_2\}$. Note that for all $n \geq N$

$$\begin{aligned}|l_1 - l_2| &= |l_1 - f(n) + f(n) - l_2| \\ &\leq |l_1 - f(n)| + |f(n) - l_2| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

This shows that $l_1 = l_2$. □

Definition 10 (Convergent Sequence)**.** A real sequence $\{x_n\}$ is said to be convergent sequence if it has a limit l. In this case the sequence $\{x_n\}$ is said to be convergent to l. We write $\lim_{n\to\infty} x_n = l$. A sequence is said to be divergent if it is not convergent.

2.3 Archimedean Property

If $x, y \in \mathbb{R}$ and $x > 0, y > 0$ then there exists a natural number n such that $ny > x$.

Example 2.12. We show that the sequence $\{\frac{1}{n}\}$ converges to 0. Let us choose a positive ϵ. By Archimedean property of $\mathbb{R}$, there exists a natural number N such that $0 < \frac{1}{N} < \epsilon$. This implies that $0 < \frac{1}{n} < \epsilon$ for all $n \geq N$. It follows that $|\frac{1}{n} - 0| < \epsilon$ for all $n \geq N$. Therefore $\lim_{n\to\infty} \frac{1}{n} = 0$.

Example 2.13. The sequence $\{\frac{n^2+1}{n^2}\}$ converges to 1: Let is take $\epsilon > 0$. Now $|\frac{n^2+1}{n^2} - 1| < \epsilon$ will hold if $\frac{1}{n^2} < \epsilon$, i.e., if $n \geqslant \frac{1}{\sqrt{n}}$. Let $N = \lfloor \frac{1}{\sqrt{\epsilon}} \rfloor + 1$. Then $|\frac{n^2+1}{n^2} - 1| < \epsilon$ for all $n \geqslant N$.

Example 2.14. Let $x_n = 2$ for all $n \in \mathbb{N}$. Then $|x_n - 2| < 2$ for all $n \in \mathbb{N}$. This implies that $\{x_n\}$ is convergent. This also shows that any constant sequence is convergent.

Theorem 2.2. A convergent is bounded.

Proof. Let $\{x_n\}$ be a convergent sequence and l be its limit. Let is choose $\epsilon = 1$. Then by definition, there exists a positive integer N such that

$$l - 1 < x_n < l + 1 \text{ for all } n \geqslant N. \tag{1}$$

Now let $B = \max\{x_1, \ldots, x_{N-1}, l+1\}$ and $b = \min\{x_1, \ldots, x_{N-1}, l-1\}$. Then from (1) we have $b \leq x_n \leq B$ for all $n \in \mathbb{N}$. Therefore $\{x_n\}$ is bounded. $\square$

Corollary 1. An unbounded sequence is not convergent. (Exercise!)

2.4 Limit Theorems

Theorem 2.3. Let $\{x_n\}$ and $\{y_n\}$ be two sequences converge to x and y respectively. Then

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

(2) if $c \in \mathbb{R}$ then $\lim_{n\to\infty} cx_n = cx$.

(3) $\lim_{n\to\infty} x_n y_n = xy$.

(4) $\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{x}{y}$, provided that $\{y_n\}$ is a sequence of non-zero real numbers and $y \neq 0$.

Proof. (1): By definition, for each $\epsilon > 0$, there exists positive integers N_1, N_2 such that

$$|x_n - x| < \epsilon/2 \text{ for all } n \geq N_1$$

and

$$|y_n - y| < \epsilon/2 \text{ for all } n \geq N_2.$$

Set $N = \max\{N_1, N_2\}$. Then we have

$$|(x_n + y_n) - (x + y)| \leq |x_n - x| + |y_n - y| < \epsilon/2 + \epsilon/2 = \epsilon \text{ for all } n \geq N.$$

This shows that $\lim_{n\to\infty}(x_n + y_n) = x + y$.

(2) and (3) are exercise. We prove (4).

(4): We first prove that if $\lim_{n\to\infty} y_n = y$ where $\{y_n\}$ is a sequence of non-zero real numbers and $y \neq 0$ then $\lim_{n\to\infty} \frac{1}{y_n} = \frac{1}{y}$.

Let $\epsilon_1 = \frac{1}{2}|y| > 0$. Since $\lim_{n\to\infty} y_n = y$ then there exists a natural number N_1 such that

$$|y_n - y| < \epsilon_1 \text{ for all } n \geq N_1.$$

Thus we have

$$||y_n| - |y|| \leq |y_n - y| < \epsilon_1 \text{ for all } n \geq N_1$$

which implies that

$$|y| - \epsilon_1 < |y_n| < |y| + \epsilon_1 \text{ for all } n \geq N_1. \tag{2}$$

From here we get $|y_n| > \frac{|y|}{2}$ for all $n \geq N_1$. Observe that for all $n \geq N_1$ we have

$$\frac{1}{|y_n|} < \frac{2}{|y|} \text{ for all } n \geqslant N_1. \tag{3}$$

Observe that

$$\left|\frac{1}{y_n} - \frac{1}{y}\right| = \frac{|y_n - y|}{|y_n|\,|y|} \leq 2\frac{|y_n - y|}{|y|^2} \text{ for all } n \geq N_1. \tag{4}$$

Let $\epsilon > 0$. Since $y_n \to y$ then there exists $N_2 \in \mathbb{N}$ such that

$$|y_n - y| < \frac{|y|^2}{2}\epsilon \text{ for all } n \geq N_2. \tag{5}$$

Let $N = \max\{N_1, N_2\}$. Thus from (4) and (5) we obtain

$$\left|\frac{1}{y_n} - \frac{1}{y}\right| < 2\frac{|y_n - y|}{|y|^2} < \epsilon \text{ for all } n \geq N. \tag{6}$$

As $\epsilon > 0$ is arbitrary then we have $\lim_{n\to\infty} \frac{1}{y_n} = \frac{1}{y}$. Therefore by (3) and the above observation we conclude that $\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{x}{y}$. $\square$

Theorem 2.4. Let $\{u_n\}$ be a convergent sequence of real numbers converging to u. Then $\lim_{n\to\infty} |u_n| = |u|$.

Proof. Let $\epsilon > 0$. There exists a natural number N such that $|u_n - u| < \epsilon$ for all $n \geq N$. Note that

$$||u_n| - |u|| \leq |u_n - u| < \epsilon \text{ for all } n \geq N.$$

This shows that $\lim_{n\to\infty} |u_n| = |u|$. $\square$

Note 2.1. The converse of the above theorem may not true in general. For example, let $u_n = (-1)^n$. The sequence $\{|u_n|\}$ converges to 1 but $\{u_n\}$ is a divergent sequence.

Theorem 2.5. Let $\{u_n\}$ and $\{v_n\}$ be two convergent sequences and suppose there exists a natural number N such that $u_n \geq v_n$ for all $n \geq N$. Then

$$\lim u_n \geq \lim v_n.$$

Proof. Let $\lim u_n = u$, $\lim v_n = v$ and $w_n = u_n - v_n$. Then $\{w_n\}$ is a convergent sequence such that $w_n \geq 0$ for all $n \geq N$ and

$$\lim w_n = u - v.$$

If possible, let $u - v < 0$. Choose $\varepsilon > 0$ such that $u - v + \varepsilon < 0$. Since $\lim w_n = u - v$, there exists a natural number N_2 such that

$$|w_n - (u - v)| < \varepsilon \quad \forall n \geq N_2.$$

Thus, for $n \geq \max\{N, N_2\}$,

$$u_n - v_n = w_n < u - v + \varepsilon < 0,$$

which contradicts $u_n - v_n \geq 0$. Hence $u \geq v$. $\square$

Theorem 2.6 (Sandwich Theorem). Let $\{u_n\}$, $\{v_n\}$, and $\{w_n\}$ be three sequences of real numbers, and suppose there exists a natural number m such that

$$u_n \leq v_n \leq w_n, \quad \forall n \geq m.$$

If $\lim u_n = \lim w_n = l$, then $\{v_n\}$ is convergent and $\lim v_n = l$.

Proof. Let $\varepsilon > 0$. Since $\lim u_n = l$, $\exists N_1$ such that

$$|u_n - l| < \varepsilon, \quad \forall n \geq N_1.$$

Since $\lim w_n = l$, $\exists N_2$ such that

$$|w_n - l| < \varepsilon, \quad \forall n \geq N_2.$$

Let $N_3 = \max\{m, N_1, N_2\}$. For $n \geq N_3$,

$$l - \varepsilon < u_n \leq v_n \leq w_n < l + \varepsilon,$$

which implies $|v_n - l| < \varepsilon$. Hence $\lim v_n = l$. □

Example 2.15. Prove that

$$\lim_{n\to\infty} \left(\frac{1}{\sqrt{n^2+1}} + \frac{1}{\sqrt{n^2+2}} + \cdots + \frac{1}{\sqrt{n^2+n}} \right) = 1.$$

Let

$$u_n = \frac{1}{\sqrt{n^2+1}} + \frac{1}{\sqrt{n^2+2}} + \cdots + \frac{1}{\sqrt{n^2+n}}.$$

We have

$$\frac{n}{\sqrt{n^2+n}} < u_n < \frac{n}{\sqrt{n^2+1}}, \quad \forall n \geq 2.$$

Note that

$$\lim_{n\to\infty} \frac{n}{\sqrt{n^2+n}} = 1, \quad \lim_{n\to\infty} \frac{n}{\sqrt{n^2+1}} = 1.$$

By the Sandwich Theorem, $\lim u_n = 1$.

2.5 Divergent Sequences

Definition 11. A real sequence $\{x_n\}$ is said to *diverge to* $+\infty$ if, for every preassigned positive number G, there exists a natural number N such that

$$x_n > G, \quad \forall n \geq N.$$

Definition 12. A real sequence $\{x_n\}$ is said to *diverge to* $-\infty$ if, for every preassigned positive number G, there exists a natural number N such that

$$x_n < -G, \quad \forall n \geq N.$$

Definition 13. A bounded sequence that is not convergent is said to be an *oscillatory sequence of finite oscillation.*

Definition 14. An unbounded sequence that is not properly divergent is said to be an *oscillatory sequence of infinite oscillation.*

Example 2.16 (Example 1)**.** The sequence $\{n\}$ diverges to $+\infty$.

Example 2.17 (Example 2)**.** The sequence $\{-n\}$ diverges to $-\infty$.

Example 2.18 (Example 3)**.** The sequence $\{(-1)^n\}$ is bounded but not convergent. It is an oscillatory sequence of finite oscillation.

Example 2.19 (Example 4)**.** The sequence $\{(-1)^n n\}$ is an unbounded sequence that is not properly divergent. It is an oscillatory sequence of infinite oscillation.

Example 2.20. Show that $\lim r^n = 0$ if $|r| < 1$.

Proof. **Case 1:** $r = 0$. In this case the sequence is $\{0, 0, 0, \dots\}$. The sequence converges to 0.

Case 2: $r \neq 0$ and $|r| < 1$. Since $|r| < 1$, we have $\frac{1}{|r|} > 1$. Let $\dfrac{1}{|r|} = 1 + a$ where $a > 0$

$$|r^n - 0| = |r|^n = \frac{1}{(1+a)^n}.$$

We have $(1 + a)^n > na$ for all $n \in \mathbb{N}$. So

$$|r^n - 0| < \frac{1}{na} \quad \text{for all } n \in \mathbb{N}.$$

Let $\varepsilon > 0$. Then $|r^n - 0| < \varepsilon$ holds if $n > \frac{1}{a\varepsilon}$. Let $N \geq \left\lceil \frac{1}{a\varepsilon} \right\rceil + 1$. Then N is a natural number and $|r^n - 0| < \varepsilon$ for all $n \geq N$. Since ε is arbitrary, we conclude that $\lim_{n\to\infty} r^n = 0$. ☐

2.6 Monotone Sequence

A real sequence $\{x_n\}$ is said to be a *monotonic increasing sequence* if

$$x_{n+1} \geq x_n \quad \forall n \in \mathbb{N}.$$

A real sequence $\{x_n\}$ is said to be a *monotonic decreasing sequence* if

$$x_{n+1} \leq x_n \quad \forall n \in \mathbb{N}.$$

A real sequence $\{x_n\}$ is said to be *monotone* if it is either increasing or decreasing.

Example 2.21. Let $x_n = 2^n$, $n \geq 1$. Then

$$x_{n+1} = 2^{n+1} = 2 \cdot 2^n > 2^n = x_n,$$

so $\{x_n\}$ is monotonic increasing.

Example 2.22. Let $x_n = \frac{1}{n}$, $n \geq 1$. Then

$$x_{n+1} - x_n = \frac{1}{n+1} - \frac{1}{n} = \frac{n - (n+1)}{n(n+1)} = -\frac{1}{n(n+1)} < 0.$$

Thus, $\{x_n\}$ is a monotonic decreasing sequence.

Example 2.23. The sequence $\{(-1)^n\}$ is neither a monotonic increasing nor a monotonic decreasing sequence. Therefore, it is not a monotone sequence.

Theorem 2.7. A monotone increasing sequence, if bounded above, is convergent and it converges to the least upper bound.

Proof. Let $\{x_n\}$ be a monotone increasing sequence bounded above and let M be its least upper bound. Then:

(1) $x_n \leq M$ for all $n \in \mathbb{N}$.

(2) For a pre-assigned $\varepsilon > 0$, there exists a natural number N such that

$$x_N > M - \varepsilon.$$

Since $\{x_n\}$ is monotone increasing, we have

$$M - \varepsilon < x_N \leq x_{N+1} \leq x_{N+2} \leq \cdots \leq M, \quad \forall n \geq N.$$

That is,

$$M - \varepsilon < x_n \leq M \quad \forall n \geq N.$$

This shows that the sequence $\{x_n\}$ converges and $\lim_{n\to\infty} x_n = M$. □

Theorem 2.8. A monotone decreasing sequence, if bounded below, is convergent and it converges to the greatest lower bound.

Proof. Similar to the previous theorem. □

2.7 Bounded Above and Bounded Below Sequences

Definition 15.
1. Let $S \subseteq \mathbb{R}$. A real number u is said to be an upper bound of S if $x \in S \Rightarrow x \leq u$.
2. A real number l is said to be a lower bound of S if $l \leq x$ for all $x \in S$.
3. S is said to be bounded above if S has an upper bound.
4. S is said to be bounded below if S has an lower bound.
5. S is said to be bounded set if S is bounded above as well as bounded below.

Example 2.24. Let $S = \{1, \frac{1}{2}, \frac{1}{3}, \ldots\}$. Then S is bounded above, 1 being an upper bound and bounded below, 0 being a lower bound.

Example 2.25. Let $S = \phi$. Then every real number is an upper bound as well as lower bound. Thus S is a bounded set.

Definition 16. Let $S \subseteq \mathbb{R}$.

1. If S is bounded above then an upper bound of S is said to be the *supremum* of S (or the *least upper bound* of S) if it is less than every upper bound of S.
2. If S is bounded below then a lower bound of S is said to be the *infimum* of S (or the *greatest lower bound* of S) if it is bigger than every lowerr bound of S.

Theorem 2.9. Let $\phi \neq S \subseteq \mathbb{R}$ be bounded above. An upper bound u of S is the supremum of S if and only if for each $\epsilon > 0$ there exists an element $s \in S$ such that $u - \epsilon < s \leq u$.

Proof. Let $u = \sup S$ and choose $\epsilon > 0$. Then $u - \epsilon$ is not an upper bound of S. Therefore there exists at least one element $s \in S$ such that $s > u - \epsilon$. Since $u = \sup S$ and $s \in S$ then we have $s \leq S$. Thus we obtain $u - \epsilon < s \leq u$.

Conversely, let u be an upper bound of S such that for each $\epsilon > 0$ there exists an element $z \in S$ such that $u - \epsilon < s \leq u$.
Claim: $u = \sup S$. We prove that u is the least upper bound of S. If possible let u_0 be an upper bound of S such that $u_0 < u$. Let $\epsilon = \frac{1}{2}(u - u_0)$. Then $\epsilon > 0$ and

$$u - \epsilon = u_0 + \epsilon. \tag{7}$$

By the stated condition, there exists an element $s^{'}$ in S such that

$$u - \epsilon < s^{'} \leqslant u \tag{8}$$

From (7) and (8) we have $u_0 + \epsilon < s^{'}$. This shows that u_0 is not an upper bound of S. This is a contradiction. Thus u is the least upper bound of S. □

Proof of the following theorem is similar. So we omit the proof.

Theorem 2.10. Let $\phi \neq S \subseteq \mathbb{R}$ be bounded above. A lower bound l of S is the infimum of S if and only if for each $\epsilon > 0$ there exists an element $s \in S$ such that $l \leqslant s < l + \epsilon$.

2.8 Properties of Supremum and Infimum

Let $\phi \neq S \subseteq \mathbb{R}$ be bounded above. Then $\sup S$ exists. Let $M = \sup S$. Then $M \in \mathbb{R}$ and M satisfies the following conditions:

1. $x \in S \Rightarrow x \leq M$,
2. for each $\epsilon > 0$ there exists an element $y(\epsilon)$ in S such that $M - \epsilon < y \leq M$.

Let $\phi \neq S \subseteq \mathbb{R}$ be bounded above. Then $\sup S$ exists. Let $m = \inf S$. Then $m \in \mathbb{R}$ and m satisfies the following conditions:

1. $x \in S \Rightarrow x \geq m$,
2. for each $\epsilon > 0$ there exists an element $z(\epsilon)$ in S such that $m \leq z < m + \epsilon$.

Theorem 2.11. A monotone increasing sequence that is unbounded above diverges to ∞.

Proof. Let $\{x_n\}$ be a monotone increasing sequence. Since it is unbounded above then for any preassigned positive G, however large, there exists a natural number N such that $x_k > G$. Since $\{x_n\}$ is monotonically increasing then

$$G < x_k \leq x_{k+1} \leq x_{k+2} \leq \dots.$$

This shows that $x_n > G$ for all $n \geq k$ and hence $\{x_n\}$ diverges to ∞. □

Theorem 2.12. A monotone decreasing sequence that is unbounded below diverges to $-\infty$.

The proof if Theorem 2.12 is similar. We we skip its proof.

2.9 Subsequence

Let $\{u_n\}$ be a sequence and $\{r_n\}$ be a strictly increasing sequence of natural numbers, i.e., $r_1 < r_2 < r_3 < \ldots$. Then the sequence $\{u_{r_n}\}$ is said to be a subsequence of the sequence of the sequence $\{x_n\}$.

Let $r : \mathbb{N} \to \mathbb{N}$ be a sequence of natural numbers such that $r_1 < r_2 < r_3 < \ldots$ and $u : \mathbb{R} \to \mathbb{R}$ be a real sequence. Then the composite mapping $u \circ r : \mathbb{N} \to \mathbb{R}$ is the subsequence of the real sequence u. The elements of the subsequence $u \circ r$ are $u_{r_1}, u_{r_2}, u_{r_3}, \ldots$.

Example 2.26. Let $u_n = \frac{1}{n}$ and $r_n = 2n, n \in \mathbb{N}$. Then $\{u_{r_n}\} = \{u_2, u_4, u_6, \ldots\} = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{6}, \ldots\}$. If $r_n = 2n - 1$ then $\{u_{r_n}\} = \{u_1, u_3, u_5, \ldots\} = \{1, \frac{1}{3}, \frac{1}{5}, \ldots\}$.

Example 2.27. Let $u_n = (-1)^n$ and $r_n = 2n, n \in \mathbb{N}$. Then $\{u_{r_n}\} = \{u_2, u_4, u_6, \ldots\} = \{1, 1, 1, \ldots\}$. If $r_n = 2n - 1$ then $\{u_{r_n}\} = \{u_1, u_3, u_5, \ldots\} = \{-1, -1, -1, \ldots\}$.

Example 2.28. Let $u_n = 1 + \frac{1}{n}$ and $r_n = n^2, n \in \mathbb{N}$. Then $\{u_{r_n}\} = \{1 + 1, 1 + \frac{1}{2^2}, 1 + \frac{1}{3^2}, \ldots\}$.

Theorem 2.13. If a sequence $\{u_n\}$ converges to l then every subsequence of $\{u_n\}$ also converges to l.

Proof. Let $\{r_n\}$ be a strictly increasing sequence of natural numbers. Then $\{u_{r_n}\}$ is a subsequence of the sequence of $\{u_n\}$. Let $\epsilon > 0$ and $\lim_{n\to\infty} u_n = l$. Then there exists a natural number N such that $|u_n - l| < \epsilon$ for all $n \geq N$. Since $\{r_n\}$ is a strictly increasing sequence of natural numbers, then there exists a natural number number N_0 such that $r_n > N$ for all $n \geq N_0$. Thus $l - \epsilon < u_{r_n} < l + \epsilon$ for all $n \geq N_0$. This shows that $\lim_{n\to\infty} u_{r_n} = l$. □

2.10 Subsequential Limit

Let $\{u_n\}$ be a real sequence. A real number l is said to be a subsequential limit of $\{u_n\}$ if there exists a subsequence of $\{u_n\}$ that converges to l.

Theorem 2.14 (Bolzano-Weierstrass Theorem)**.** Every bounded sequence of real numbers has a convergent subsequence.

2.11 Cauchy Sequence

Definition 17. A sequence $\{u_n\}$ is said to be a *Cauchy sequence* if for every $\varepsilon > 0$ there exists a natural number K such that

$$|u_m - u_n| < \varepsilon \quad \text{for all } m, n \geq K.$$

Replacing m by $n + p$ where $p = 1, 2, 3, \ldots$, the above condition can equivalently be stated as

$$|u_{n+p} - u_n| < \varepsilon \quad \text{for all } n \geq K,\ p = 1, 2, 3, \ldots$$

Theorem 2.15. A Cauchy sequence of real numbers is convergent.

Proof. Let $\{u_n\}$ be a Cauchy sequence. First, we prove that the sequence $\{u_n\}$ is bounded. Let $\varepsilon = 1$. Then there exists a natural number K such that

$$|u_m - u_n| < 1 \quad \text{for all } m, n \geq K.$$

Equivalently,

$$u_{K-1} - 1 < u_n < u_{K-1} + 1 \quad \text{for all } n \geq K.$$

Let

$$B = \max\{u_1, u_2, \ldots, u_{K-1}, u_{K-1} + 1\}, \quad b = \min\{u_1, u_2, \ldots, u_{K-1}, u_{K-1} - 1\}.$$

Then

$$b \leq u_n \leq B \text{ for all } \ n \in \mathbb{N},$$

which proves that the sequence $\{u_n\}$ is bounded. By the Bolzano–Weierstrass theorem, $\{u_n\}$ has a convergent subsequence. Let l be the limit of that convergent subsequence. Then l is a subsequential limit of $\{u_n\}$.

We now prove that the sequence $\{u_n\}$ converges to l. Let us choose $\varepsilon > 0$. Then there exists a natural number K such that

$$|u_m - u_n| < \tfrac{\varepsilon}{2}, \quad \forall m, n \geq K. \tag{9}$$

Since l is a subsequential limit of $\{u_n\}$, there exists a natural number $n_0 \geq K$ such that

$$|u_{n_0} - l| < \tfrac{\varepsilon}{2}. \tag{10}$$

From (9), for all $n \geq K$ we have

$$|u_n - u_{n_0}| < \tfrac{\varepsilon}{2}.$$

Therefore, using the triangle inequality,

$$|u_n - l| \leq |u_n - u_{n_0}| + |u_{n_0} - l| < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon, \quad \forall n \geq K.$$

Hence,

$$\lim_{n\to\infty} u_n = l.$$

Thus, every Cauchy sequence of real numbers is convergent. □

Note that

$$|u_n - l| = |u_n - u_q + u_q - l| \leq |u_n - u_q| + |u_q - l| < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon.$$

This shows that $\lim u_n = l$. This completes the proof.

Theorem 2.16. A convergent sequence is a Cauchy sequence.

Proof. Let $\{u_n\}$ be a convergent sequence and

$$\lim u_n = l.$$

For a pre-assigned $\varepsilon > 0$, there exists a natural number K such that

$$|u_n - l| < \tfrac{\varepsilon}{2} \quad \text{for all } n \geq K.$$

Let m, n be natural numbers with $m, n \geq K$. Then we have

$$|u_m - l| < \tfrac{\varepsilon}{2} \quad \text{and} \quad |u_n - l| < \tfrac{\varepsilon}{2}.$$

Now,

$$|u_m - u_n| \leq |u_m - l| + |l - u_n| < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon, \quad \text{for all } m, n \geq K.$$

That is, $|u_m - u_n| < \varepsilon$ for all $m, n \geq K$. This implies that $\{u_n\}$ is a Cauchy sequence. This completes the proof. □

Example 2.29. Prove that the sequence $\left\{\frac{1}{n}\right\}$ is a Cauchy sequence. Let $u_n = \frac{1}{n}$. Let $\varepsilon > 0$. Then there exists a natural number K such that

$$\frac{1}{K} < \frac{\varepsilon}{2} \quad \text{(by Archimedean property).}$$

Now, for $m, n \geq K$, we have

$$|u_m - u_n| = \left|\frac{1}{m} - \frac{1}{n}\right| \leq \frac{1}{m} + \frac{1}{n} \leq \frac{2}{K} < \varepsilon.$$

This proves that $\{u_n\}$ is a Cauchy sequence.

Example 2.30. Prove that the sequence $\{(-1)^n\}$ is not a Cauchy sequence.

Let $u_n = (-1)^n$. Then

$$|u_m - u_n| = |(-1)^m - (-1)^n| = \begin{cases} 0, & \text{if } m, n \text{ are both odd or both even,} \\ 2, & \text{if one of } m, n \text{ is odd and the other is even.} \end{cases} \tag{11}$$

Let $\varepsilon = \frac{1}{2}$. Then it is not possible to find a natural number K such that

$$|u_m - u_n| < \varepsilon \quad \text{for all } m, n \geq K.$$

Hence $\{u_n\}$ is not a Cauchy sequence.

3 Infinite Series

Let $\{u_n\}$ be a sequence. The sequence $\{S_n\}$ is defined by

$$S_1 = u_1, \quad S_2 = u_1 + u_2, \quad S_3 = u_1 + u_2 + u_3, \quad \ldots.$$

It is represented by the symbol

$$u_1 + u_2 + u_3 + \ldots$$

which is said to be an *infinite series* generated by $\{u_n\}$. The series is denoted by

$$\sum_{n=1}^{\infty} u_n,$$

where S_n is said to be the n-th term of the series. The elements of the sequence $\{S_n\}$ are called the *partial sums* of the series.

An infinite series $\sum u_n$ is said to be *convergent* or *divergent* according as the sequence $\{S_n\}$ is convergent or divergent. In case of convergence, if

$$\lim_{n\to\infty} S_n = S,$$

then S is said to be the *sum of the series* $\sum u_n$. If, however,

$$\lim_{n\to\infty} S_n = \infty \quad (\text{or } -\infty),$$

the series $\sum u_n$ is said to *diverge to* ∞ (or $-\infty$).

Example 3.1. Let us consider the series

$$\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\cdots$$

Here

$$u_n=\frac{1}{n(n+1)}, \quad \text{and the series is } \sum u_n.$$

Let

$$\begin{aligned}
S_n &= u_1+u_2+\cdots+u_n \\
&= \frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\cdots+\frac{1}{n(n+1)} \\
&= \left(1-\frac{1}{2}\right)+\left(\frac{1}{2}-\frac{1}{3}\right)+\cdots+\left(\frac{1}{n}-\frac{1}{n+1}\right) \\
&= 1-\frac{1}{n+1}.
\end{aligned}$$

Therefore $\lim_{n\to\infty} S_n = 1$. Thus the series $\sum u_n$ is convergent and the sum of the series is 1.

Example 3.2. Let us consider the series

$$1+2+3+\cdots.$$

Let the sequence of partial sums be $\{S_n\}$. Then

$$S_n = 1+2+\cdots+n=\frac{n(n+1)}{2}.$$

Thus $\lim_{n\to\infty} S_n=\infty$. Hence the series diverges.

3.1 Harmonic Series

Consider

$$1+\frac{1}{2}+\frac{1}{3}+\cdots$$

Let

$$\sum_{n=1}^{\infty} u_n \quad \text{be the series, where } u_n=\frac{1}{n}.$$

Let $\{S_n\}$ be the sequence of partial sums. Then

$$S_n=u_1+u_2+\cdots+u_n.$$

Then

$$S_1=1, S_2=1+\frac{1}{2},$$

$$S_4=1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}>1+\frac{1}{2}+\left(\frac{1}{4}+\frac{1}{4}\right)=1+2\cdot\frac{1}{2},$$

$$S_8=1+\frac{1}{2}+\left(\frac{1}{3}+\frac{1}{4}\right)+\left(\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}\right).$$

Now we can see that

$$S_8 > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) = 1 + 3 \cdot \frac{1}{2}.$$

Similarly

$$S_{16} > 1 + 4 \cdot \frac{1}{2}.$$

In general we have

$$S_{2^n} > 1 + n \cdot \frac{1}{2}.$$

Therefore

$$\lim_{n\to\infty} S_{2^n} = \infty.$$

Hence the sequence $\{S_n\}$ is a monotone increasing sequence. Since

$$S_{n+1} - S_n = u_{n+1} > 0 \quad \forall n \in \mathbb{N},$$

the sequence $\{S_n\}$ diverges to ∞. Therefore, the series $\sum_{n=1}^{\infty} u_n$ is divergent.

3.2 Geometric Series

Example 3.3. Let us consider the series

$$1 + a + a^2 + a^3 + \cdots \quad \text{where } |a| < 1.$$

Let $S_n = 1 + a + a^2 + \cdots + a^{n-1}$. Then $S_n = \dfrac{1-a^n}{1-a} = \dfrac{1}{1-a} - \dfrac{a^n}{1-a}$. From (1), we have $\lim_{n\to\infty} S_n = \dfrac{1}{1-a}$ because $\lim_{n\to\infty} a^n = 0$. Therefore the series is convergent and the sum of the series $= \dfrac{1}{1-a}$.

Example 3.4. Let us consider the series $1 + a + a^2 + \cdots \quad$ where $|a| \geq 1$. Let $S_n = 1 + a + a^2 + \cdots + a^{n-1}$.

Case 1: Suppose $a = 1$. In this case $S_n = n$ and $\lim_{n\to\infty} S_n = \infty$. Therefore the series diverges.

Case 2: Suppose $a > 1$. In this case $S_n = \dfrac{a^n - 1}{a - 1}$ and $\lim_{n\to\infty} S_n = \infty$ because $\lim_{n\to\infty} a^n = \infty$. Therefore the series is divergent.

Case 3: Suppose $a = -1$. Then

$$S_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even} \end{cases}$$

The set $\{S_n\}$ is divergent. Therefore the series is divergent.

Case 4. Suppose $q = -1$. In this case the sequence $\{s_n\}$ is divergent and therefore the series is divergent.

3.3 Cauchy's principle of convergence

Theorem 3.1. A necessary and sufficient condition for the convergence of a series $\sum u_n$ is that corresponding to a pre-assigned $\varepsilon > 0$ there exists a natural number m such that

$$|u_{n+1} + \cdots + u_{n+p}| < \varepsilon \quad \text{for } n \geq m \text{ and for all } p \in \mathbb{N}.$$

Proof. Let $s_n = u_1 + \cdots + u_n$. Let $\sum u_n$ be convergent. Then the sequence $\{s_n\}$ is convergent. Therefore the Cauchy principle of convergence for the sequence, corresponding to a pre-assigned positive ε there exists a natural number p such that

$$|S_{n+p} - S_n| < \varepsilon \quad \text{for } n \geq m, \ \forall \text{ natural no. } p.$$

This implies that

$$|u_{n+1} + \cdots + u_{n+p}| < \varepsilon \quad \forall n \geq m+1 \ \text{ for every natural no. } p.$$

Conversely, let us assume that for a pre-assigned $\varepsilon > 0$ there exists a natural no. p such that

$$|u_{n+1} + \cdots + u_{n+p}| < \varepsilon \quad \forall n \geq m+1 \ \text{ for every natural no. } p. \tag{1}$$

From (1), we have

$$|S_{n+p} - S_n| < \varepsilon \quad \forall n \geq m \ \text{ for every natural no. } p.$$

This implies that the sequence $\{S_n\}$ is convergent by Cauchy's principle of convergence. Therefore, $\sum u_n$ is convergent. This completes the proof. □

Theorem 3.2. A necessary condition for the convergence of a series $\sum u_n$ is $\lim u_n = 0$.

Proof. Let $\sum u_n$ be convergent. Then for a pre-assigned $\varepsilon > 0$ there exists a natural no. N such that

$$|u_{n+1} + \cdots + u_{n+p}| < \varepsilon \quad \forall n \geq N \ \text{ and for every natural no. } p.$$

Taking $p = 1$, we have $|u_{n+1}| < \varepsilon$ for $n \geq N$. This implies $\lim u_n = 0$. This completes the proof.
□ □

Remark. Note that the converse of the above statement is not true in general. Let us consider the series $\sum u_n$ where $u_n = \frac{1}{n}$. Here $\lim u_n = 0$. But we will show that $\sum u_n$ is a divergent series. Note that

$$|S_{n+p} - S_n| = \left| \frac{1}{n+1} + \cdots + \frac{1}{n+p} \right|. \tag{1}$$

If we take $p = n$, then from (1), we have

$$|S_{2n} - S_n| = \frac{1}{n+1} + \cdots + \frac{1}{2n} > \frac{1}{2n} + \cdots + \frac{1}{2n} = \frac{1}{2}.$$

Therefore $|S_{n+p} - S_n|$ cannot be made less than a chosen positive $\varepsilon < \frac{1}{2}$ for every natural no. p.

Example 3.5. Prove that the series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

is convergent. Let us consider the series $\sum u_n$, where

$$u_n = \frac{(-1)^{n+1}}{n}.$$

Let $S_n = u_1 + \cdots + u_n$. Then

$$\begin{aligned} |S_{n+p} - S_n| &= \left| \frac{1}{n+1} - \frac{1}{n+2} + \cdots + (-1)^{p-1}\frac{1}{n+p} \right| \\ &= \left| \frac{1}{n+1} - \left(\frac{1}{n+2} - \frac{1}{n+3} + \cdots \right) \right| \\ &< \frac{1}{n+1}. \end{aligned}$$

Let $\varepsilon > 0$. Then $|S_{n+p} - S_n| < \varepsilon$ holds if $n > \frac{1}{\varepsilon} - 1$. Let $N = \left[\frac{1}{\varepsilon} - 1\right] + 2$. Then N is a natural no. and $|S_{n+p} - S_n| < \varepsilon$ for $n \geq N$ and $p = 1, 2, \ldots$. This shows that the series is convergent.

Example 3.6. Let us consider the series $\sum u_n$ where

$$u_n = \frac{n}{n+1}.$$

This series is divergent, since $\lim u_n \neq 0$. Thus, $\sum u_n$ is divergent because a necessary condition for convergence is not satisfied for convergence of the series $\sum u_n$ it is necessary that $\lim u_n = 0$.

Example 3.7 (Introducing Brackets)**.** Let $\sum_{n\geq 1} u_n = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ldots$. Let us introduce the brackets and the series takes the form

$$1 + \frac{1}{2} + (\frac{1}{3} + \frac{1}{4}) + (\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}) + (\frac{1}{9} + \cdots + \frac{1}{16}) + \ldots.$$

Then $v_1 = 1, v_2 = \frac{1}{2}, v_3 = \frac{1}{3} + \frac{1}{4}, v_4 = \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}$ and so on. Thus the new series $\sum_{n\geq 1} v_n$ is obtained from $\sum_{n\geq 1} u_n$ by introducing the brackets.

Theorem 3.3. Let $\sum_{n\geq 1} u_n$ be a series of positive real numbers and $\sum_{n\geq 1} v_n$ be obtained from $\sum_{n\geq 1} u_n$ by grouping its terms. Then $\sum_{n\geq 1} u_n$ converges if and only if $\sum_{n\geq 1} v_n$ converges.

Proof. Exercise! □

Note 3.1. The above theorem does not hold if $\sum_{n\geq 1} u_n$ is a series of arbitrary terms. Let us consider the series

$$1 - 1 + 1 - 1 + 1 - 1 + \ldots. \tag{12}$$

By introducing brackets we get the series

$$(1-1) + (1-1) + (1-1) + \ldots. \tag{13}$$

Again by introducing new brackets we get the series

$$1 - (1-1) - (1-1) - (1-1) - \ldots. \tag{14}$$

It can be noted that the series in (12) diverges but the series in (13) and (14) converge to 0 and 1 respectively.

3.4 Re-arrangement of Terms

Let $\sum_{n\geq 1} u_n$ be a given series. If a new series $\sum_{n\geq 1} v_n$ is obtained by using each term of the series $\sum_{n\geq 1} u_n$ exactly once, the order of the terms being distributed then $\sum_{n\geq 1} v_n$ is called a re-arrangement of $\sum_{n\geq 1} u_n$.

Let $f : \mathbb{N} \to \mathbb{N}$ be a bijection map. $\sum_{n\geq 1} u_{f(n)}$ is a re-arrangement of $\sum_{n\geq 1} u_n$ and conversely if $\sum_{n\geq 1} v_n$ is a re-arrangement of the series $\sum_{n\geq 1} u_n$ then $v_n = u_{f(n)}$ for some bijection $f : \mathbb{N} \to \mathbb{N}$.

Example 3.8. Let $f : \mathbb{N} \to \mathbb{N}$ be defined by

$$f(n) = \begin{cases} n+1 \text{ if } n \text{ is odd} \\ n-1 \text{ if } n \text{ is even.} \end{cases}$$

Here $f(1) = 2, f(2) = 1, f(3) = 4, f(4) = 3$ and so on. Note that

$$\sum_{n\geq 1} u_{f(n)} = u_2 + u_1 + u_4 + u_3 + \dots$$

is a re-arrangement of $\sum_{n\geq 1} u_n$.

Theorem 3.4. Let $\sum_{n\geq 1} u_n$ be a series of positive real number and the series $\sum_{n\geq 1} u_n$ converges to l. Then any re-arrangement of $\sum_{n\geq 1} u_n$ are convergent and the re-arrangement of $\sum_{n\geq 1} u_n$ converges to l.

Proof. Let $\sum_{n\geq 1} v_n$ be a re-arrangement if $\sum_{n\geq 1} u_n$. Then $v_n = u_{f(n)}$ for some bijection $f : \mathbb{N} \to \mathbb{N}$. Let $s_n = u_1 + \cdots + u_n$ and $t_n = v_1 + \cdots + v_n$. Since $u_n > 0$, the sequence $\{s_n\}$ is a monotone increasing sequence. As $\sum_{n\geq 1} u_n$ converges to l then we have $\lim_{n\to\infty} s_n = l$. Therefore, the sequence $\{s_n\}$ is bounded above and $s_n \leq l$ for all $n \in \mathbb{N}$. Now

$$\begin{aligned} t_n &= v_1 + \cdots + v_n \\ &= u_{f(1)} + \cdots + u_{f(n)} \\ &\leq u_1 + \cdots + u_{m(n)} \end{aligned} \tag{15}$$

where $m(n) = \max\{f(1), \dots, f(n)\}$. But $u_1 + \cdots + u_{m(n)} = s_{m(n)} \leq l$. Thus the sequence $\{t_n\}$ is bounded above and being a monotone increasing, it is convergent. Let $\lim_{n\to\infty} t_n = t$. Then $t \leq l$. Note that

$$\begin{aligned} s_n &= u_1 + \cdots + u_n \\ &= v_{f^{-1}(1)} + \cdots + v_{f^{-1}(n)} \\ &\leq v_1 + \cdots + v_{k(n)} \end{aligned} \tag{16}$$

where $k(n) = \max\{f^{-1}(1), \dots, f^{-1}(n)\}$. From (16) we observe that

$$v_1 + \cdots + v_{k(n)} = t_{k(n)} \leq t,$$

i.e., $s_n \leq t$. This implies that $\lim_{n\to\infty} s_n \leq t$, i.e., $l \leq t$. This shows that $t = l$. This completes the proof. $\square$

Theorem 3.5 (Comparison test)**.** Let $\sum u_n$ and $\sum v_n$ be two series of positive real numbers, and there a natural number m such that exists a constant $K > 0$ such that $u_n \leq K v_n$ for all $n \geq m$, where K is a fixed positive real number. Then

1. $\sum u_n$ is convergent, if $\sum v_n$ is convergent.

2. $\sum v_n$ is divergent, if $\sum u_n$ is divergent.

Proof. Exercise. □

Theorem 3.6 (Limit Form of Comparison Test)**.** Let $\{u_n\}$ and $\{v_n\}$ be two series of positive real numbers. If

$$\lim_{n\to\infty} \frac{u_n}{v_n} = l, \quad 0 < l < \infty,$$

then the two series $\sum u_n$ and $\sum v_n$ converge or diverge together.

Note 3.2. 1. If $\lim_{n\to\infty} \frac{u_n}{v_n} = 0$, then $\sum u_n$ is convergent, if $\sum v_n$ is convergent (but not conversely).

2. If $\lim_{n\to\infty} \dfrac{u_n}{v_n} = \infty$, then $\sum u_n$ is divergent if $\sum v_n$ is divergent, but not conversely

3.5 p-Series

Theorem 3.7. The series

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

converges if $p > 1$ and diverges if $p \le 1$.

Proof. **Case 1:** $p > 1$**:** Suppose $p > 1$. Let $u_n = \frac{1}{n^p}$. The series $\sum u_n$ can be arranged by grouping terms as follows:

$$1 + \left(\frac{1}{2^p} + \frac{1}{3^p}\right) + \left(\frac{1}{4^p} + \frac{1}{5^p} + \frac{1}{6^p} + \frac{1}{7^p}\right) + \left(\frac{1}{8^p} + \cdots + \frac{1}{15^p}\right) + \cdots.$$

Let

$$v_1 = 1, \quad v_2 = \frac{1}{2^p} + \frac{1}{3^p}, \quad v_3 = \frac{1}{4^p} + \frac{1}{5^p} + \frac{1}{6^p} + \frac{1}{7^p}, \quad v_4 = \frac{1}{8^p} + \cdots + \frac{1}{15^p}, \text{ etc.}$$

Now we estimate:

$$v_2 \le \frac{2}{2^p}, \qquad v_3 \le \frac{4}{4^p}, \qquad v_4 \le \frac{8}{8^p}, \quad \text{and in general, } v_n \le \frac{2^{n-1}}{(2^{n-1})^p} = \frac{1}{(2^{n-1})^{p-1}}.$$

Thus the series is bounded by

$$1 + \frac{1}{2^{p-1}} + \frac{1}{(2^{p-1})^2} + \cdots,$$

which is a geometric series with common ratio $r = \frac{1}{2^{p-1}} < 1$ for $p > 1$. Hence, $\sum \frac{1}{n^p}$ converges for $p > 1$.
Case 2: $p = 1$**.** In this case the series becomes

$$\sum \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

which is the harmonic series. Let

$$S_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

Then

$$S_{2n} - S_n = \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n}.$$

This sum has n terms, each greater than $\frac{1}{2n}$, so

$$S_{2n} - S_n > \frac{n}{2n} = \frac{1}{2}.$$

Thus the sequence $\{S_n\}$ is not Cauchy, and hence it is not convergent. Therefore $\sum \frac{1}{n}$ diverges.
Case 3: $0 < p < 1$. Then

$$\frac{1}{n^p} > \frac{1}{n}, \quad \text{for large } n.$$

Hence

$$\sum \frac{1}{n^p} > \sum \frac{1}{n}.$$

But we know that $\sum \frac{1}{n}$ diverges. Therefore, $\sum \frac{1}{n^p}$ is also divergent.
Case 4: $p < 0$. Then $\lim 1/n^p \neq 0$ and therefore by necessary condition for convergent of a series, we conclude that $\sum \frac{1}{n^p}$ id not Convergent. This completes the proof. □

Example 3.9. Test of Convergence of the series

$$\frac{1+2}{2^3} + \frac{1+2+3}{3^3} + \cdots .$$

Let $\sum n = 1^\infty u_n$ be the given series. Then

$$u_n = \frac{1+2+\cdots+n}{(n+1)^3} = \frac{(n+1)(n+2)}{2(n+1)^3} = \frac{(n+2)}{2(n+1)^2}.$$

Now let

$$v_n = \frac{1}{n}.$$

Then we have

$$\lim_{n\to\infty} \frac{u_n}{v_n} = \lim_{n\to\infty} \frac{2n^3+3n^2+n}{6(n+1)^3} \cdot \frac{(n+1)^2}{n^2}.$$

It follows from here that

$$\lim_{n\to\infty} \frac{u_n}{v_n} = \lim_{n\to\infty} \frac{n(n+2)}{2(n+1^2)} = \frac{1}{2}.$$

Therefore, $\sum u_n$ is divergent by comparison test, since $\sum v_n$ is divergent.

Example 3.10. Test of convergence of the series

$$\frac{1}{1.2^2} + \frac{2}{2.3^2} + \frac{3}{3.4^2} + \cdots$$

Let $\sum u_n$ be the given series. Then

$$u_n = \frac{1}{n(n+1)^2}.$$

Let

$$v_n = \frac{1}{n^3}.$$

Then

$$\lim_{n\to\infty} \frac{u_n}{v_n} = 1.$$

Since $\sum v_n = \sum \frac{1}{n^3}$ is convergent, therefore by the comparison test, $\sum u_n$ is also convergent.

Example 3.11. Test of convergence of the series $\sum u_n$, where

$$u_n = \sqrt{n^4+1} - \sqrt{n^4-1}.$$

Then

$$u_n = \frac{2}{\sqrt{n^4+1} + \sqrt{n^4-1}}.$$

Let

$$v_n = \frac{1}{n^2}.$$

Then

$$\lim_{n\to\infty} \frac{u_n}{v_n} = 1.$$

But $\sum v_n = \sum \frac{1}{n^2}$ is convergent, by comparison test $\sum u_n$ is also convergent.

Theorem 3.8 (Comparison Test)**.** Let $\sum u_n$ and $\sum v_n$ be two series of non-negative terms. Suppose there is a natural number m such that

$$\frac{u_{n+1}}{u_n} \leq \frac{v_{n+1}}{v_n}, \quad \text{for all } n \geq m.$$

Then

1. $\sum u_n$ is convergent if $\sum v_n$ is convergent.
2. $\sum v_n$ is divergent if $\sum u_n$ is divergent.

Proof. Exercise. □

Theorem 3.9 (D'Alembert's Ratio Test)**.** Let $\sum u_n$ be a series of positive terms and let

$$\lim_{n\to\infty} \frac{u_{n+1}}{u_n} = l.$$

Then $\sum u_n$ is convergent if $l < 1$, and $\sum u_n$ is divergent if $l > 1$.

Remark. The test fails to give a decision if $l = 1$.

Example 3.12. Let $u_n = \frac{1}{n}$. Then $\sum u_n$ is divergent series and

$$\lim_{n\to\infty} \frac{u_{n+1}}{u_n} = 1.$$

Theorem 3.10 (Cauchy's Root Test)**.** Let $\sum u_n$ be a series of positive terms. Let

$$\lim_{n\to\infty} \sqrt[n]{u_n} = l.$$

Then $\sum u_n$ is convergent if $l < 1$, and $\sum u_n$ is divergent if $l > 1$.

Remark. If $l = 1$, the test fails.

Example 3.13. Test the convergence of the series

$$1 + \frac{3}{2!} + \frac{5}{3!} + \frac{7}{4!} + \cdots .$$

Let $\sum u_n$ be the given series, where

$$u_n = \frac{2n-1}{n!}.$$

Then

$$\lim_{n\to\infty} \sqrt[n]{u_n} = \lim_{n\to\infty} \sqrt[n]{\frac{2n-1}{n!}} = 0.$$

By Cauchy's root test, $\sum u_n$ is convergent.

Example 3.14. Examine the convergence of the series

$$x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots , \quad x > 0.$$

Let $\sum u_n$ be the given series. Since $x > 0$, $\sum u_n$ is a series of positive terms, where

$$u_n = \frac{x^n}{n}, \qquad \frac{u_{n+1}}{u_n} = \frac{x^{n+1}}{n+1} \cdot \frac{n}{x^n} = \frac{nx}{n+1}.$$

Thus,

$$\lim_{n\to\infty} \frac{u_{n+1}}{u_n} = x.$$

$\sum u_n$ is convergent if $x < 1$, and $\sum u_n$ is divergent if $x > 1$. When $x = 1$, the series becomes

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

and thus is divergent.

Example 3.15. Test of convergence of the series

$$1 + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \cdots .$$

Let

$$u_n = \frac{1}{2^n + (-1)^n}.$$

Now,

$$\lim_{n\to\infty} \sqrt[n]{u_n} = \lim_{n\to\infty} \frac{1}{\sqrt[n]{2^n + (-1)^n}} = \frac{1}{2}.$$

Therefore, the series is convergent by Cauchy's root test.

Theorem 3.11 (Cauchy's Condensation Test)**.** Let $\{f(n)\}$ be a monotone decreasing sequence of positive real numbers and $a > 1$. Then the series $\sum f(n)$ and $\sum a^n f(a^n)$ converge and diverge together.

Example 3.16. Test the convergence of the series $\frac{1}{n}$.

Let $f(n) = \frac{1}{n}$. Then $\{f(n)\}$ is a monotone decreasing sequence of positive real numbers. Note that $2^n f(2^n) = 1$ and therefore $\sum 2^n f(2^n)$ is divergent. It follows from Cauchy's condensation test that the series $\sum \frac{1}{n}$ is divergent.

Example 3.17. Discuss the convergence of the series $\sum_{n\geqslant 2}\frac{1}{n(\log n)^p}$ for some $p>0$.

Let $f(n)=\frac{1}{n(\log n)^p}, n\geqslant 2$. As $\{\log n\}$ is an increasing sequence and $p>0$ then $(\log(n+1))^p > (\log n)^p$ and therefore $(n+1)(\log(n+1))^p > n(\log n)^p$. Therefore $\{f(n)\}_{n\geqslant 2}$ is a monotone decreasing sequence of positive real numbers. Note that $\sum 2^n f(2^n)=\sum\frac{1}{(n\log 2)^p}$ and thus it converge when $p>1$ and diverge when $p\leqslant 1$. Hence by Cauchy's condensation test, the series $\sum_{n\geqslant 2} f(n)$ converge when $p>1$ and diverge when $p\leqslant 1$.

3.6 Absolutely Convergent Series

Let $\sum u_n$ be a series of positive and negative real numbers. Let $u_n^{'}=|u_n|$. Then $\sum u_n^{'}$ is a series of positive real numbers. If $\sum u_n^{'}$ is convergent then $\sum u_n$ is said to be an *absolutely convergent series*.

Theorem 3.12. An absolutely convergent series is convergent.

Proof. Let $\sum u_n$ be a series of positive and negative real numbers ans be absolutely convergent. Then $\sum |u_n|$ is a convergent series of positive terms. Let us choose $\epsilon>0$. Then there exists a natural number m such that

$$|u_{n+1}|+\cdots+|u_{n+p}|<\epsilon \text{ for all } n\geqslant m, \text{ and } p\in\mathbb{N}.$$

Note that

$$|u_{n+1}+\cdots+u_{n+p}|\leqslant |u_{n+1}|+\cdots+|u_{n+p}|<\epsilon \text{ for all } n\geqslant m, \text{ and } p\in\mathbb{N}.$$

By Cauchy's principle of convergence $\sum u_n$ is convergent. □

Example 3.18. The series $1-\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}-\frac{1}{5^2}-\frac{1}{6^2}+\dots$ is convergent as it is absolutely convergent.

Example 3.19. For a fixed value of x, the series $\sum\frac{\sin nx}{n^2}$ is absolutely convergent.

Note that $\sum\frac{\sin nx}{n^2}$ is a series of arbitrary terms and $\left|\frac{\sin nx}{n^2}\right|\leqslant\frac{1}{n^2}$. Since $\frac{1}{n^2}$ is a convergent series, $\sum\left|\frac{\sin nx}{n^2}\right|$ is a convergent series by comparison test. Hence $\sum\frac{\sin nx}{n^2}$ is absolutely convergent.

Theorem 3.13. If $\sum u_n$ be absolutely convergent and $\{v_n\}$ be a bounded sequence then the series $\sum u_n v_n$ is absolutely convergent.

Proof. By hypothesis, $|v_n|\leqslant K$ for all $n\in\mathbb{N}$, for some $K>0$. Note that

$$|u_{n+1}v_{n+1}|+\cdots+|u_{n+p}v_{n+p}|\leqslant K(|u_{n+1}|+\cdots+|u_{n+p}|). \tag{17}$$

Since $\sum u_n$ is absolutely convergent then $\sum |u_n|$ is convergent. Therefore for any given $\epsilon>0$ there exists a natural number m such that

$$|u_{n+1}|+\cdots+|u_{n+p}|<\epsilon/K \text{ for all } n\geqslant m \text{ for any } p\in\mathbb{N}. \tag{18}$$

Therefore for (18) we have

$$|u_{n+1}v_{n+1}|+\cdots+|u_{n+p}v_{n+p}|<\epsilon \text{ for all } n\geqslant m \text{ for any } p\in\mathbb{N}.$$

Therefore by Cauchy's principle of convergence the series $\sum |u_n v_n|$ is convergent and hence $\sum u_n v_n$ is absolutely convergent. □

Example 3.20. Test the convergence of the series $\frac{2}{1^3}-\frac{3}{2^3}+\frac{4}{3^3}-\frac{5}{4^3}+\dots$.

Let $\sum u_n$ be the given series. Then $u_n=(-1)^{n+1}\frac{n+1}{n^3}=\frac{(-1)^{n+1}}{n^2}(1+\frac{1}{n})=a_n b_n$ where $a_n=\frac{(-1)^{n+1}}{n^2}$ and $b_n=1+\frac{1}{n}$. Note that $\sum a_n$ is absolutely convergent and $\{b_n\}$ is bounded. Therefore the series $\sum a_n b_n$ is absolutely convergent.

3.7 Alternating Series

If $u_n > 0$ for all $n \in \mathbb{N}$, the series $\sum_{n \geqslant 1} (-1)^{n+1} u_n$ is called *alternating series.*

Theorem 3.14 (Leibnitz's Test)**.** If $\{u_n\}$ be a monotone decreasing sequence of positive real numbers and $\lim_{n \to \infty} u_n = 0$ then the alternating series $u_1 - u_2 + u_3 - u_4 + \dots$ is convergent.

Example 3.21. The series $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$ is convergent by Leibnitz's test.

4 Limit

4.1 Neighbourhood and Limit Point

First we start with the definition of neighbourhood and limit point.

Definition 18. Let $c \in \mathbb{R}$. A subset S of $\mathbb{R}$ is called a neighbourhood of c if there exists an open interval (a, b) such that $c \in (a, b) \subset S$.

Definition 19. Let $S \subseteq \mathbb{R}$. A point $p \in \mathbb{R}$ is said to be a limit point of S if every neighbourhood of p contains a point of S other than p, that is, for each $\epsilon > 0$,

$$(N(p, \epsilon) \setminus \{p\}) \cap S \neq \phi.$$

Example 4.1. Let $S_1 = (0, 1), S_2 = (0, 1], S_3 = [0, 1)$ and $S_4 = [0, 1]$. Here 0 and 1 are limit points for S_1, S_2, S_3 and S_4.

Example 4.2. Let $S = \{\frac{1}{n} : n \in \mathbb{N}\}$. Then 0 is the only limit point of S (Why? Exercise!).

4.2 Limit of a Function

Definition 20. Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let c be a limit point of D. A real number l is said to be a limit of f at c if corresponding to any neighbourhood V of l there exists a neighbourhood W of c such that $f(x) \in V$ for all $x \in (W \setminus \{c\}) \cap D$. In symbol, we write it as $\lim_{x \to c} f(x) = l$.

Definition 21 (Equivalent Definition)**.** Let $D \subseteq \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ be a function. Let c be a limit point of D. A real number l is said to be a limit of f at c if for any $\epsilon > 0$ there exists $\delta > 0$ such that

$$l - \epsilon < f(x) < l + \epsilon \text{ for all } x \in (c - \delta, c + \delta) \setminus \{c\}.$$

Example 4.3. Show that $\lim_{x \to 2} f(x) = 4$ where $f(x) = \frac{x^2-4}{x-2}, x \neq 2$.

Here the domain D of f is $\mathbb{R} \setminus \{2\}$ and 2 is the limit point of D. When $x \in D$ then

$$\begin{aligned} |f(x) - 4| &= \left| \frac{x^2 - 4}{x - 2} - 4 \right| \\ &= |x - 2|. \end{aligned}$$

Let us choose $\epsilon > 0$. Note that $|f(x) - 4| < \epsilon$ whenever $|x - 2| < \epsilon$. Therefore $|f(x) - 4| < \epsilon$ for all $x \in ((2 - \epsilon, 2 + \epsilon) \setminus \{2\}) \cap D$. Hence $\lim_{x \to 2} f(x) = 4$.

Example 4.4. Show that $\lim_{x \to 0} f(x) = 0$ where $f(x) = \sqrt{x}$. Notice that here the domain D is $\{x \in \mathbb{R} : x \geqslant 0\}$ and 0 is a limit point of D. Choose $\epsilon > 0$. When $x \geqslant 0$, we have $|f(x) - 0| = \sqrt{x}$. Thus $|f(x) - 0| < \epsilon$ for all x satisfying $0 < x < \epsilon^2$ where $\delta = \epsilon^2$. This shows that $|f(x) - 0| < \epsilon$ for all $x \in (N(0, \delta) \setminus \{0\}) \cap D$.

Theorem 4.1. Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let c be a limit point of D. Then f can have at most one limit at c.

Proof. Suppose, on the contrary, there exists two limits l and m at c. Since $l \neq m$ wlog we may assume $m > l$. Let $\epsilon = \frac{m-l}{2}$. Then $(l-\epsilon, l+\epsilon) \cap (m-\epsilon, m+\epsilon) = \phi$. Since l is a limit of f at c, then there exists a $\delta_1 > 0$ such that

$$l - \epsilon < f(x) < l + \epsilon \text{ for all } x \in ((c-\delta_1, c+\delta_1) \setminus \{c\}) \cap D.$$

Also, since m is a limit of f at c, then then there exists a $\delta_2 > 0$ such that

$$m - \epsilon < f(x) < m + \epsilon \text{ for all } x \in ((c-\delta_2, c+\delta_2) \setminus \{c\}) \cap D.$$

Let $\delta = \min\{\delta_1, \delta_2\}$. Then

$$l - \epsilon < f(x) < l + \epsilon \text{ and } m - \epsilon < f(x) < m + \epsilon \text{ for all } x \in ((c-\delta, c+\delta) \setminus \{c\}) \cap D.$$

This is a contradiction, since $(l-\epsilon, l+\epsilon) \cap (m-\epsilon, m+\epsilon) = \phi$. Therefore $l = m$ and hence this completes the proof. □

4.3 Sequential Criterion

Theorem 4.2. Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let c be a limit point of D and $l \in \mathbb{R}$. Then $\lim_{x\to c} f(x) = l$ if and only if for every sequence $\{x_n\}$ in $D \setminus \{c\}$ converging to c, the sequence $\{f(x_n)\}$ converges to l.

Proof. Let $\lim_{x\to c} f(x) = l$. Then for any pre-assigned $\epsilon > 0$, there exists $\delta > 0$ such that for all

$$l - \epsilon < f(x) < l + \epsilon \text{ for all } x \in (c-\delta, c+\delta) \setminus \{c\}. \tag{19}$$

Let $\{x_n\}$ be a sequence in $D \setminus \{c\}$ converging to c. Since $\lim_{n\to\infty} x_n = c$, then there exists $k \in \mathbb{N}$ such that $c - \delta < x_n < c + \delta$ for all $n \geqslant k$. Therefore from (19) we have

$$l - \epsilon < f(x_n) < l + \epsilon \text{ for all } n \geqslant k.$$

This shows that $\lim_{n\to\infty} f(x_n) = l$.

Conversely, let for every sequence $\{x_n\}$ in $D \setminus \{c\}$ converging to c, $\lim_{n\to\infty} f(x_n) = l$. We prove that $\lim_{x\to c} f(x) = l$. If not let, there exists a neighbourhood V of l such that for every neighbourhood W of c there exists at least one element $x_w \in (W \setminus \{c\}) \cap D$ for which $f(x_w)$ does not belong tp V. Let $W_1 = N(c, 1)$. Then there exists an element $x_1 \in (N(c,1) \setminus \{c\}) \cap D$ such that $f(x_1) \notin V$. Let $W_2 = N(c, 1/2)$. Then there exists an element $x_2 \in (N(c, 1/2) \setminus \{c\}) \cap D$ such that $f(x_2) \notin V$. Proceeding in this manner, we obtain a sequence $\{x_n\}$ in $D \setminus \{c\}$ such that $\lim_{n\to\infty} x_n = c$. Since $x_n \in W_n = N(c, 1/n)$ for all $n \in \mathbb{N}$, but the sequence $f(x_n) \notin V$, then $f(x_n)$ does not converge to l. This is a contradiction to the hypothesis and therefore $\lim_{x\to c} f(x) = l$. This completes the proof. □

Example 4.5. Prove that $\lim_{x\to 0} f(x)$ does not exist where $f(x) = \sin\frac{1}{x}, x \neq 0$.

Here the domain D is $\{x \in \mathbb{R} : x \neq= 0\} = \mathbb{R} \setminus \{0\}$. So 0 is a limit point of D. Let us consider the sequence $\{x_n\}$ in D defined by $x_n = \frac{2}{(4n-3)\pi}, n \in \mathbb{N}$. The sequence is $\{\frac{2}{\pi}, \frac{2}{5\pi}, \frac{2}{9\pi}, \dots\}$ and this converges to 0. The sequence $\{f(x_n)\}$ is $\{\sin\frac{\pi}{2}, \sin\frac{5\pi}{2}, \sin\frac{9\pi}{2}, \dots\}$ and converges to 1. Now consider $y_n = \frac{1}{n\pi}, n \in \mathbb{N}$. Then $y_n \to 0$. But the sequence $\{f(y_n)\}$ converges to 0. Thus we have two different sequences $\{x_n\}$ and $\{y_n\}$ in D both converging to 0 but the sequences $\{f(x_n)\}$ and $\{f(y_n)\}$ converge to different limits. Thus $\lim_{x\to 0} \sin\frac{1}{x}$ does not exist.

Example 4.6. Show that $\lim_{x\to 0}[x]$ does not exist.

Let $f(x) = [x]$. The domain of f is $\mathbb{R}$. Note that

$$f(x) = \begin{cases} -1, \text{ if } -1 < x < 0 \\ 0, \text{ if } 0 \leqslant x < 1. \end{cases}$$

Let us consider the sequence $x_n = \frac{1}{n+1}$ in $N(0,1), n \in \mathbb{N}$. Then $x_n \to 0$. By definition $\{f(x_n)\}$ is also converges to 0. Let us consider the sequence $y_n = -\frac{1}{n+1}$. Then $y_n \to 0$ and by definition the sequence $\{f(y_n)\}$ converges to -1. Thus we have two different sequences $\{x_n\}$ and $\{y_n\}$ in D both converging to 0 but the sequences $\{f(x_n)\}$ and $\{f(y_n)\}$ converge to different limits. Thus $\lim_{x\to 0}[x]$ does not exist.

Theorem 4.3. Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let c be a limit point of D. If f has a limit $l \in \mathbb{R}$ at c then f is bounded on $N(c) \cap D$ for some neighbourhood $N(c)$ of c.

Proof. Let us choose $\epsilon = 1$. Then there exists a $\delta > 0$ such that

$$|f(x) - l| < 1 \text{ for all } x \in (N(c,\delta) \setminus \{c\}) \cap D.$$

Note that

$$||f(x)| - |l|| \leqslant |f(x) - l| < 1 \text{ for all } x \in (N(c,\delta) \setminus \{c\}) \cap D.$$

It follows that

$$|f(x)| < |l| + 1 \text{ for all } x \in (N(c,\delta) \setminus \{c\}) \cap D.$$

This shows that f is bounded on $N(c,\delta) \cap D$. If however, $c \in D$, let $B = \max\{|f(c)|+1, |l|+1\}$. Then $|f(x)| < B$ for all $x \in N(c,\delta) \cap D$. This shows that f is bounded on $N(c,\delta) \cap D$. This completes the proof. $\square$

Let us define a sequence $\{u_n\}$ by $u_n = \sup\{f(x) : x \in A_n\}$. Then $f(x) \leq u_n$ for all $n \in \mathbb{N}$. Since $\lim u_n = l$ and $\lim f(x) \leq b$, it follows that $l \leq b$.

Theorem 4.4. Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let c be a limit point of D. If $f(x) \geq a$ for all $x \in D \setminus \{c\}$ and $\lim\limits_{x\to c} f(x) = l$, then $l \geq a$.

Proof. Exercise. $\square$

Theorem 4.5 (Sandwich Theorem)**.** Let $D \subseteq \mathbb{R}$ and f, g, h be functions defined on D to $\mathbb{R}$. Let c be a limit point of D. If $f(x) \leq g(x) \leq h(x)$ for all $x \in D \setminus \{c\}$ and if $\lim\limits_{x\to c} f(x) = \lim\limits_{x\to c} h(x) = l$, then $\lim\limits_{x\to c} g(x) = l$.

Proof. It follows from the previous theorem. $\square$

Theorem 4.6 (Cauchy's Principle)**.** Let $D \subseteq \mathbb{R}$ be a function and let c be a limit point of D. A necessary and sufficient condition for the existence of $\lim\limits_{x\to c} f(x)$ as $x \to c$ is that for every $\epsilon > 0$, there exists $\delta > 0$ such that for all $x, y \in (N(c,\delta) \cap D) \setminus \{c\}$,

$$|f(x) - f(y)| < \epsilon.$$

Proof. Exercise. $\square$

Example 4.7. A function $f : (0,1) \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} x, & \text{if } x \text{ is rational,} \\ 1-x, & \text{if } x \text{ is irrational.} \end{cases}$$

Using Cauchy's principle, we prove that $\lim_{x\to 0} f(x)$ does not exist, where $x \in (0,1)$. The domain of f is $D = (0,1)$. Let $\epsilon = 1$. Let a first point x be rational and a second point y be irrational, and let x and y be irrational points such that

$$N(\mathbb{Q} \cap (0,\delta) \setminus \{5\}) \cap N((0,1) \setminus \mathbb{Q} \cap (0,\delta)) = \emptyset.$$

Then

$$|f(x) - f(y)| = |x - (1-y)| = |x + y - 1| \geq 1.$$

Thus the limit does not exist.

Example 4.8. Let $f(x) = \frac{x}{x^2+1}$ and suppose $\lim_{x\to 0} f(x)$ exists. Let $D = \mathbb{R}$. Then $f(0) = 0$. Let $D_1 = (-\infty, 0)$ and $D_2 = (0, +\infty)$. Clearly, $D = D_1 \cup D_2 \cup \{0\}$. Suppose the limits $\lim_{x\to 0^-} f(x)$ and $\lim_{x\to 0^+} f(x)$ exist and are defined above. Then

$$\lim_{x\to 0} f(x) = l \quad \text{if and only if} \quad \lim_{x\to 0^-} f(x) = l \quad \text{and} \quad \lim_{x\to 0^+} f(x) = l.$$

Theorem 4.7 (Infinite Limit)**.** Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let c be a limit point of D. We write $\lim_{x\to c} f(x) = \infty$ if for every $G > 0$, there exists $\delta > 0$ such that for all $x \in (N(c,\delta) \cap D) \setminus \{c\}$, we have $f(x) > G$

Example 4.9. Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let c be a limit point of D. Suppose the limit $\lim_{x\to c} f(x)$ does not exist. Let

$$D_1 = \{x \in D : x < c\}, \quad D_2 = \{x \in D : x > c\}.$$

Then D_1 and D_2 are defined above. Let $\epsilon > 0$ such that for all $\delta > 0$, there exist $x_1 \in (N(c,\delta) \cap D_1) \setminus \{c\}$ and $x_2 \in (N(c,\delta) \cap D_2) \setminus \{c\}$ such that

$$|f(x_1) - f(x_2)| \geq \epsilon.$$

Example 4.10. Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let c be a limit point of D and $L \in \mathbb{R}$. We say that f tends to L (respectively ∞ or $-\infty$) as $x \to c$ if

$$\lim_{x\to c} f(x) = L \quad (\text{respectively } \lim_{x\to c} f(x) = \pm\infty).$$

Example 4.11. Suppose that

$$\lim_{x\to c} f(x) = \infty.$$

In every neighborhood $N(c,\delta)$ of c, no matter how small $\delta > 0$ is chosen, there exists $x \in N(c,\delta) \cap D$ with $x \neq c$ such that

$$f(x) > G$$

for any given $G > 0$. This shows that the definition above implies $\lim_{x\to c} f(x) = \infty$.

Definition 22. Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let $c \in D$ be a limit point of D. We say that f tends to ∞ (respectively $-\infty$) for some $c \in D$ if there exists $\delta > 0$ such that for all $x \in (N(c,\delta) \cap D) \setminus \{c\}$,

$$f(x) > G \quad (\text{respectively } f(x) < G)$$

for any $G > 0$.

4.4 Limits at Infinity

Definition 23. Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let $C \subset (c, \infty) \cap D$ (respectively $C \subset (-\infty, c) \cap D$). We say that f tends to l (respectively f tends to l) as $x \to \infty$ (respectively $x \to -\infty$) if for every $\varepsilon > 0$ there exists $K > 0$ such that

$$|f(x) - l| < \varepsilon \quad \forall x \in C, \ |x| > K.$$

In this case we write

$$\lim_{x\to\infty} f(x) = l \quad \text{(respectively } \lim_{x\to-\infty} f(x) = l).$$

Example 4.12. Let $f(x) = \frac{1}{x}$, $D = (0, \infty)$. Prove that $\lim_{x\to\infty} f(x) = 0$.

$$\lim_{x\to\infty} \frac{1}{x} = 0.$$

Given $\varepsilon > 0$, choose $G = \frac{1}{\varepsilon}$. Then for all $x > G$, we have

$$\left| \frac{1}{x} - 0 \right| = \frac{1}{x} < \varepsilon.$$

Thus the result follows.

4.5 Infinite Limits at Infinity

Definition 24. Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let $C \subset (c, \infty) \cap D$ (respectively $C \subset (-\infty, c) \cap D$). We say that f tends to ∞ (respectively f tends to $-\infty$) as $x \to \infty$ (respectively $x \to -\infty$) if for every $G > 0$ there exists $K > 0$ such that

$$f(x) > G \quad \text{(respectively } f(x) < -G) \quad \forall x \in C, \ |x| > K.$$

We write

$$\lim_{x\to\infty} f(x) = \infty \quad \text{(respectively } \lim_{x\to-\infty} f(x) = -\infty).$$

Example 4.13. Show that $\lim_{x\to\infty} \sqrt{x} = \infty$.

$$\lim_{x\to\infty} \sqrt{x} = \infty.$$

Given $G > 0$, let us choose $K = G^2$. Then for $x > K$, we have

$$\sqrt{x} > \sqrt{K} = \sqrt{G^2} = G.$$

This proves the result.

5 Continuity

Definition 25. Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let $c \in D$. The function f is said to be *continuous at* c if

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

Definition 26 (Equivalent Definition of Continuity)**.** The function f is continuous at $c \in D$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - f(c)| < \varepsilon \quad \text{whenever } x \in D, |x - c| < \delta.$$

5.1 Continuity at a Point

Theorem 5.1. Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let $c \in D$. Then f is continuous at c if and only if

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

Proof. Suppose f is continuous at c. Then for a preassigned $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - f(c)| < \varepsilon \quad \forall x \in N(c, \delta) \cap D.$$

This shows that $\lim_{x \to c} f(x) = f(c)$.

Conversely, let $\lim_{x \to c} f(x) = f(c)$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - f(c)| < \varepsilon \quad \forall x \in N(c, \delta) \cap D.$$

Hence f is continuous at c. □

5.2 Continuity on a Set

Definition 27. Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. f is said to be continuous on $A \subset D$ if f is continuous at every point of A.

Theorem 5.2 (Sequential Criterion for Continuity)**.** Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ be a function. Let $c \in D$. Then f is continuous at c if and only if for every sequence (x_n) in D converging to c, we have

$$\lim_{n \to \infty} f(x_n) = f(c).$$

Example 5.1. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = x$. Then for $c \in \mathbb{R}$,

$$|f(x) - f(c)| = |x - c|.$$

Given $\varepsilon > 0$, let us choose $\delta = \varepsilon$. Then whenever $|x - c| < \delta$, we have $|f(x) - f(c)| < \varepsilon$. Thus f is continuous at c.

Example 5.2. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \cos\frac{1}{x} & , x \neq 0 \\ 0 & , x = 0. \end{cases}$$

Prove that f is not continuous at 0.

Let us consider a sequence $\{x_n\}$ where $x_n = \frac{1}{2\pi n}$ for $n \in \mathbb{N}$. Then $\lim_{n \to \infty} x_n = 0$ and $f(x_n) = 1$ for all $n \in \mathbb{N}$. Therefore $\lim_{n \to \infty} f(x_n) = 1 \neq f(0)$.

Theorem 5.3. Let $D \subset \mathbb{R}$ and f, g be functions on D to $\mathbb{R}$. Let $c \in \mathbb{R}$ and f and g be continuous functions. Then the following hold:

1. $f + g$ is continuous at c.
2. if $k \in \mathbb{R}$ then kf is continuous at c.
3. fg is continuous at c.
4. $\frac{f}{g}$ is continuous at c if $g(x) \neq 0$ for all $x \in \mathbb{D}$.

Proof. Similar to the properties of limits. $\square$

Theorem 5.4. Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ be continuous at c. Then $|f|$ is continuous at c.

Proof. Note that $|f| : D \to \mathbb{R}$ is defined by $|f|(x) = |f(x)|$ for all $x \in D$. Observe that

$$||f(x)| - |f(c)|| \leqslant |f(x) - f(c)|. \tag{20}$$

Let us choose $\epsilon > 0$. Since f is continuous at c, then there exists $\delta > 0$ such that

$$|f(x) - f(c)| < \epsilon \text{ for all } x \in N(c, \delta) \cap D. \tag{21}$$

Thus from (20) and (21) we have

$$||f(x)| - |f(c)|| < \epsilon \text{ for all } x \in N(c, \delta) \cap D.$$

This shows that $|f|$ is continuous at c. $\square$

Note 5.1. If $|f|$ is continuous on D then f may not be continuous on D. For example, let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1 & , x \in \mathbb{Q} \\ -1 & , x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Then $|f|$ is continuous on $\mathbb{R}$ but f is not continuous on $\mathbb{R}$.

5.3 Examples of Discontinuity

Example 5.3. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 0 & x \neq 0, \\ 1 & x = 0. \end{cases}$$

Then f is not continuous at 0.

Indeed, consider the sequence $x_n = \frac{1}{n}$, then $x_n \to 0$. But $f(x_n) = 0$ for all n, hence $\lim f(x_n) = 0 \neq f(0) = 1$. Thus f is not continuous at 0.

5.4 Properties of Continuous Functions

Lemma 5.5. Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$ be a set. Then f is continuous at $c \in D$ if f is defined at c and

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

Proof. Note that $f : D \to \mathbb{R}$ is defined. Let $\varepsilon > 0$. Then there exists $\delta > 0$ such that

$$|f(x) - f(c)| < \varepsilon \quad \text{whenever } |x - c| < \delta, \ x \in D.$$

This shows that f is continuous at c. $\square$

Note 5.2. $f|_D$ is continuous on D. For example, let $f : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q}, \\ -1 & x \notin \mathbb{Q}. \end{cases}$$

Then f is not continuous on $\mathbb{R}$, but f is continuous on $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$.

Theorem 5.6. Let $A \subset \mathbb{R}$ and $f : A \to \mathbb{R}$, $g : B \to \mathbb{R}$ be functions with $f(A) \subset B$. If f is continuous at $c \in A$ and g is continuous at $f(c)$, then the composition $g \circ f : A \to \mathbb{R}$ is continuous at c.

Proof. Let U be a neighborhood of $g(f(c))$. Since g is continuous at $f(c)$, there exists a neighborhood V of $f(c)$ such that

$$g(V \cap B) \subset U.$$

Since f is continuous at c, there exists a neighborhood W of c corresponding to the neighborhood V of $f(c)$ such that

$$f(W \cap A) \subset V.$$

Thus $f(W \cap A) \subset V \cap B \implies g(f(W \cap A)) \subset g(V \cap B) \subset U$. This shows that $g \circ f$ is continuous at c. This completes the proof. ☐

5.5 Continuity of some important functions

1. **Polynomial function:** Let $p(x) = a_n x^n + \cdots + a_0$, $x \in \mathbb{R}$ and $a_i \in \mathbb{R}$ for $i = 0, 1, \ldots, n$. Then p is a polynomial and p is continuous on $\mathbb{R}$.

2. **Rational function:** Let $p(x)$ and $q(x)$ be polynomials. Let $r(x) = \frac{p(x)}{q(x)}$. If $\deg(q(x)) = m$, then by the fundamental theorem of algebra, $q(x)$ has at most m roots $\alpha_1, \ldots, \alpha_m$ in the field of complex numbers. Hence $r(x)$ is continuous on $\mathbb{R} \setminus \{\alpha_1, \ldots, \alpha_m\}$.

3. **Trigonometric functions:** Let $c \in \mathbb{R}$. Let $f(x) = \sin x$, $x \in \mathbb{R}$. Then

$$|\sin x - \sin c| = |2 \cos \tfrac{x+c}{2} \sin \tfrac{x-c}{2}| \leq 2|\sin \tfrac{x-c}{2}| \leq |x - c|.$$

   So if $|x - c| < \delta$, then $|\sin x - \sin c| < \varepsilon$ for $\delta = \varepsilon$. Thus $\sin x$ is continuous at c. Hence $\sin x$ is continuous on $\mathbb{R}$.

4. **Trigonometric functions (contd):**

   (a) Let $f(x) = \cos x$, $x \in \mathbb{R}$. Let $c \in \mathbb{R}$. Note that

$$|\cos x - \cos c| = 2\Big|\sin \tfrac{x+c}{2} \sin \tfrac{x-c}{2}\Big| \leq 2|\sin \tfrac{x-c}{2}| \leq |x - c|.$$

   Thus $\cos x$ is continuous on $\mathbb{R}$.

   (b) Let $f(x) = \tan x = \frac{\sin x}{\cos x}$. It is not defined at the points $x = \left(\frac{\pi}{2} + n\pi\right), n \in \mathbb{Z}$ where $\cos x = 0$. Hence $\tan x$ is continuous on $\mathbb{R} \setminus \left\{\frac{\pi}{2} + n\pi : n \in \mathbb{Z}\right\}$.

   (c) Let $f(x) = \cot x = \frac{\cos x}{\sin x}$. It is not defined at $x = n\pi$, $n \in \mathbb{Z}$. So $\cot x$ is continuous on $\mathbb{R} \setminus \{n\pi : n \in \mathbb{Z}\}$.

5. **Exponential function:** Let $a > 0$ and $f(x) = a^x$, $x \in \mathbb{R}$. Let $c \in \mathbb{R}$. Let $g(x)$ be any sequence in $\mathbb{R}$ such that $g \to c$. Then

$$\lim_{x \to c} a^x = a^c.$$

   This shows that f is continuous at c. Since c is arbitrary, f is continuous on $\mathbb{R}$.

   Example: $f(x) = e^x$ is continuous on $\mathbb{R}$.

6. **Logarithmic function:** Let $f(x) = \log x$, $x > 0$. Let $c > 0$. Let $g(x)$ be any sequence such that $x_n > 0$ and $x_n \to c$. Then

$$\lim_{n\to\infty} \log x_n = \log c.$$

This shows that f is continuous at c. Since c is arbitrary, f is continuous on $(0, \infty)$.

7. **Square root function:** Let $f(x) = \sqrt{x}$, $x \geq 0$. Let $c \geq 0$, $f(c) = \sqrt{c}$. When $x \geq 0$,

$$|f(x) - f(c)| = |\sqrt{x} - \sqrt{c}| = \frac{|x-c|}{\sqrt{x}+\sqrt{c}} \leq \frac{|x-c|}{\sqrt{c}} \quad (x \neq c).$$

Let $\varepsilon > 0$. Choose $\delta = \varepsilon\sqrt{c}$. Then whenever $|x - c| < \delta$ and $x \geq 0$,

$$|f(x) - f(c)| < \varepsilon.$$

Hence $f(x) = \sqrt{x}$ is continuous at $c \geq 0$. So f is continuous on $[0, \infty)$.

5.6 Discontinuities

Definition 28. Let f be defined on an interval (a, b) and let $x \in (a, b)$. We say f is discontinuous at x if f is not continuous at x.

Example 5.4. For all sequences $\{x_n\}$ in (a, b) such that $x_n \to x$, if $\lim f(x_n) \neq f(x)$, then f is discontinuous at x.

To obtain the definition at $a \leq x \leq b$, we restrict ourselves to $\{x_n\}$ in (a, x).

5.7 Types of Discontinuities

If x is a point of the domain of f at which f is not continuous, we say that f is discontinuous at x.

Definition 29. Let f be defined on (a, b). Consider a point x such that $a \leqslant x < b$. We write $f(x+) = q$ if $f(t_n) \to q$ as $n \to \infty$ for all sequence $\{t_n\}$ in (x, b) such that $t_n \to x$. To obtain the definition of $f(x-)$ for $a < x \leqslant b$, we restrict ourselves to the sequence $\{t_n\}$ in (a, x).

It is clear that at any point $x \in (a, b)$, the limit $\lim_{t\to x} f(t)$ exists if and only if $f(x+) = f(x-) = \lim_{t\to x} f(t)$.

Definition 30. Let f be defined on (a, b). If f is discontinuous at a point x and if $f(x+)$ and $f(x-)$ exists, then f is said to have a *discontinuity of first kind* or a *simple discontinuity*. Otherwise, the discontinuity is said to be of *second kind*.

There are two ways in which a function can have a simple discontinuity:

(1) $\lim_{x\to c^-} f(x)$ and $\lim_{x\to c^+} f(x)$ exist and are finite, but not equal (in which case the value $f(x)$ is immaterial).

(2) $\lim_{x\to c^-} f(x)$ and $\lim_{x\to c^+} f(x)$ exist and are finite, and equal, but different from $f(c)$.

If $\lim_{x\to c^-} f(x)$ or $\lim_{x\to c^+} f(x)$ does not exist, then f has a discontinuity of the second kind.

Example 5.5. Define
$$f(x) = \begin{cases} 1 & x \in \mathbb{Q}, \\ 0 & x \notin \mathbb{Q}. \end{cases}$$
We show that f has a discontinuity of the second kind at every point, since neither $\lim_{x\to c^-} f(x)$ nor $\lim_{x\to c^+} f(x)$ exist.

Case 1: Let c be a rational point. Let $\{x_n\}$ be a sequence of irrational points such that $x_n \to c$. Then $f(x_n) = 0$ for all n, hence $\lim f(x_n) = 0$. But $f(c) = 1$. Thus by sequential criterion f is not continuous at c.

Case 2: Let c be any irrational point. Then there exists a sequence $\{y_n\}$ of rational numbers such that $y_n \to c$. Then $f(y_n) = 1$ for all n, hence $\lim f(y_n) = 1 \neq f(c) = 0$. So f is not continuous at c.

Thus we conclude that neither $\lim_{x\to c^-} f(x)$ nor $\lim_{x\to c^+} f(x)$ exist. Hence f is not continuous at any $c \in \mathbb{R}$. This shows that f has discontinuity of the second kind at every point.

Example 5.6. Define
$$f(x) = \begin{cases} x, & x \in \mathbb{Q}, \\ 0, & x \notin \mathbb{Q}. \end{cases}$$
Show that f is continuous at 0 and has a discontinuity of the second kind at every other point.

Note that
$$|f(x) - f(0)| = |f(x)| = \begin{cases} |x|, & x \in \mathbb{Q} \\ 0, & , x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$
Let $\varepsilon > 0$. Then $|f(x) - f(0)| < \varepsilon$ whenever $|x| < \varepsilon$. Therefore f is continuous at 0. Let $c \in \mathbb{R}$ and $c \neq 0$. Choose a rational sequence $\{x_n\}$ with $x_n \to c$. Then $\lim f(x_n) = \lim x_n = c$. Choose an irrational sequence $\{y_n\}$ with $y_n \to c$. Then $\lim f(y_n) = 0$. Since $\lim f(x_n) \neq \lim f(y_n)$, the limit $\lim_{x\to c} f(x)$ does not exist. Thus f has discontinuity of the second kind at every $c \neq 0$.

Example 5.7. Define
$$f(x) = \begin{cases} x + 2, & -3 < x < -2 \\ -x - 2, & -2 \leqslant x < 0 \\ x + 2, & 0 \leqslant x < 1. \end{cases}$$
Let $\{x_n\}$ be a sequence in $(-2, 0)$ such that $y_n \to 0$. Then by definition, we have $f(x_n) \to -2$. Thus $f(0-) = -2$ Let $\{y_n\}$ be a sequence in $(0, 1)$ such that $y_n \to 0$. Then by definition, we have $f(y_n) \to 2$. This shows that $f(0+) = 2$ But $f(0-) \neq f(0+)$. This shows that f has simple discontinuity. Check that f is continuous at every other point of $(-3, 1)$.

Theorem 5.7. Let $I = [a, b]$ be a closed and bounded interval and $f : I \to \mathbb{R}$ be cts on I. Then f is bounded on I.

Proof. If possible, let f be not bounded on I. Then for each natural number n, there exists a point $x_n \in [a, b]$ such that $|f(x_n)| > n$. Thus we obtain a sequence $\{x_n\} \subset [a, b]$ such that $|f(x_n)| > n, \forall n \in \mathbb{N}$. Since $[a, b]$ is a bounded interval, $\{x_n\}$ is bounded. By Bolzano-Weierstrass theorem, there is a subsequence which is convergent. Let $\{x_{n_k}\}$ be the subsequence with limit l. Since $[a, b]$ is closed, it implies $l \in [a, b]$. This shows that f is cts on l. By sequential criteria, we have $\lim\limits_{k\to\infty} f(x_{n_k}) = f(l)$. Therefore the sequence $\{f(x_{n_k})\}$ is bounded, but by construction, $|f(x_{n_k})| > n_k$. This is a contradiction. $\square$

Theorem 5.8. Let $I = [a, b]$ and let $f : I \to \mathbb{R}$ be cts on I. Then there is a point $c, d \in [a, b]$ such that

$$f(c) = \sup_{x \in I} f(x), \quad f(d) = \inf_{x \in I} f(x).$$

Proof. The set $f(I) = \{f(x) : x \in I\}$ is bounded. Since this is a non-empty bounded subset of $\mathbb{R}$, $\sup f(I)$ and $\inf f(I)$ exist. We shall prove that $f(c) = M$ for some $c \in [a, b]$. If not, let $f(x) < M$ for all $x \in [a, b]$. Then $M - f(x) > 0$ for all $x \in [a, b]$. Let $\phi(x) = \frac{1}{M - f(x)}$, $\forall x \in [a, b]$. Then ϕ is cts on $[a, b]$ and therefore ϕ is bounded on $[a, b]$. Let B be the upper bound of ϕ on $[a, b]$. Then $B > 0$ and

$$0 < \frac{1}{M - f(x)} < B, \quad \forall x \in [a, b].$$

Therefore $f(x) < M - \frac{1}{B}$, $\forall x \in [a, b]$. This is a contradiction that $M = \sup f(I)$. Hence there is a point $c \in [a, b]$ such that $f(c) = M$. □

Similarly, one can also show that there exists a point $d \in [a, b]$ such that $f(d) = m$.

Theorem 5.9 (Bolzano)**.** Let $[a, b]$ be a closed and bounded interval and $f : [a, b] \to \mathbb{R}$ be cts on $[a, b]$. If $f(a)$ and $f(b)$ are of opposite sign, then there exists at least a point $c \in (a, b)$ such that $f(c) = 0$.

Theorem 5.10 (Intermediate Value Theorem)**.** Let $I = [a, b]$ and $f : I \to \mathbb{R}$ be cts on I. If $f(a) \neq f(b)$ then f attains every value between $f(a)$ and $f(b)$ at least once in (a, b).

Proof. WLOG, we may assume that $f(a) < f(b)$. Let M be a real number such that $f(a) < M < f(b)$. Let $\phi : [a, b] \to \mathbb{R}$ be defined by

$$\phi(x) = f(x) - M.$$

Then ϕ is cts on $[a, b]$ and $\phi(a) < 0$ and $\phi(b) > 0$. By Bolzano's theorem, there exists at least one point $c \in (a, b)$ such that $\phi(c) = 0$. Therefore $f(c) - M = 0$, that is, $f(c) = M$. This completes the proof. □

5.8 Uniform Continuity

Definition 31. Let I be any interval. A function $f : I \to \mathbb{R}$ is said to be uniformly cts on I if corresponding to a pre-assigned $\epsilon > 0$, there exists $\delta > 0$ such that for any two points $x_1, x_2 \in I$,

$$|f(x_1) - f(x_2)| < \epsilon \quad \text{whenever} \quad |x_1 - x_2| < \delta.$$

Note 5.3.

1. The definition of uniform continuity shows that uniform continuity is a property of the function on an interval (or on a set) but continuity is a property of the function at a point. Thus we conclude that continuity of a function is a local property while uniform continuity of a function is a global property.

2. It follows from the definition of uniform continuity that if a function f is uniformly cts on an interval I, then it is also uniformly cts on any subinterval $I \subset I$.

Example 5.8. Show that f defined by

$$f(x) = \frac{1}{x}, \quad x \in [1, \infty)$$

is uniformly continuous.

Solution: Let $c \geq 1$. Then for all $x \geq 1$,

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

Since $|x| \geq 1$, we have

$$|f(x) - f(c)| \leq |x - c|.$$

Let $\varepsilon > 0$. Choose $\delta = \varepsilon$. Then whenever $|x - c| < \delta$, we get

$$|f(x) - f(c)| < \varepsilon.$$

This proves that f is uniformly continuous on $[1, \infty)$.

Example 5.9. Let $f(x) = x^2$, $x \in \mathbb{R}$. Show that f is uniformly continuous on $[a, b]$, $a > 0$, but not uniformly continuous on $[a, \infty)$, $a > 0$.

Proof. Let $\varepsilon > 0$. Note that

$$|f(x_1) - f(x_2)| = |x_1^2 - x_2^2| = |x_1 + x_2||x_1 - x_2|.$$

Since $a \leq x_1, x_2 \leq b$, we have

$$|x_1 + x_2| \leq 2b.$$

Thus,

$$|f(x_1) - f(x_2)| \leq 2b|x_1 - x_2|.$$

If we choose $\delta = \frac{\varepsilon}{2b}$, then whenever $|x_1 - x_2| < \delta$, we get

$$|f(x_1) - f(x_2)| < \varepsilon.$$

Hence f is uniformly continuous on $[a, b]$. Now, consider $[a, \infty)$. Let $\varepsilon > 0$. Then for large x, $|x_1 + x_2|$ becomes arbitrarily large, so no single δ works for all pairs. Thus f is not uniformly continuous on $[a, \infty)$. $\square$

Theorem 5.11. If f is uniformly continuous on an interval I, then f is continuous on I.

Proof. Since f is uniformly continuous on I, for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|x_1 - x_2| < \delta \quad \Rightarrow \quad |f(x_1) - f(x_2)| < \varepsilon, \quad x_1, x_2 \in I.$$

Let $c \in I$. Choose $x_1 = x$, $x_2 = c$. Then whenever $|x - c| < \delta$,

$$|f(x) - f(c)| < \varepsilon.$$

This shows that f is continuous at c. Since c was arbitrary, f is continuous on I. $\square$

6 Differentiation

Definition 32. Let $I = [a, b]$ be an interval and $f : I \to \mathbb{R}$. Let c be an interior point of I. We say f is *differentiable at c* if the limit

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

exists. If the limit equals l, then l is called the derivative of f at c, denoted by $f'(c)$.

Since c is an interior point, both one-sided limits must exist and be equal:

$$\lim_{x \to c^-} \frac{f(x) - f(c)}{x - c}, \quad \lim_{x \to c^+} \frac{f(x) - f(c)}{x - c}.$$

2. If $c = a$ (a left endpoint of I), then f is differentiable at a if the right-hand limit

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

exists. This is called the right derivative of f at a, denoted $f'_+(a)$.
3. If $c = b$ (a right endpoint of I), then f is differentiable at b if the left-hand limit

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

exists. This is called the left derivative of f at b, denoted $f'_-(b)$.

Example 6.1. Let $f(x) = |x|$. At $x = 0$,

$$\lim_{x \to 0^+} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0^+} \frac{|x|}{x} = 1,$$
$$\lim_{x \to 0^-} \frac{f(x) - f(0)}{x - 0} = \lim_{x \to 0^-} \frac{|x|}{x} = -1.$$

Since the left and right derivatives are not equal, f is not differentiable at 0.

Theorem 6.1 (Chain Rule)**.** Let f be defined on an open interval (a, b), let g be defined on $f(a, b)$ and consider the composition function $g \circ f$ on (a, b) by the equation $(g \circ f)(x) = g(f(x))$. Assume there is a point c in (a, b) such that $f(c)$ is an interior point of $f(a, b)$. If f is differentiable at $f(c)$ then $g \circ f$ is differentiable at c and we have

$$(g \circ f)^{'}(c) = g^{'}(f(c))f^{'}(c).$$

Theorem 6.2. Let f be defined on (a, b) and assume that for some c in (a, b) we have $f^{'}(c) > 0$ or $f^{'}(c) = \infty$. Then there is a neighbourhood $N_\delta(c) \subset (a, b)$ in which

$$f(x) > f(c) \text{ if } x > c \text{ and } f(x) < f(c) \text{ if } x < c.$$

Proof. If $f^{'}(c)$ is finite and positive we can write $f(x) - f(c) = (x - c)f^*(c)$ where f^* is continuous at c and $f^*(c) = f^{'}(c) > 0$. By the sign preserving property of continuous functions, there exists a neighbourhood $N_\delta(c) \subset (a, b)$ in which $f^*(x)$ has same sign as of $f^*(c)$ and this means that $f(x) - f(c)$ has the same sign as of $x - c$.

If $f^{'}(c) = \infty$, then there exists a neighbourhood $N_\delta(c)$ in which $\frac{f(x) - f(c)}{x - c} > 1$ where $x \neq 0$. In this neighbourhood the quotient is again positive and the conclusion follows as before. $\square$

A similar to above theorem holds if $f^{'}(c) < 0$ or $f^{'}(c) = -\infty$ at some interior point of c of (a, b).

6.1 Local Extrema

Definition 33. Let f be a real-valued function defined on a set $S \subseteq \mathbb{R}$. Then f is said to have a *local maximum* at $a \in S$ (respectively, a *local minimum* at a) if there exists a neighbourhood $N_S(a)$ such that

$$f(x) \leq f(a) \quad \text{(respectively, } f(x) \geq f(a)) \quad \forall x \in N_S(a) \cap S.$$

Theorem 6.3. Let f be defined on (a, b) and assume that f has a local maximum or a local minimum at an interior point $c \in (a, b)$. If f has a derivative (finite) at c, then $f'(c) = 0$.

Proof. If $f'(c)$ is positive or negative, then f cannot have a local extremum at c, because the function would be increasing or decreasing near c. Similarly, $f'(c)$ cannot be $\pm\infty$. However, since f has a derivative at c, the only remaining possibility is $f'(c) = 0$. $\square$

Remark. The converse of the above theorem is not true. The condition $f'(c) = 0$ alone is not enough to guarantee that f has an extremum at c. For example, for $f(x) = x^3$, we have $f'(0) = 0$ but f is increasing in every neighbourhood of 0.

6.2 Rolle's Theorem

Theorem 6.4 (Rolle's Theorem)**.** Assume that f has a finite derivative at each point of the open interval (a, b) and is continuous on the closed interval $[a, b]$. If $f(a) = f(b)$, then there exists at least one point $c \in (a, b)$ such that $f'(c) = 0$.

Proof. Assume that f' is never 0 on (a, b). We shall obtain a contradiction. Since f is continuous on $[a, b]$, it attains its maximum M and minimum m somewhere in $[a, b]$.

Neither extreme value can be attained at an interior point (otherwise f' would vanish there), so both must be attained at the endpoints. Since $f(a) = f(b)$, this implies that $M = m$, and hence f is constant on $[a, b]$. This is a contradiction. Therefore, there must exist at least one point $c \in (a, b)$ such that $f'(c) = 0$. $\square$

6.3 Mean Value Theorem

Theorem 6.5 (Lagrange's Mean Value Theorem (LMVT))**.** Suppose that a function f is continuous on $[a, b]$ and differentiable on (a, b). Then there exists a point $c \in (a, b)$ such that

$$f(b) - f(a) = f'(c)(b - a).$$

Proof. Let $\phi : [a, b] \to \mathbb{R}$ be defined by

$$\phi(x) = f(x) + Ax, \quad x \in [a, b],$$

where A is a constant.

Since f is continuous on $[a, b]$ and differentiable on (a, b), the function ϕ is also continuous and differentiable on the same intervals. We choose A such that $\phi(a) = \phi(b)$:

$$f(a) + Aa = f(b) + Ab.$$

This gives

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

For this choice of A, the function ϕ satisfies all the required conditions for Rolle's Theorem, and hence there exists some $c \in (a, b)$ such that $\phi'(c) = 0$. That is,

$$f'(c) + A = 0 \quad \Rightarrow \quad f'(c) = \frac{f(b) - f(a)}{b - a}.$$

This completes the proof. $\square$

Note 6.1 (Geometric Interpretation)**.** The geometric view of the Mean Value Theorem is that there is some point on the curve $y = f(x)$ at which the tangent line is parallel to the secant line joining the points $(a, f(a))$ and $(b, f(b))$.

Consider two distinct points $(a, f(a))$ and $(b, f(b))$. The line connecting these points is the *secant* of the curve, which is parallel to the tangent line at some point $(c, f(c))$ on the curve.

The slope of the secant joining these points is equal to the slope of the tangent at the point $(c, f(c))$. Thus,

$$\text{slope of the tangent at } c = \text{slope of the secant} = \frac{f(b) - f(a)}{b - a}.$$

Theorem 6.6 (Cauchy's Mean Value Theorem (CMVT))**.** Let $f, g : [a, b] \to \mathbb{R}$ be differentiable such that $g^{'}(x) \neq 0$ for all $x \in [a, b]$. Then there exists $c \in (a, b)$ such that

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

Proof. Exercise! $\square$

Note 6.2 (Geometric Interpretation)**.** Given two functions f and g which are continuous on $[a, b]$ differentiable on (a, b), if the slope of the line joining $(a, f(a))$ and $(b, f(b))$ is equal to the slope of the line joining $(a, g(a))$ and $(b, g(b))$, then is there a point c in (a, b) such that at that point, both f and g have the same derivative?

Not only does the CMVT say that this is true but it also generalizes this idea when the slope of the two lines are not equal.

In general, when the slope of the two lines are not equal, we can multiply the function g by a constant r so that slope of the line joining $(a, f(a))$ and $(b, f(b))$ is equal to the slope of the line joining $(a, rg(a))$ and $(b, rg(b))$. This constant r is precisely $\frac{f(b)-f(a)}{g(b)-g(a)}$.

Now since the slopes of the two lines are equal, the lines are parallel. So, the distance between $(a, f(a))$ and $(a, rg(a))$ is equal to the distance between $(b, f(b))$ and $(b, rg(b))$. Now, if you consider the function $h(x) = f(x) - rg(x)$ on $[a, b]$, which is essentially the distance between the points $f(x)$ and $rg(x)$, we will get that $h(a) = h(b)$. Since h is a differentiable function, we can apply Rolle's theorem and say that there is a point c such that $h^{'}(c) = 0$ i.e. $f^{'}(c) = rg^{'}(c) = \frac{f(b)-f(a)}{g(b)-g(a)}$.

So, geometrically CMVT says that given two functions f and g, which are continuous on $[a, b]$ and differentiable on (a, b), we can find a point c such that the derivative of f and of a particularly 'stretched' or 'shrunk' g at c is the same.

7 Riemann Integral

Definition 34 (Partition)**.** Let $[a, b]$ be a closed and bounded interval. A partition P of $[a, b]$ is a finite ordered set

$$P = \{x_0, x_1, \ldots, x_n\} \quad \text{with } a = x_0 < x_1 < \cdots < x_n = b.$$

The family of all such partitions of $[a, b]$ is denoted by $\mathcal{P}([a, b])$.

7.1 Riemann Integrability

Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Consider a partition $P = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$. Since f is bounded on $[a, b]$, define

$$M = \sup_{x \in [a,b]} f(x), \quad m = \inf_{x \in [a,b]} f(x),$$

and for each subinterval $[x_{r-1}, x_r]$, define

$$M_r = \sup_{x \in [x_{r-1}, x_r]} f(x), \quad m_r = \inf_{x \in [x_{r-1}, x_r]} f(x), \quad \text{for } r = 1, 2, \ldots, n.$$

Then $m \leq m_r \leq M_r \leq M$ for all r. The sum

$$U(f, P) = \sum_{r=1}^{n} M_r(x_r - x_{r-1})$$

is called the *upper sum* of f with respect to partition P, and

$$L(f, P) = \sum_{r=1}^{n} m_r(x_r - x_{r-1})$$

is called the *lower sum*. Note that

$$m(x_r - x_{r-1}) \leq m_r(x_r - x_{r-1}) \leq M_r(x_r - x_{r-1}) \leq M(x_r - x_{r-1})$$

Thus

$$m(b - a) \leq L(f, P) \leq U(f, P) \leq M(b - a).$$

Definition 35. The supremum of the set $\{L(P, f) : P \in \mathcal{P}([a, b])\}$ exists and it is called the *lower integral* of f on $[a, b]$ and is denoted by

$$\underline{\int_a^b} f(x)\, dx.$$

The infimum of the set $\{U(P, f) : P \in \mathcal{P}([a, b])\}$ exists and it is called the *upper integral* of f on $[a, b]$ and is denoted by

$$\overline{\int_a^b} f(x)\, dx.$$

A function f is said to be *Riemann integrable* on $[a, b]$ if

$$\underline{\int_a^b} f = \overline{\int_a^b} f$$

and we denote it by $\int_a^b f$.

Example 7.1. Let $f : [a, b] \to \mathbb{R}$ be defined by $f(x) = c$, where $x \in [a, b]$. Prove that $f \in \mathcal{R}[a, b]$. Let f be constant on $[a, b]$. Let us take a partition P of $[a, b]$ defined by $P = \{x_0, x_1, \ldots, x_n\}$, where $a = x_0 < x_1 < \cdots < x_n = b$. Let

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x), \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x).$$

Then $M_i = c = m_i \Rightarrow M_i = m_i$. Now,

$$U(P, f) = M_1(x_1 - x_0) + \cdots + M_n(x_n - x_{n-1}) = c(b - a),$$

and similarly,

$$L(P, f) = c(b - a).$$

This is true for all partitions P of $[a, b]$. Thus,

$$\underline{\int_a^b} f = \overline{\int_a^b} f = c(b - a).$$

Example 7.2 (Dirichlet's Function)**.** Let $f : [0, 1] \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1, & \text{if } x \text{ is rational,} \\ 0, & \text{if } x \text{ is irrational.} \end{cases}$$

Show that f is not Riemann integrable on $[0, 1]$. Let us take a partition P of $[0, 1]$ defined by $P = \{x_0, x_1, \dots, x_n\}$, where $0 = x_0 < x_1 < \cdots < x_n = 1$. Then

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x), \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x).$$

Since rationals and irrationals are both dense in $\mathbb{R}$, in any interval $[x_{i-1}, x_i]$, we have:

$$M_i = 1, \quad m_i = 0 \quad \forall i = 1, 2, \dots, n.$$

Thus,

$$U(P, f) = M_1(x_1 - x_0) + \cdots + M_n(x_n - x_{n-1}) = 1,$$
$$L(P, f) = m_1(x_1 - x_0) + \cdots + m_n(x_n - x_{n-1}) = 0.$$

Therefore,

$$\underline{\int_0^1} f \neq \overline{\int_0^1} f \Rightarrow f \notin \mathcal{R}[0, 1].$$

7.2 Refinement and Condition of Integrability

Definition 36 (Refinement of a Partition)**.** Let $[a, b]$ be a closed and bounded interval and let $P = \{x_0, x_1, \dots, x_n\}$ be a partition of $[a, b]$. A partition Q of $[a, b]$ is said to be a *refinement* of P if $P \subseteq Q$.

Lemma 7.1. Let $f : [a, b] \to \mathbb{R}$ be bounded on $[a, b]$. If Q is any refinement of P, then

$$U(P, f) \geq U(Q, f) \quad \text{and} \quad L(P, f) \leq L(Q, f).$$

Proof. If Q is a refinement of P, then Q can be obtained from P by adjoining a finite number of additional points to P, one at a time. By repeating this argument finitely many times, we obtain

$$U(P, f) \geq U(Q, f), \quad L(P, f) \leq L(Q, f).$$

□

Theorem 7.2 (Condition of Integrability)**.** Let f be bounded on $[a,b]$. Then f is integrable on $[a,b]$ if and only if for each $\varepsilon > 0$, there exists a partition P of $[a,b]$ such that

$$U(P,f) - L(P,f) < \varepsilon.$$

In that case,

$$\int_a^b f = \underline{\int_a^b} f = \overline{\int_a^b} f.$$

Proof. Since

$$\overline{\int_a^b} f = \inf\{U(P,f) : P \in \mathcal{P}[a,b]\},$$

there exists a partition P'' of $[a,b]$ such that

$$\overline{\int_a^b} f \leq U(P'',f) < \overline{\int_a^b} f + \frac{\varepsilon}{2}.$$

Let $P' = P \cup P''$. Then $L(P',f) \leq L(P,f)$ and $U(P,f) \leq U(P',f)$. Thus,

$$\int_a^b f - \frac{\varepsilon}{2} < L(P',f) \leq L(P,f) \leq U(P,f) \leq U(P',f) < \int_a^b f + \frac{\varepsilon}{2}.$$

Hence,

$$U(P,f) - L(P,f) < \varepsilon.$$

Since $\underline{\int_a^b} f = \overline{\int_a^b} f$, the function f is integrable.

To prove the converse, we first observe that for any partition P of $[a,b]$,

$$L(P,f) \leq \underline{\int_a^b} f \leq \overline{\int_a^b} f \leq U(P,f).$$

Therefore,

$$\overline{\int_a^b} f - \underline{\int_a^b} f \leq U(P,f) - L(P,f).$$

□

Theorem 7.3. Let $f : [a,b] \to \mathbb{R}$ be a monotone bounded function on $[a,b]$. Then f is integrable on $[a,b]$.

Proof. Let f be monotone increasing on $[a,b]$. Thus $f(a)$ is the lower bound and $f(b)$ is the upper bound of f on $[a,b]$. Let $P = (x_0, x_1, \ldots, x_n)$ be a partition of $[a,b]$ with $||P|| < \frac{\epsilon}{f(b)-f(a)}$. Let $M_r = \sup f$ and $m_r = \inf f$ for $r = 1, 2, \ldots, n$. Then

$$\begin{aligned}
U(P,f) - L(P,f) &= \sum_{r=1}^{n} (M_r - m_r)(x_r - x_{r-1}) \\
&= \sum_{r=1}^{n} (f(x'_r) - f(x''_r))(x_r - x_{r-1}) \\
&\leq \|P\| \sum_{r=1}^{n} |f(x'_r) - f(x''_r)| \\
&\leq \|P\|(f(b) - f(a)) < \varepsilon.
\end{aligned}$$

Therefore, for a chosen $\varepsilon > 0$, there exists a partition P of $[a, b]$ such that

$$U(P, f) - L(P, f) < \varepsilon.$$

Similarly, we can prove it if f is monotone decreasing. $\square$

Theorem 7.4. If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$, then f is integrable on $[a, b]$.

Proof. Since f is continuous on $[a, b]$, it is bounded on $[a, b]$. As f is continuous on the closed interval $[a, b]$, it is uniformly continuous on $[a, b]$. Hence, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that whenever $|x - y| < \delta$, we have $|f(x) - f(y)| < \frac{\varepsilon}{b-a}$. Choosing a partition P of $[a, b]$ with $\|P\| < \delta$, we obtain

$$U(P, f) - L(P, f) < \varepsilon.$$

Hence, f is Riemann integrable on $[a, b]$. $\square$

7.3 Fundamental Theorem

Let $f : [a, b] \to \mathbb{R}$ be integrable on $[a, b]$. Then for each $x \in [a, b]$, f is integrable on $[a, x]$, so $\int_a^x f(t)\,dt$ exists and it depends on x. We can define a function F on $[a, b]$ by

$$F(x) = \int_a^x f(t)\,dt.$$

Theorem 7.5. If $f : [a, b] \to \mathbb{R}$ be integrable on $[a, b]$ then the function F defined by $F(x) = \int_a^x f(t)\,dt, x \in [a, b]$ is continuous on $[a, b]$.

Proof. Let x_1, x_2 be any two points in $[a, b]$. Note that

$$|F(x_2) - F(x_1)| = \left| \int_a^{x_2} f(t)\,dt - \int_a^{x_1} f(t)\,dt \right| = \left| \int_{x_1}^{x_2} f(t)\,dt \right|.$$

Since f is integrable, f is bounded on $[a, b]$. Therefore there exists a real number $k > 0$ such that $|f(x)| < k$ for all $x \in [a, b]$. Thus

$$|F(x_2) - F(x_1)| \leqslant k|x_2 - x_1|.$$

Let us choose $\epsilon > 0$. Then for $\delta = \epsilon/k$ we have $|F(x_2) - F(x_1)| < \epsilon$ for all x_1, x_2 satisfying $|x_2 - x_1| < \delta$. This shows that F is uniformly continuous on $[a, b]$ and hence F is continuous. This completes the proof. $\square$

Definition 37. A function ϕ is called an *antiderivative* or *primitive* of a function f on an interval I, if $\phi^{'}(x) = f(x)$ for all $x \in I$.

It is interesting to observe that if ϕ is an antinderivative of f then $\phi + c$ is also an antiderivative of f for any $c \in \mathbb{R}$. Therefore if f admits an antiderivative then it admits infinitely many antiderivative.

Theorem 7.6. If $f : [a, b] \to \mathbb{R}$ is continuous on $[a, b]$ and $\phi : [a, b] \to \mathbb{R}$ be an antiderivative of f on $[a, b]$, then

$$\int_a^b f = \phi(b) - \phi(a).$$

Proof. Since f is continuous on $[a,b]$, f is integrable on $[a,b]$. Let $F(x) = \int_a^x f(t)dt$, $x \in [a,b]$. Since f is continuous on $[a,b]$, F is differentiable on $[a,b]$, and $F'(x) = f(x)$ for all $x \in [a,b]$. So F is an antiderivative of f on $[a,b]$.

Since ϕ is an antiderivative of f on $[a,b]$, for all $x \in [a,b]$, $\phi(x) = F(x) + c$, where c is a constant. So $\phi(a) = F(a) + c = c$, since $F(a) = 0$. Therefore $\phi(x) = F(x) + \phi(a)$, for all $x \in [a,b]$.

Consequently,

$$\int_a^b f = F(b) = \phi(b) - \phi(a).$$

This completes the proof. $\square$

Note. The theorem states that if f be *continuous* on $[a,b]$ then the integral $\int_a^b f$ can be evaluated in terms of an antiderivative of f on $[a,b]$.

Theorem 7.7 (Fundamental Theorem of Integral Calculus)**.** Let $f : [a,b] \to \mathbb{R}$ be a function such that

(i) f is integrable on $[a,b]$,

(ii) f possesses an antiderivative ϕ on $[a,b]$,

then

$$\int_a^b f = \phi(b) - \phi(a).$$

Proof. Let $P = (x_0, x_1, \ldots, x_n)$ where $a = x_0 < x_1 < \cdots < x_n = b$ be a partition of $[a,b]$. Let $M_r = \sup_{x \in [x_{r-1}, x_r]} f(x)$, $m_r = \inf_{x \in [x_{r-1}, x_r]} f(x)$, for $r = 1, 2, \ldots, n$. Since $\phi'(x) = f(x)$ for all $x \in [a,b]$, ϕ satisfies all conditions of Lagrange's Mean Value Theorem on $[x_{r-1}, x_r]$, for $r = 1, \ldots, n$. Therefore for $r = 1, \ldots, n$,

$$\phi(x_r) - \phi(x_{r-1}) = \phi'(\xi_r)(x_r - x_{r-1}) = f(\xi_r)(x_r - x_{r-1}) \quad \text{for some } \xi_r \in (x_{r-1}, x_r).$$

The summation gives

$$\sum_{r=1}^{n} f(\xi_r)(x_r - x_{r-1}) = \phi(b) - \phi(a).$$

But $m_r \le f(\xi_r) \le M_r$ for $r = 1, 2, \ldots, n$.

Therefore

$$\sum_{r=1}^{n} m_r(x_r - x_{r-1}) \le \phi(b) - \phi(a) \le \sum_{r=1}^{n} M_r(x_r - x_{r-1}).$$

and this implies that

$$L(P,f) \le \phi(b) - \phi(a) \le U(P,f).$$

This holds for all partitions P of $[a,b]$. So $\phi(b) - \phi(a)$ is an upper bound of the set $\{L(P,f) : P \in \mathcal{P}[a,b]\}$. As the supremum of the set is $\underline{\int_a^b} f$, it follows that

$$\underline{\int_a^b} f \le \phi(b) - \phi(a) \quad \cdots \text{(i)}$$

Also, $\phi(b)-\phi(a)$ is a lower bound of the set $\{U(P,f) : P \in \mathcal{P}[a,b]\}$. As the infimum of the set is $\overline{\int_a^b} f$, it follows that

$$\overline{\int_a^b} f \geq \phi(b)-\phi(a) \quad \cdots \text{(ii)}$$

Hence from (i) and (ii),

$$\underline{\int_a^b} f \leq \phi(b)-\phi(a) \leq \overline{\int_a^b} f.$$

Since f is integrable on $[a,b]$, $\underline{\int_a^b} f = \overline{\int_a^b} f = \int_a^b f$.

Consequently,

$$\int_a^b f = \phi(b)-\phi(a).$$

This finishes the proof. $\square$

8 Differential Calculus of Several Variables

8.1 Regions in the Plane

Definition 38 (Region in the Plane)**.** Let D be a subset of the plane $\mathbb{R}^2$ and let $(a,b) \in \mathbb{R}^2$ be any point. An ε-*disk* around (a,b) is the set of all points $(x,y) \in \mathbb{R}^2$ whose distance from (a,b) is less than ε.

$$\text{That is, } D_\varepsilon(a,b) = \{(x,y) \in \mathbb{R}^2 : \sqrt{(x-a)^2+(y-b)^2} < \varepsilon\}.$$

(1) A point (a,b) is an *interior point* of D if and only if some ε-disk around (a,b) is contained in D.

(2) A point $(a,b) \in D$ is an *isolated point* of D if and only if (a,b) is the only point of D that is contained in some ε-disk around (a,b).

(3) A point (a,b) is a *boundary point* of D if and only if every ε-disk around (a,b) contains points from D and points not from D.

(4) A set R is an *open subset* of $\mathbb{R}^2$ if and only if all points of D are its interior points.

(5) A set D is a *closed subset* of $\mathbb{R}^2$ if and only if it contains all its boundary points.

(6) The set $\overline{D}$ denotes the set of boundary points of D; it is the *closure* of D.

(7) A set D is a *bounded subset* of $\mathbb{R}^2$ if and only if D is contained in some ε-disk (around some point).

8.2 Level Curves and Level Surfaces

Definition 39 (Level Curves and Surfaces)**.** A *level curve* of f is the set of points (x, y) in the domain of f for which $f(x, y)$ is constant, i.e.,

$$f(x, y) = c, \text{ where } c \text{ is constant in the range of } f.$$

The level curve is the projection of the contour curve on the xy-plane (with the same constant value c). Similarly, for a function $f(x, y, z)$ of three variables, the *level surfaces* are the sets of points (x, y, z) such that

$$f(x, y, z) = c,$$

for values c in the range of f.

Example 8.1. Consider the function

$$f(x, y) = 100 - x^2 - y^2.$$

Its domain is $\mathbb{R}^2$ and its range is the interval $(-\infty, 100]$. The level curve $f(x, y) = 0$ is

$$\{(x, y) : x^2 + y^2 = 100\}.$$

The level curve $f(x, y) = 51$ is

$$\{(x, y) : x^2 + y^2 = 49\}.$$

8.3 Limit of a Function of Two Variables

Definition 40 (Limit of a Function)**.** Let $f : D \to \mathbb{R}$ be a function, where D is a region in the plane. Let $(a, b) \in \overline{D}$. The limit of $f(x, y)$ as (x, y) approaches (a, b) is L if and only if corresponding to each $\varepsilon > 0$, there exists $\delta > 0$ such that for all $(x, y) \in D$ with

$$0 < \sqrt{(x-a)^2 + (y-b)^2} < \delta,$$

we have

$$|f(x, y) - L| < \varepsilon.$$

In this case, we write

$$\lim_{(x,y)\to(a,b)} f(x, y) = L.$$

If no real number L satisfies this condition, then the *limit of f at (a, b) does not exist.*

Intuitive Understanding of Limits: The intuitive understanding of the notion of limit is as follows: The distance between $f(x, y)$ and L can be made arbitrarily small by making the distance between (x, y) and (a, b) sufficiently small but not necessarily zero. It is often difficult to show that limit of a function does not exist at a point. We will come back to this question soon.

When the limit exists, we write it in many alternative ways:

The limit of $f(x, y)$ as (x, y) approaches (a, b) is L.

$$f(x, y) \to L \quad \text{as } (x, y) \to (a, b).$$

$$\lim_{(x,y)\to(a,b)} f(x, y) = L.$$

Example 8.2. Determine if $\lim_{(x,y)\to(0,0)} \frac{4xy^2}{x^2+y^2}$ exists.

Observe that the region D of f is $\mathbb{R}^2 \setminus \{(0,0)\}$. And $f(0,0)$ is not given, for $(x,y) \neq (0,0)$ it is defined. We guess that if the limit exists, it would be 0. To see that it is the case, we start with any $\varepsilon > 0$. We want to choose a $\delta > 0$ such that the following sentence becomes true:
If $0 < \sqrt{x^2+y^2} < \delta$, then

$$\left|\frac{4xy^2}{x^2+y^2}\right| < \varepsilon.$$

Since $|x| < \sqrt{x^2+y^2}$ and $y^2 < x^2+y^2$, we have

$$\left|\frac{4xy^2}{x^2+y^2}\right| \leq 4|x|\frac{y^2}{x^2+y^2} \leq 4|x| < 4\sqrt{x^2+y^2} < 4\delta.$$

So, we choose $\delta = \varepsilon/4$. Then indeed

$$\left|\frac{4xy^2}{x^2+y^2}\right| < \varepsilon.$$

Hence

$$\lim_{(x,y)\to(0,0)} \frac{4xy^2}{x^2+y^2} = 0.$$

Observation: Suppose we have obtained a δ corresponding to some ε. If we take ε_1 which is larger than the earlier ε, then the same δ will satisfy the requirement in the definition of limit. Similarly, if we choose another $\delta_1 \leq \delta$, the limit requirement is also satisfied. Thus, while showing that the limit of a function is such and such at a point, we are free to choose a pre-assigned upper bound for our ε.

Similarly, suppose for some $\epsilon > 0$, we have already obtained a δ such that the limit requirement is satisfied. If we choose another δ, say δ_1, which is smaller than δ, then the limit requirement is also satisfied. Thus, we are free to choose a pre-assigned upper bound for our δ provided it is convenient to us and it works.

Example 8.3. Consider $f(x,y) = \sqrt{1-x^2-y^2}$ where $D = \{(x,y) : x^2+y^2 \leq 1\}$. We guess that $\lim_{(x,y)\to(0,0)} f(x,y) = 1$. To show that the guess is right, let $\varepsilon > 0$. Observe that $0 \leq f(x,y) \leq 1$. Using our observation, assume that $0 < \varepsilon < 1$. Choose $\delta = \sqrt{1-(1-\varepsilon)^2}$. Then whenever $\sqrt{x^2+y^2} < \delta$, we have:

$$|f(x,y) - 1| = |\sqrt{1-(x^2+y^2)} - 1| < \varepsilon.$$

That is, $f(x,y) \to 1$ as $(x,y) \to (0,0)$.

Theorem 8.1 (Uniqueness of Limit)**.** Let $f(x,y)$ be a real-valued function defined on a region $D \subset \mathbb{R}^2$. Let $(a,b) \in \overline{D}$. If $\lim_{(x,y)\to(a,b)} f(x,y)$ exists, then it is unique.

Proof. Suppose $f(x,y) \to l$ and also $f(x,y) \to m$ as $(x,y) \to (a,b)$. Let $\varepsilon > 0$. Then there exist $\delta_1, \delta_2 > 0$ such that

$$0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_1 \Rightarrow |f(x,y) - l| < \varepsilon/2,$$

and

$$0 < \sqrt{(x-a)^2+(y-b)^2} < \delta_2 \Rightarrow |f(x,y) - m| < \varepsilon/2.$$

Choose a point (α,β) so that $0 < \sqrt{(\alpha-a)^2+(\beta-b)^2} < \min(\delta_1,\delta_2)$. Then

$$|l - m| \leq |l - f(\alpha,\beta)| + |f(\alpha,\beta) - m| < \varepsilon.$$

Hence $l = m$. □

Theorem 8.2. Suppose that $f(x, y) \to L_1$ as $(x, y) \to (a, b)$ along a path C_1 and $f(x, y) \to L_2$ as $(x, y) \to (a, b)$ along a path C_2. If $L_1 \neq L_2$, then the limit of $f(x, y)$ as $(x, y) \to (a, b)$ does not exist.

Example 8.4. Consider $f(x, y) = \dfrac{x^2 - y^2}{x^2 + y^2}$ for $(x, y) \neq (0, 0)$. What is its limit at $(0, 0)$?
When $y = 0$, $\lim\limits_{x \to 0} f(x, 0) = \lim\limits_{x \to 0} \dfrac{x^2}{x^2} = 1$. That is, $f(x, y) \to 1$ as $(x, y) \to (0, 0)$ along the x-axis.
When $x = 0$, $\lim\limits_{y \to 0} f(0, y) = \lim\limits_{y \to 0} \dfrac{-y^2}{y^2} = -1$. That is, $f(x, y) \to -1$ as $(x, y) \to (0, 0)$ along the y-axis. Hence,

$$\lim_{(x,y) \to (0,0)} f(x, y) \text{ does not exist.}$$

Example 8.5. Consider $f(x, y) = \dfrac{xy}{x^2 + y^2}$ for $(x, y) \neq (0, 0)$. What is its limit at $(0, 0)$?
Along the x-axis, $y = 0$; then $f(x, 0) = 0$. Hence, $\lim\limits_{x \to 0} f(x, 0) = 0$.
Along the y-axis, $x = 0$; then $f(0, y) = 0$. Hence, $\lim\limits_{y \to 0} f(0, y) = 0$.
Does it say that $\lim_{(x,y) \to (0,0)} f(x, y) = 0$? No.
Along the line $y = x$, $f(x, x) = \dfrac{x^2}{2x^2} = \dfrac{1}{2}$. Hence, $\lim\limits_{(x,y) \to (0,0)} f(x, y)$ does not exist.

Example 8.6. Consider $f(x, y) = \dfrac{x^2 y^2}{x^2 + y^2}$ for $(x, y) \neq (0, 0)$. What is its limit at $(0, 0)$?
If $y = mx$ for some $m \in \mathbb{R}$, then

$$f(x, mx) = \frac{x^4 m^2}{x^2(1 + m^2)} = \frac{m^2 x^2}{1 + m^2}.$$

Hence, along all straight lines, $\lim\limits_{x \to 0} f(x, mx) = 0$.
If $x = y^2$, $y \neq 0$, then

$$f(y^2, y) = \frac{y^4 y^2}{y^4 + y^2} = \frac{y^6}{y^4 + y^2} = \frac{y^4}{y^2 + 1} \to 0 \text{ as } y \to 0.$$

Hence, $\lim\limits_{(x,y) \to (0,0)} f(x, y) = 0$.

Theorem 8.3. Let $L, M, c \in \mathbb{R}$; $\lim\limits_{(x,y) \to (a,b)} f(x, y) = L$, and $\lim\limits_{(x,y) \to (a,b)} g(x, y) = M$. Then

1. Constant Multiple: $\lim\limits_{(x,y) \to (a,b)} cf(x, y) = cL$.

2. Sum: $\lim\limits_{(x,y) \to (a,b)} [f(x, y) + g(x, y)] = L + M$.

3. Product: $\lim\limits_{(x,y) \to (a,b)} [f(x, y)g(x, y)] = LM$.

4. Quotient: If $M \neq 0$ and $g(x, y) \neq 0$ in an open disk around (a, b), then

$$\lim_{(x,y) \to (a,b)} \frac{f(x, y)}{g(x, y)} = \frac{L}{M}.$$

5. Power: If $r \in \mathbb{R}$, $L^r \in \mathbb{R}$, and $\lim\limits_{(x,y) \to (a,b)} f(x, y) = L$, then

$$\lim_{(x,y) \to (a,b)} [f(x, y)]^r = L^r.$$

8.4 Continuity

Let $f(x,y)$ be a real-valued function defined on a subset D of $\mathbb{R}^2$. We say that $f(x,y)$ is *continuous at a point* $(a,b) \in D$ if and only if for each $\varepsilon > 0$, there exists $\delta > 0$ such that for all points $(x,y) \in D$ with

$$\sqrt{(x-a)^2 + (y-b)^2} < \delta \Rightarrow |f(x,y) - f(a,b)| < \varepsilon.$$

Equivalently, f is continuous at (a,b) if:

1. $f(a,b)$ is well-defined, i.e., $(a,b) \in D$;
2. $\lim\limits_{(x,y)\to(a,b)} f(x,y)$ exists; and
3. $\lim\limits_{(x,y)\to(a,b)} f(x,y) = f(a,b)$.

The function $f(x,y)$ is said to be *continuous on a subset* D if and only if $f(x,y)$ is continuous at all points in the subset. Therefore, constant multiples, sums, differences, products, quotients, and rational powers of continuous functions are continuous whenever they are well defined. Polynomials in two variables are continuous functions. Rational functions (ratios of polynomials) are continuous wherever they are defined.

Example 8.7 (Example 1.9)**.**

$$f(x,y) = \begin{cases} \dfrac{3x^2y}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0) \end{cases}$$

is continuous on $\mathbb{R}^2$.
At any point other than the origin, $f(x,y)$ is a rational function; therefore, it is continuous. To see that $f(x,y)$ is continuous at the origin, let $\epsilon > 0$ be given. Take $\delta = \epsilon/3$. Assume that $\sqrt{x^2+y^2} < \delta$. Then

$$\left|\frac{3x^2y}{x^2+y^2} - f(0,0)\right| \leqslant \left|\frac{3(x^2+y^2)\sqrt{x^2+y^2}}{x^2+y^2}\right| \leqslant 3\sqrt{x^2+y^2} < 3\delta = \epsilon.$$

Therefore, $f(x,y)$ is continuous at $(0,0)$.

Example 8.8. Let $f(x,y) = \frac{x^2-y^2}{x^2+y^2}$ be defined on $D = \mathbb{R}^2 \setminus \{(0,0)\}$. Then $f(x,y)$ is continuous on D.
Now consider the function

$$g(x,y) = \begin{cases} \dfrac{x^2-y^2}{x^2+y^2}, & (x,y) \neq (0,0), \\ 0, & (x,y) = (0,0). \end{cases}$$

By Example 8.4,

$$\lim_{(x,y)\to(0,0)} g(x,y)$$

does not exist. Hence $g(x,y)$ is not continuous at $(0,0)$.

Composition of Continuous Functions: As in the single variable case, composition of continuous functions is continuous.

Theorem 8.4. Let $f : D \to \mathbb{R}$ be continuous at (a, b) with $f(a, b) = c$. Let $g : I \to \mathbb{R}$ be continuous at c, where I is some interval in $\mathbb{R}$. Then $g(f(x, y))$ from D to $\mathbb{R}$ is continuous at (a, b).

Proof is left as an exercise.

Example 8.9.

- e^{x-y} is continuous at all points in the plane.
- $\dfrac{\cos xy}{1+x^2}$ and $\ln(1 + x^2 + y^2)$ are continuous on $\mathbb{R}^2$.
- At what points is $\tan^{-1}(y/x)$ continuous?

  The function y/x is continuous everywhere except where $x = 0$. The function $\tan^{-1} t$ is continuous everywhere on $\mathbb{R}$. Hence, $\tan^{-1}(y/x)$ is continuous everywhere except where $x = 0$.
- The function $(x^2+y^2+z^2-1)^{-1}$ is continuous everywhere except on the sphere $x^2+y^2+z^2 = 1$, where it is not defined.

1.4 Partial Derivatives

Let $f(x, y)$ be a real-valued function defined on a region $D \subseteq \mathbb{R}^2$. Let $(a, b) \in D$.
If C is the curve of intersection of the surface $z = f(x, y)$ with the plane $y = b$, then the slope of the tangent line to C at $(a, b, f(a, b))$ is the partial derivative of $f(x, y)$ with respect to x at (a, b).

The *partial derivative of $f(x, y)$ with respect to x at the point (a, b)* is

$$f_x(a, b) = \frac{\partial f}{\partial x}(a, b) = \lim_{h\to 0} \frac{f(a + h, b) - f(a, b)}{h},$$

provided this limit exists. Similarly, the partial derivative with respect to y at (a, b) is

$$f_y(a, b) = \frac{\partial f}{\partial y}(a, b) = \lim_{h\to 0} \frac{f(a, b + h) - f(a, b)}{h}.$$

Example 8.10. Find $f_x(1, 1)$ where $f(x, y) = 4 - x^2 - 2y^2$.

$$f_x(1, 1) = \lim_{h\to 0} \frac{(4 - (1 + h)^2 - 2) - (4 - 1 - 2)}{h} = \lim_{h\to 0} \frac{-2h - h^2}{h} = -2.$$

Thus $f_x(1, 1) = -2$.

Example 8.11. Find f_x and f_y where $f(x, y) = y \sin(xy)$.

$$f_x(x, y) = y^2 \cos(xy), \qquad f_y(x, y) = \sin(xy) + xy \cos(xy).$$

Example 8.12. The plane $x = 1$ intersects the surface $z = x^2 + y^2$ in a parabola. Find the slope of the tangent to the parabola at the point $(1, 2, 5)$.

$$\frac{\partial z}{\partial y}(1, 2) = (2y)(1, 2) = 4.$$

Alternatively, since $z = x^2 + y^2$, at $x = 1$ we get $z = 1 + y^2$, so

$$\left.\frac{dz}{dy}\right|_{y=2} = 2y = 4.$$

For a function $f(x, y)$, partial derivatives of second order are:

$$f_{xx} = \frac{\partial^2 f}{\partial x^2}, \quad f_{xy} = \frac{\partial^2 f}{\partial y\, \partial x}, \quad f_{yy} = \frac{\partial^2 f}{\partial y^2}, \quad f_{yx} = \frac{\partial^2 f}{\partial x\, \partial y}.$$

Similarly, higher-order partial derivatives are defined, for example

$$f_{xxyy} = \frac{\partial^4 f}{\partial y^2\, \partial x^2} = \frac{\partial^4 f}{\partial x^2\, \partial y^2}.$$

Note 8.1. Observe that $f_x(a, b)$ is not the same as

$$\lim_{(x,y)\to(a,b)} f_x(x, y).$$

To see this, let

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

Then $f_x(y) = 0$ for all $x > 0$ and $f_x(y) = 0$ for $x < 0$, but

$$\lim_{(x,y)\to(0,0)} f_x(x, y)$$

does not exist.

Theorem 8.5 (Clairaut)**.** Let D be a region in $\mathbb{R}^2$. Let the function $f : D \to \mathbb{R}$ have continuous first and second-order partial derivatives on D. Then $f_{xy} = f_{yx}$.

Example 8.13. Let

$$f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

Here, $f(x, 0) = 0 = f(0, y)$. So $f_x(0, 0) = 0 = f_y(0, 0)$. And limit of $f(x, y)$ as $(x, y) \to (0, 0)$ does not exists. Hence $f(x, y)$ is not continuous at $(0, 0)$.
Further, we find that $f_{xx}(x, 0) = 0 = f_{yy}(0, y)$. What about $f_{xy}(0, 0)$?

$$f_x(0, y) = \lim_{h\to 0} \frac{f(h, y) - f(0, y)}{h} = \lim_{h\to 0} \frac{y}{h^2 + y^2} = \frac{1}{y}.$$

Thus, $f_x(0, y)$ is not continuous at $y = 0$. Notice that the second order partial derivatives $f_{xy}(0, 0)$ and $f_{yx}(0, 0)$ do not exist.

1.5 Increment Theorem

In order to see the connection between continuity of a function and the partial derivatives, the associated geometry may help. Let $z = f(x, y)$, where f_x, f_y are continuous on the region D, the domain of f. Let $(a, b) \in D$. Let C_1 and C_2 be the curves of intersection of the planes $x = a$ and $y = b$ with S (the surface $z = f(x, y)$).

Let T_1 and T_2 be tangent lines to the curves C_1 and C_2 at the point $P(a, b, f(a, b))$. The *tangent plane* to the surface S at P is the plane containing T_1 and T_2. The tangent plane to S at P consists of all possible tangent lines at P to the curves C that lie on S and pass through P. This plane approximates S at P most closely. Write the z-coordinate of P is c. Then $P = (a, b, c)$. Equation of any plane passing through P is

$$z - c = A(x - a) + B(y - b).$$

When $y = b$, the tangent plane represents the tangent to the intersection curve at P. Thus $A = f_x(a, b)$, the slope of the tangent line. Similarly, $B = f_y(a, b)$. Hence the equation of the tangent plane to the surface $z = f(x, y)$ at the point $P(a, b, c)$ is

$$z - c = f_x(a, b)(x - a) + f_y(a, b)(y - b),$$

provided that f_x, f_y are continuous at (a, b).

Example 8.14. Find the equation of the tangent plane to the elliptic paraboloid $z = 2x^2 + y^2$ at $(1, 1, 3)$.
Here, $z_x = 4x$, $z_y = 2y$. So, $z_x(1, 1) = 4$, $z_y(1, 1) = 2$. Then the equation of the tangent plane is

$$z - 3 = 4(x - 1) + 2(y - 1).$$

It simplifies to $z = 4x + 2y - 3$.

Write the equation in the linear approximation form:

$$f(x, y) \approx f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

This formula holds true for all points $(x, y, f(x, y))$ on the tangent plane at $(a, b, f(a, b))$. Writing in the increment form,

$$f(a + h, b + k) \approx f(a, b) + f_x(a, b)h + f_y(a, b)k.$$

This gives rise to the total increment

$$f(a + h, b + k) - f(a, b).$$

The total increment can be written in a more suggestive form. Towards this, write

$$\Delta f = f(a + h, b + k) - f(a + h, b) + f(a + h, b) - f(a, b).$$

By MVT, there exist $c \in [a, a + h]$ and $e \in [b, b + k]$ such that

$$f(a + h, b) - f(a, b) = h[f_x(c, b) - f_x(a, b)] + f_x(a, b)$$

and

$$f(a + h, b + k) - f(a + h, b) = k[f_y(a + h, d, f_y(a, b))] + kf_y(a, b).$$

Let $\epsilon_1 = f_x(d, b) - f_x(a, b)$ and $\epsilon_2 = f_y(a + h, c) - f_y(a, b)$. When both $h \to 0$ and $k \to 0$, we have $\epsilon_1 \to$ and $\epsilon_2 \to 0$. Hence the total increment can be written as

$$\Delta f = f_x(c, b)h + f_y(a + h, e)k.$$

Theorem 8.6 (Increment Theorem)**.** Let D be a region in $\mathbb{R}^2$. Let the function $f : D \to \mathbb{R}$ have continuous first-order partial derivatives in D. Then f is continuous on D, and the increment

$$\Delta f = f(a+h, b+k) - f(a,b)$$

can be written as

$$\Delta f = f_x(a,b)h + f_y(a,b)k + \varepsilon_1 h + \varepsilon_2 k,$$

where $\varepsilon_1 \to 0$ and $\varepsilon_2 \to 0$ as both $\Delta x \to 0$ and $\Delta y \to 0$.

Recall that for a function of one variable, its *differential* is defined as $dg = g'(x)dx$. Let $f(x,y)$ be a given function. The *differential* of f, also called the *total differential*, is

$$df = f_x(x,y)dx + f_y(x,y)dy.$$

Here, $dx = \Delta x$ and $dy = \Delta y$ are the increments in x and y, respectively. Then df is a linear approximation to the total increment Δf.

8.5 Chain Rules

We apply the increment theorem to partially differentiate composite functions.

Theorem 8.7 (Chain Rule 1)**.** Let $x(t)$ and $y(t)$ be differentiable functions. Let $f(x,y)$ have continuous first order partial derivatives. Then

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}.$$

Proof. Using the increment theorem at a point P, we obtain

$$\frac{\Delta f}{\Delta t} = \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t} + \epsilon_1\frac{\Delta x}{\Delta t} + \epsilon_2\frac{\Delta y}{\Delta t}.$$

As $\Delta t \to 0$, we have $\Delta x \to 0, \Delta y \to 0, \epsilon_1 \to 0, \epsilon_2 \to 0$. Then the result follows. $\square$

For example, if $z = xy$ and $x = \sin t, y = \cos t$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = \cos^2 t - \sin^2 t.$$

Check: $z(t) = \sin t \cos t = \frac{1}{2}\sin 2t$. So, $z'(t) = \cos 2t = \cos^2 t - \sin^2 t$.

Theorem 8.8 (Chain Rule 2)**.** Let $f(x,y,z)$ have continuous first order partial derivatives. Suppose $x = x(u,v)$ and $y = y(u,v)$ are functions such that x, y, and v are also continuous. Then

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}, \qquad \frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}.$$

Proof of this follows a similar line to that of Chain Rule 1. The pattern is clear if we use the subscript notation:

$$f_u = f_x x_u + f_y y_u, \qquad f_v = f_x x_v + f_y y_v.$$

Example 8.15. Let $z = e^x \sin y$, $x = st^2, y = s^2t$. Then

$$\frac{\partial z}{\partial s} = (e^x \sin y)t^2 + (e^x \cos y)2st = te^{st^2}(t\sin(s^2t) + 2s\cos(s^2t)).$$

$$\frac{\partial z}{\partial t} = (e^x \sin y)2st + (e^x \cos y)s^2 = se^{st^2}(2t\sin(s^2t) + 2s\cos(s^2t)).$$

Substitute expressions for x and y to get

$$z_s = 2se^{s^2+t^2}(\sin(s^2 - t^2) + \cos(s^2 - t^2)).$$

Example 8.16. Given that $z = f(x, y)$ has continuous second order partial derivatives and that $x = r^2 + s^2$, $y = 2rs$, find z_{rr}.
We have $x_r = 2r$ and $y_r = 2s$. Then

$$\begin{aligned}
z_r &= z_x x_r + z_y y_r = 2rz_x + 2sz_y.\\
z_{xr} &= z_{xx}x_r + z_{xy}y_r = 2rz_{xx} + 2sz_{xy}.\\
z_{yr} &= z_{yx}x_r + z_{yy}y_r = 2rz_{yx} + 2sz_{yy}.\\
z_{rr} &= \frac{\partial z_r}{\partial r}\\
&= 2z_x + 2rz_{xr} + 2sz_{yr}\\
&= 2z_x + 2r(2rz_{xx} + 2sz_{xy}) + 2s(2rz_{yx} + 2sz_{yy})\\
&= 2z_x + 4r^2 z_{xx} + 8rsz_{xy} + 4s^2 z_{yy}.
\end{aligned}$$

Functions can be differentiated implicitly. If F is defined within a sphere S containing a point (a, b, c), where $F(a, b, c) = 0$, $F_c \neq 0$, and F_a, F_b, F_c are continuous inside the sphere, then the equation $F(x, y, z) = 0$ defines a function $z = f(x, y)$ in a sphere containing (a, b, c) and contained in S. Moreover, the function $z = f(x, y)$ can now be differentiated partially with

$$z_x = -\frac{F_x}{F_z}, \quad z_y = -\frac{F_y}{F_z}.$$

1.7 Directional Derivative

Recall that if $f(x, y)$ is a function, then $f_x(x_0, y_0)$ is the rate of change in f with respect to change in x at (x_0, y_0), that is, in the direction $\hat{\imath}$. Similarly, $f_y(x_0, y_0)$ is the rate of change at (x_0, y_0) in the direction $\hat{\jmath}$. How do we find the rate of change of $f(x, y)$ at (x_0, y_0) in the direction of any unit vector $\hat{u}$?

Consider the surface S with the equation $z = f(x, y)$. Let $z_0 = f(x_0, y_0)$ be a point $P(x_0, y_0, z_0)$ on S. The vertical plane that passes through P in the direction of a vector $\hat{u}$ cuts S in a curve C. The slope of the tangent line T to the curve C at the point P is the rate of change of $z = f(x, y)$ in the direction of $\hat{u}$.

Let $f(x, y)$ be a function defined in a region D. Let $(x_0, y_0) \in D$. The *directional derivative* of $f(x, y)$ in the direction of a unit vector $\hat{u} = a\hat{\imath} + b\hat{\jmath}$ at (x_0, y_0) is given by

$$(D_{\hat{u}}f)(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0 + ha, y_0 + hb) - f(x_0, y_0)}{h}.$$

Example 8.17. Find the derivative of $z = x^2 + y^2$ at $(1, 2)$ in the direction of $\hat{a} = \frac{1}{\sqrt{2}}(\hat{\imath} + \hat{\jmath})$.

$$D_{\hat{a}}f(1,2) = \lim_{h\to 0} \frac{f(1+h/\sqrt{2}, 2+h/\sqrt{2}) - f(1,2)}{h} = \frac{2h/\sqrt{2} + 2h/\sqrt{2}}{h} = \frac{6}{\sqrt{2}}.$$

Hence $D_{\hat{a}}f(1,2) = 3\sqrt{2}$.

Theorem 8.9. Let $f(x,y)$ have continuous first order partial derivatives. Then $f(x,y)$ has a directional derivative at (x,y) in any direction $\hat{u} = a\hat{\imath} + b\hat{\jmath}$, and it is given by

$$D_{\hat{u}}f(x,y) = f_x(x,y)a + f_y(x,y)b.$$

Proof. Let (x_0, y_0) be a point in the domain of $f(x,y)$. Define the function $g(h) = f(x_0+ah, y_0+bh)$. Then $g(h)$ is a continuously differentiable function of h. Now

$$g'(h) = f_x(x_0+ah, y_0+bh)a + f_y(x_0+ah, y_0+bh)b.$$

Then

$$g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b.$$

Hence

$$D_{\hat{u}}f(x_0, y_0) = g'(0) = f_x(x_0, y_0)a + f_y(x_0, y_0)b.$$

This completes the proof. $\square$

Example 8.18. Find the directional derivative of $f(x,y) = x^3 - 3xy + 4y^2$ in the direction of the line that makes an angle of $\pi/6$ with the x-axis. Here, the direction is given by the unit vector $\hat{u} = \cos(\pi/6)\hat{\imath} + \sin(\pi/6)\hat{\jmath} = \frac{\sqrt{3}}{2}\hat{\imath} + \frac{1}{2}\hat{\jmath}$.

$$D_{\hat{u}}f(x,y) = \frac{\sqrt{3}}{2}f_x + \frac{1}{2}f_y = \frac{\sqrt{3}}{2}(3x^2 - 3y) + \frac{1}{2}(-3x+8y) = \frac{1}{2}[3\sqrt{3}x^2 - 3\sqrt{3}y - 3x + 8y].$$

The formula for the directional derivative in the direction of the unit vector $\hat{u} = a\hat{\imath} + b\hat{\jmath}$ can be written as

$$D_{\hat{u}}f = f_x a + f_y b = (f_x\hat{\imath} + f_y\hat{\jmath}) \cdot (a\hat{\imath} + b\hat{\jmath}).$$

The vector operator $\nabla = \frac{\partial}{\partial x}\hat{\imath} + \frac{\partial}{\partial y}\hat{\jmath}$ is called the *gradient* and the *gradient of* $f(x,y)$ is

$$\nabla f = f_x\hat{\imath} + f_y\hat{\jmath}.$$

Therefore,

$$D_{\hat{u}}f = \nabla f \cdot \hat{a}.$$

That is, at (x_0, y_0), the directional derivative is given by

$$D_{\hat{a}}f(x_0, y_0) = \nabla f(x_0, y_0) \cdot \hat{a}.$$

Example 8.19. How much does the value of $y\sin x + 2yz$ change if the point (x,y,z) moves 0.1 units from $(0,1,0)$ toward $(2,2,-2)$?

Let $f(x,y,z) = y\sin x + 2yz$. $P(0,1,0)$, $Q(2,2,-2)$. Then $\overrightarrow{PQ} = \hat{\imath} + \hat{\jmath} - 2\hat{k}$. The unit vector in the direction of $\overrightarrow{PQ}$ is $\hat{u} = \frac{1}{3}(2\hat{\imath} + \hat{\jmath} - 2\hat{k})$.

We find $D_{\hat{a}}f$ at P, which requires ∇f:

$$\nabla f = (y\cos x)\hat{\imath} + (\sin x + 2z)j + 2y\hat{k}.$$

Then

$$D_{\hat{u}}f(P) = \nabla f(P) \cdot \hat{a} = (1, 0, 2) \cdot \frac{1}{3}(2, 1, -2) = \frac{2-4}{3} = -\frac{2}{3}.$$

Thus, the change of f in the direction of $\hat{u}$ in moving 0.1 units is approximately

$$\Delta f \approx D_{\hat{u}}f(P) \times 0.1 = -\frac{2}{3} \times 0.1 = -0.067.$$

Theorem 8.10. Let $f(x, y)$ have continuous first order partial derivatives. The maximum value of the directional derivative $D_{\hat{a}}f$ is $|\nabla f|$ and it is achieved when the unit vector **a** has the same direction as that of ∇f.

Proof. Note that

$$D_{\hat{a}}f = \nabla f \cdot \hat{a} = |\nabla f||\hat{a}| \cos\theta = |\nabla f| \cos\theta,$$

where θ is the angle between ∇f and $\hat{a}$. Since the maximum of $\cos\theta$ is 1, the maximum of $D_{\hat{a}}f$ is $|\nabla f|$. The maximum is achieved when $\theta = 0$, i.e., when the directions of ∇f and $\hat{a}$ coincide. □

This also says the following:

- $f(x, y)$ increases most rapidly in the direction of its gradient.
- $f(x, y)$ decreases most rapidly in the opposite direction of its gradient.
- $f(x, y)$ remains constant in any direction orthogonal to its gradient.

8.6 Normal to Level Curve and Tangent Planes

Let $z = f(x, y)$ be a given surface. Assume that f_x and f_y are continuous. Recall that a level curve to this surface is a curve in the plane where $f(x, y)$ is constant. Suppose $f(x, y) = c$ defines a level curve. Differentiating, we have

$$f_x \frac{dx}{dt} + f_y \frac{dy}{dt} = 0.$$

Since $\frac{d\hat{r}}{dt}$ is the tangent to the curve, ∇f is the normal to the level curve. That is, the gradient ∇f is normal to the level curve that passes through (x_0, y_0).

In higher dimensions, if $f(x_1, \ldots, x_n)$ is a function of n independent variables defined on $D \subseteq \mathbb{R}^n$, then its gradient at any point is

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right).$$

The directional derivative at any point **x** in the direction of a unit vector $\hat{a} = (a_1, \ldots, a_n)$ is

$$D_{\hat{a}}f = \lim_{h\to 0} \frac{f(\hat{x} + h\hat{a}) - f(\hat{x})}{h} = \nabla f \cdot \hat{a} = f_{x_1}a_1 + \cdots + f_{x_n}a_n.$$

The algebraic rules for the gradient are as follows:

(1) Constant multiple: $\nabla(kf) = k(\nabla f), \quad k \in \mathbb{R}$.

(2) Sum: $\nabla(f + g) = \nabla f + \nabla g$.

(3) Difference: $\nabla(f - g) = \nabla f - \nabla g$.

(4) Product: $\nabla(fg) = f(\nabla g) + g(\nabla f)$.

(5) Quotient: $\nabla\left(\frac{f}{g}\right) = \frac{g(\nabla f) - f(\nabla g)}{g^2}$.

In $\mathbb{R}^3$, let

$$\vec{r}(t) = x(t)\,\hat{i} + y(t)\,\hat{j} + z(t)\,\hat{k}$$

be a smooth curve on the level surface

$$f(x, y, z) = c$$

Then

$$f(x(t), y(t), z(t)) = c \quad \text{for all } t$$

Differentiating with respect to t, we get

$$\nabla f \cdot \frac{d\vec{r}}{dt} = 0$$

Thus, the velocity vectors $\frac{d\vec{r}}{dt}$ of all such smooth curves passing through a point P on the level surface are orthogonal to the gradient ∇f at P.

Let $f(x, y, z)$ have continuous partial derivatives f_x, f_y, and f_z. The *tangent plane* at $P(x_0, y_0, z_0)$ on the level surface $f(x, y, z) = c$ is the plane through P which is orthogonal to ∇f at P. Its equation is

$$f_x(x_0, y_0, z_0)(x - x_0) + f_y(x_0, y_0, z_0)(y - y_0) + f_z(x_0, y_0, z_0)(z - z_0) = 0$$

The *normal line* to the level surface $f(x, y, z) = c$ at $P(x_0, y_0, z_0)$ is the line through P parallel to ∇f. Its parametric equations are

$$\begin{cases} x = x_0 + f_x(x_0, y_0, z_0)t \\ y = y_0 + f_y(x_0, y_0, z_0)t \\ z = z_0 + f_z(x_0, y_0, z_0)t \end{cases}$$

The equation of the tangent plane to the surface $z = f(x, y)$ at (a, b) can be obtained as follows. Write the surface as

$$F(x, y, z) = 0 \quad \text{where } F(x, y, z) = f(x, y) - z$$

Then

$$F_x = f_x, \quad F_y = f_y, \quad F_z = -1$$

The equation of the tangent plane is therefore

$$f_x(a, b)(x - a) + f_y(a, b)(y - b) - (z - f(a, b)) = 0$$

or equivalently,

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

Example 8.20. Find the tangent plane and the normal line of the surface $x^2 + y^2 + z = 9$ at the point $(1, 2, 4)$.

First, check that the point $(1, 2, 4)$ lies on the surface.

$$f(1, 2, 4) = 1^2 + 2^2 + 4 = 9.$$

Next, $f_x(1, 2, 4) = 2x = 2(1) = 2, \quad f_y(1, 2, 4) = 2y = 4, \quad f_z(1, 2, 4) = 1$. The tangent plane is given by

$$2(x - 1) + 4(y - 2) + (z - 4) = 0.$$

The normal line at $(1, 2, 4)$ is given by

$$x = 1 + 2t, \qquad y = 2 + 4t, \qquad z = 4 + t.$$

8.7 Taylor's Theorem

Theorem 8.11 (Taylor's Formula for One Variable)**.** Let $n \in \mathbb{N}$. Suppose that $f^{(n)}(x)$ is continuous on $[a, b]$ and differentiable on (a, b). Then there exists a point $c \in (a, b)$ such that

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

Proof. For $x = a$, the formula clearly holds. So, let $x \in (a, b]$. For any $t \in [a, x]$, define

$$p(t) = f(a) + f'(a)(t-a) + \frac{f''(a)}{2!}(t-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(t-a)^n.$$

Here we treat x as a fixed point and t as a variable. Now define

$$g(t) = f(t) - p(t) - \frac{f(x) - p(x)}{(x-a)^{n+1}}(t-a)^{n+1}.$$

We see that

$$g(a) = 0, \quad g(x) = 0, \quad g'(a) = 0, \quad \ldots, \quad g^{(n)}(a) = 0.$$

By Rolle's Theorem, there exists $c_1 \in (a, x)$ such that $g'(c_1) = 0$. Since $g(a) = 0$, applying Rolle's Theorem again gives a $c_2 \in (a, c_1)$ such that $g''(c_2) = 0$. Continuing this process, we obtain points

$$c_1, c_2, \ldots, c_{n+1}, \quad \text{with } c_{k+1} \in (a, c_k),$$

such that

$$g^{(n+1)}(c_{n+1}) = 0.$$

Computing $g^{(n+1)}(t)$, we find

$$g^{(n+1)}(t) = f^{(n+1)}(t) - \frac{f(x) - p(x)}{(x-a)^{n+1}}(n+1)!.$$

Setting $t = c_{n+1}$ gives

$$f(x) - p(x) = \frac{f^{(n+1)}(c_{n+1})}{(n+1)!}(x-a)^{n+1}.$$

Hence,

$$f(x) = p(x) + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1},$$

where $c = c_{n+1} \in (a, x)$. Since $p(t)$ is a polynomial of degree at most n, we have

$$p^{(n+1)}(t) = 0.$$

Then

$$g^{(n+1)}(t) = f^{(n+1)}(t) - \frac{f(x) - p(x)}{(x-a)^{n+1}}(n+1)!.$$

Evaluating at $t = c_{n+1}$, we obtain

$$f^{(n+1)}(c_{n+1}) - \frac{f(x) - p(x)}{(x-a)^{n+1}}(n+1)! = 0.$$

That is,

$$\frac{f(x)-p(x)}{(x-a)^{n+1}} = \frac{f^{(n+1)}(c_{n+1})}{(n+1)!}.$$

Consequently,

$$g(t) = f(t) - p(t) - \frac{f^{(n+1)}(c_{n+1})}{(n+1)!}(t-a)^{n+1}.$$

Evaluating at $t = x$ and using the fact that $g(x) = 0$, we get

$$f(x) = p(x) + \frac{f^{(n+1)}(c_{n+1})}{(n+1)!}(x-a)^{n+1}.$$

Since x is an arbitrary point in $(a, b]$, this completes the proof. $\square$

We have a similar result for functions of several variables.

Theorem 8.12 (Taylor's Formula for Several Variables)**.** Let $D \subset \mathbb{R}^2$ be a region, and let (a, b) be an interior point of D. Suppose $f : D \to \mathbb{R}$ has continuous partial derivatives of order up to $n + 1$ in some open disk D_0 centered at (a, b) and contained in D. Then for any point $(a + h, b + k) \in D_0$, there exists $\theta \in (0, 1)$ such that

$$\begin{aligned} f(a+h, b+k) = f(a, b) &+ \sum_{m=1}^{n} \frac{1}{m!}(h\partial_x + k\partial_y)^m f(a, b) \\ &+ \frac{1}{(n+1)!}(h\partial_x + k\partial_y)^{n+1} f(a+\theta h, b+\theta k), \end{aligned}$$

where ∂_x and ∂_y denote partial differentiation with respect to x and y, respectively.

8.8 Extreme Values

We extend the notions of local maxima and local minima to a function of two variables. Let D be a region in $\mathbb{R}^2$, (a, b) be an interior point of D, and let $f : D \to \mathbb{R}$. We say that $f(x, y)$ has a *local maximum* at (a, b) if $f(x, y) \leq f(a, b)$ for all $(x, y) \in D$ near (a, b). We say that $f(x, y)$ has a *local minimum* at (a, b) if $f(x, y) \geq f(a, b)$ for all $(x, y) \in D$ near (a, b). The number $f(a, b)$ is then called a *point of local maximum (or minimum)* value of $f(x, y)$, and the point (a, b) is called a *point of local maximum (or minimum).* We say that f has an *absolute maximum* at $(a, b) \in D$ if

$$f(x, y) \leq f(a, b) \text{ for all } (x, y) \in D.$$

Similarly, f has an *absolute minimum* at $(a, b) \in D$ if

$$f(x, y) \geq f(a, b) \text{ for all } (x, y) \in D.$$

Notice that a local extremum point must be an interior point, whereas an absolute extremum point need not be. An interior point (a, b) of D is a *critical point* of $f(x, y)$ if either

$$f_x(a, b) = f_y(a, b) = 0$$

or at least one of $f_x(a, b)$, $f_y(a, b)$ does not exist.

Theorem 8.13. Let D be a region in $\mathbb{R}^2$, $f : D \to \mathbb{R}$. Let (a, b) be an interior point of D. If $f(x, y)$ has a local extremum at (a, b), then (a, b) is a critical point of $f(x, y)$.

Proof. Suppose f has a local maximum at (a, b). Then, along the line $x = a$, the function $g(y) = f(a, y)$ has a local maximum at $y = b$. Hence $g'(b) = 0$, which gives $f_y(a, b) = 0$. Similarly, by taking $h(x) = f(x, b)$, we get $f_x(a, b) = 0$. □

Geometrically, it says that if at an interior point (a, b) there exists a tangent plane to the surface $z = f(x, y)$, and if this point happens to be an extremum, then the tangent plane is parallel to the xy-plane, i.e. $f_x(a, b) = f_y(a, b) = 0$.

Let D be a region in $\mathbb{R}^2$. Let $f : D \to \mathbb{R}$ have continuous partial derivatives f_x and f_y. Let (a, b) be a critical point of $f(x, y)$. The point $(a, b, f(a, b))$ on the surface is called a *saddle point* of f if in every open disk centered at (a, b), there are points (x_1, y_1) and (x_2, y_2) such that

$$f(x_1, y_1) < f(a, b) < f(x_2, y_2).$$

At a saddle point, the function has neither a local maximum nor a local minimum; the surface crosses its tangent plane.

For a function $f(x, y)$, its *Hessian* is defined by

$$H(f) = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - f_{xy}^2.$$

Suppose $f(x, y)$ has second-order continuous partial derivatives in an open disk centered at (a, b) inside its domain. If $H(f)(a, b) > 0$, then the surface $z = f(x, y)$ curves the same way in all directions near (a, b).

Theorem 8.14. Let $f : D \to \mathbb{R}$ have continuous first and second partial derivatives in an open disk centered at $(a, b) \in D$. Suppose (a, b) is a critical point of $f(x, y)$. Then

(1) If $H(f)(a, b) > 0$ and $f_{xx}(a, b) < 0$, then $f(x, y)$ has a local maximum at (a, b).

(2) If $H(f)(a, b) > 0$ and $f_{xx}(a, b) > 0$, then $f(x, y)$ has a local minimum at (a, b).

(3) If $H(f)(a, b) < 0$, then $f(x, y)$ has a saddle point at (a, b).

(4) If $H(f)(a, b) = 0$, then nothing can be said in general.

Proof. Let $(a + h, b + k)$ be in an open disk centered at (a, b) and contained in D. By Taylor's formula,

$$f(a + h, b + k) = f(a, b) + (hf_x + kf_y)_{(a,b)} + \frac{1}{2}\left(h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy}\right)_{(a,b)} + o(h^2 + k^2).$$

Since (a, b) is a critical point, $f_x(a, b) = f_y(a, b) = 0$, so

$$f(a + h, b + k) - f(a, b) = \frac{1}{2}\left(f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2\right)_{(a,b)} + o(h^2 + k^2).$$

Hence, the sign of $f(a + h, b + k) - f(a, b)$ depends on the quadratic form inside parentheses, which is determined by $H(f)(a, b)$ and $f_{xx}(a, b)$. □Next, set $f_x(a, b) = f_y(a, b) = 0$. Use Equation (1.1) to obtain

$$f(a + h, b + k) - f(a, b) = \tfrac{1}{2}\left(h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy}\right)_{(a,b)} + \epsilon(h, k),$$

where $\epsilon(h,k) \to 0$ as $(h,k) \to (0,0)$. Multiply both sides of Equation (1.1) by $2/h^2$, add and subtract $(f(a,b+h) - f(a,b-h))/h^2$ and rearrange to get

$$2\left[\frac{f(a+h,b+k)+f(a-h,b-k)}{2} - f(a,b)\right] = \tfrac{1}{2}H(f)(a,b)(h^2+k^2),$$

where $H(f)(a,b) = f_{xx}(a,b)f_{yy}(a,b) - f_{xy}^2(a,b)$. The RHS is positive. Therefore, by continuity of functions involved,

$$f(a+h,b+k)+f(a-h,b-k) > 2f(a,b)$$

for (h,k) in some neighborhood of $(0,0)$. That is, f has a local minimum at (a,b). We break this into three subcases:

(3A) Let $H(f)(a,b) > 0$ and $f_{xx}(a,b) > 0$. Then $f(x,y)$ has a local minimum at (a,b).

(3B) Let $H(f)(a,b) > 0$ and $f_{xx}(a,b) < 0$. Then $f(x,y)$ has a local maximum at (a,b).

(3C) Let $H(f)(a,b) < 0$. Use Equation (1.1) to get

$$f(a+h,b+k) - f(a,b) = \tfrac{1}{2}\left(h^2 f_{xx} + 2hk f_{xy} + k^2 f_{yy}\right)_{(a,b)}.$$

First set $h = k$. Using Equation (1.1) again, we have

$$\lim_{h\to 0}\frac{f(a+h,b+h)-f(a,b)}{h^2} = \tfrac{1}{2}\left(f_{xx} + 2f_{xy} + f_{yy}\right)_{(a,b)}.$$

Next, set $h = -k$. Using Equation (1.1) again, we have

$$\lim_{h\to 0}\frac{f(a+h,b-h)-f(a,b)}{h^2} = \tfrac{1}{2}\left(f_{xx} - 2f_{xy} + f_{yy}\right)_{(a,b)}.$$

Since $H(f)(a,b) < 0$, these two limits have opposite signs. Due to continuity of f, $f(a+h,b+h) - f(a,b)$ and $f(a+h,b-h) - f(a,b)$ will have opposite signs in any neighborhood of (a,b). Thus $f(a,b)$ is a saddle point of $f(x,y)$.

Notice that if $H(f)(a,b) > 0$ and $f_{xx}(a,b) = 0$, these two possibilities (maximum or minimum) are not possible. Moreover, under this situation, since $H(f)(a,b) = f_{xx}f_{yy} - f_{xy}^2 = 0$, both f_{xx} and f_{yy} are replaced by the corresponding critical condition between first and second derivatives. Therefore, if $H(f)(a,b) = 0$, the test fails. □

Example 8.21. Find the extreme values of $f(x,y) = x^2 + y^2 - 2x - 4y + 4$.
Domain of f is $\mathbb{R}^2$. Making two bordered polylines. The first and second partial derivatives are $f_x = 2x - 2$, $f_y = 2y - 4$. The critical points are where $f_x = f_y = 0$. That is, $x = 1$, $y = 2$. Now compute $H(f)(x,y) = f_{xx}f_{yy} - f_{xy}^2 = 4 > 0$. Also $f_{xx} = 2 > 0$. Thus f has a local minimum at $(1,2)$, and the minimum value is

$$f(1,2) = 1 + 4 - 2 - 8 + 4 = -1.$$

Thus, f has an absolute minimum -1 at $(1,2)$.

8.9 Lagrange Multipliers

Our requirement is to minimize or maximize a certain function $f(x,y,z)$ subject to the constraint $g(x,y,z) = 0$. The constraint represents a surface in three dimensional space. Let S be a surface given by $g(x,y,z) = 0$. Let $f(x,y,z)$ have an extreme value at $P(x_0,y_0,z_0)$ on the surface S.

Let C be a curve given by $\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k}$ that lies on S and passes through P. Suppose for $t = t_0$, we get the point P, that is, $P = \vec{r}(t_0)$.

The composite function $h(t) = (f \circ g)(t) = f(x(t), y(t), z(t))$ represents the values that f takes on C. Since f has an extreme value at $P(t = t_0)$, the function $h(t)$ has an extreme value at $t = t_0$. Then $h^{'}(t_0) = 0$. That is,

$$0 = h^{'}(t_0) = f_x(P)x^{'}(t_0) + f_y(P)y^{'}(t_0) + f_z(P)z^{'}(t_0) = (\nabla f)(P) \,.\, \vec{r}(t_0).$$

For every such curve C, $(\nabla g)(P)$ is orthogonal to $\vec{r}^{\,'}(t_0)$. Thus, $(\nabla f)(P)$ is parallel to $(\nabla g)(P)$. If $(\nabla g)(P) \neq 0$, then

$$(\nabla g + \lambda \nabla g)(x_0, y_0, z_0) = 0 \text{ for some } \lambda \mathbb{R}.$$

Breaking into components, we have, at (x_0, y_0, z_0)

$$f_x + \lambda g_x = 0, f_y + \lambda g_y = 0, f_z + \lambda g_z = 0, g = 0.$$

We write the following result as our next theorem.

Theorem 8.15. Let $D \subset \mathbb{R}^2$ be a region. Let $f, g : D \to \mathbb{R}^2$ have continuous first order partial derivatives. If $g_x^2 + g_y^2 > 0$ for all $(x, y) \in D$, then each point (a, b) on the curve $g(x, y) = 0$, where $f(x, y)$ has maxima or minima corresponds to a solution (a, b, λ) of the system of equations

$$f_x(a, b) + \lambda g_x(a, b) = 0, f_y(a, b) + \lambda g_y(a, b) = 0, g(a, b) = 0.$$

Notice that if we set $F(x, y, z, \lambda) := f(x, y, z) + \lambda g(x, y, z) = 0$, then

$$F_x = f_x + \lambda g_x = 0, \quad F_y = f_y + \lambda g_y = 0, \quad F_z = f_z + \lambda g_z = 0.$$

Moreover, $g(x, y, z) = 0$ also comes from $F_\lambda = 0$.

We can now formulate the method of solving a constrained optimization problem.
Requirement: Find extrema of the function

$$f(x_1, \ldots, x_n)$$

subject to the conditions

$$g_1(x_1, \ldots, x_n) = 0, \quad \ldots, \quad g_m(x_1, \ldots, x_n) = 0.$$

Method: Set the auxiliary function

$$F(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) := f(x_1, \ldots, x_n) + \lambda_1 g_1(x_1, \ldots, x_n) + \cdots + \lambda_m g_m(x_1, \ldots, x_n).$$

Equate to zero the partial derivatives of F with respect to $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m$. It results in $m + n$ equations in $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m$.

Determine $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m$ from these equations. The required extremum points may be found from among these values of $x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m$. Remember that the method succeeds under the condition that such extreme values exist where $\nabla g_j \neq 0$ for any j. Further, the points of extremum thus obtained are only possible points of extremum. They need not be points of actual extremum. Other considerations may be required to determine whether any of such points is an actual maximum or minimum of $f(x, y)$ while (x, y) varies over the curve $g(x, y) = 0$.

Example 8.22. Find the maximum value of

$$f(x, y, z) = x + 2y + 3z$$

on the curve of intersection of the plane

$$g(x, y, z) := x - y + z - 1 = 0$$

and the cylinder

$$h(x, y, z) := x^2 + y^2 - 1 = 0.$$

The auxiliary function is

$$F(x, y, z, \lambda, \mu) := f + \lambda g + \mu h = x + 2y + 3z + \lambda(x - y + z - 1) + \mu(x^2 + y^2 - 1).$$

Setting $F_x = F_y = F_z = F_\lambda = F_\mu = 0$, for (x_0, y_0, z_0), we have

$$\begin{cases} 1 + \lambda + 2x_0\mu = 0, \\ 2 - \lambda + 2y_0\mu = 0, \\ 3 + \lambda = 0, \\ x_0 - y_0 + z_0 - 1 = 0, \\ x_0^2 + y_0^2 - 1 = 0. \end{cases}$$

We obtain

$$\lambda = -3, \quad x_0 = \frac{1}{\mu}, \quad y_0 = -\frac{5}{2\mu},$$

and

$$\frac{1}{\mu^2} + \frac{25}{4\mu^2} = 1 \quad \Rightarrow \quad \mu^2 = \frac{29}{4}.$$

Hence the possible extreme points are

$$x_0 = \pm\frac{2}{\sqrt{29}}, \quad y_0 = \mp\frac{5}{\sqrt{29}}, \quad z_0 = 1 \pm \frac{7}{\sqrt{29}}.$$

The corresponding values of $f(x_0, y_0, z_0)$ show that the maximum value of f is

$$f_{\max} = 3 + \sqrt{29}.$$

Notice that if $\mu = 0$, then $1 + \lambda = 0 = 2 - \lambda$ leads to inconsistency. Also, the conditions that $\nabla g \neq 0$ and $\nabla h \neq 0$ are satisfied automatically for the given constraints.

9 Multiple Integrals

9.1 Volume of a Solid of Revolution

The solid formed by rotating a plane region about a straight line in the same plane is called a *solid of revolution*. The line is called the *axis of revolution*.

Suppose the region is bounded above by the curve $y = f(x)$ and below by the x-axis, where $a \leqslant x \leqslant b$. To find the volume of the solid so generated, we divide the interval $[a, b]$ into n equal parts. Let the partition be

$$a = x_0 < x_1 < \cdots < x_n = b.$$

On the r-th subinterval, we approximate the solid by a right circular cylinder whose radius is $f(x_r^*)$ for some $x_r^* \in [x_{r-1}, x_r]$. If the axis is the x-axis, the cross-sectional area perpendicular to the axis is a circle. Then the volume of the solid of revolution is approximated by the sum

$$\sum_{r=1}^{n} \pi[f(x_r^*)]^2(x_r - x_{r-1}).$$

Then the volume of the solid of revolution is the limit of this sum, where

$$\Delta x = x_r - x_{r-1} \to 0.$$

Observe that the cross-sectional area is $A(x) = \pi[f(x)]^2$. Hence, if $f(x)$ is continuous on $[a, b]$, the required volume is

$$V = \lim_{\Delta x \to 0} \sum_{r=1}^{n} A(x_r^*)\Delta x = \int_a^b A(x)\,dx = \pi \int_a^b [f(x)]^2 dx.$$

If the axis of revolution is a straight line other than the x-axis, similar formulas can be obtained for the volume.

Example 9.1. Find the volume of the solid generated by revolving the region between the curve $y = \sqrt{x}$, $0 \le x \le 4$ and the x-axis about the x-axis.
The region between the curve and the x-axis is revolved about the x-axis. As shown in the figure, the required volume is

$$V = \pi \int_0^4 x dx = \pi \left[\frac{x^2}{2}\right]_0^4 = 8\pi.$$

Example 9.2. Find the volume of the sphere $x^2 + y^2 + z^2 = a^2, a > 0$.
The region bounded by the upper semicircle $y = \sqrt{r^2 - x^2}$, the x-axis, and the ordinates $x = -r$ and $x = r$ is revolved about the x-axis. The resulting solid is a sphere of radius r. Hence, the required volume is

$$V = \pi \int_{-r}^{r} (\sqrt{r^2 - x^2})^2 dx = \pi \int_{-r}^{r} (r^2 - x^2)\,dx = \frac{4}{3}\pi r^3.$$

9.2 Approximating Volume

We now consider solids which are not necessarily solids of revolution. First, we take a typical simpler case, when a given solid has all plane faces except one, which is a portion of a surface given by a function $f(x, y)$.

Let $f(x, y)$ be defined on the rectangle $R : a \le x \le b, c \le y \le d$. For simplicity, take $f(x, y) \ge 0$. The graph of f is the surface $z = f(x, y)$. We approximate the volume of the solid

$$S = \{(x, y, z) : (x, y) \in R, 0 \le z \le f(x, y)\}$$

by partitioning R and then adding up the volumes of the solid rods. So, consider a *partition* of R as

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j], \quad 1 \le i \le m,\ 1 \le j \le n,\ a = x_0,\ b = x_m,\ c = y_0,\ d = y_n.$$

Denote by $A(R_{ij})$ the area of the rectangle R_{ij}. Denote by $\|P\| = \max A(R_{ij})$ the *norm of* P. Choose *sample points* $(x_i^*, y_j^*) \in R_{ij}$. An approximation to the volume of S is the *Riemann sum*

$$S_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i^*, y_j^*) A(R_{ij}).$$

If the limit of S_{mn} exists as $\|P\| \to 0$, then this limit is called the *double integral* of $f(x, y)$. It is denoted by

$$\iint_R f(x, y)\, dA.$$

Whenever the integral exists, it is also enough to consider *uniform partitions*, that is,

$$x_i - x_{i-1} = \frac{b-a}{m} = \Delta x, \quad y_j - y_{j-1} = \frac{d-c}{n} = \Delta y.$$

In this case, we write $A(R_{ij}) = \Delta x \Delta y$. Then

$$\iint_R f(x, y)\, dA = \lim_{\|P\| \to 0} S_{mn} = \lim_{m,n \to \infty} \sum_{i=1}^{m} \sum_{j=1}^{n} f(x_i^*, y_j^*) \Delta x \Delta y.$$

Since $f(x, y) \geq 0$, the value of this integral is the volume of the solid S bounded by the rectangle R and the surface $z = f(x, y)$. When the integral of $f(x, y)$ exists, we say that f is *Riemann integrable* or just *integrable*. Riemann sum is well defined even if f is not a positive function. However, the double integral computes the signed volume. Analogous to the single-variable case, we have the following result:

Theorem 9.1. Each continuous function defined on a closed bounded rectangle is integrable.

Volumes of solids can also be calculated by using iterated integrals. For example, to find the volume V of the solid raised over the rectangle $R : [0, 2] \times [0, 1]$ and bounded above by the plane $z = 4 - x - y$, we proceed as follows (similar to solids of revolution):

Suppose $A(x)$ is the cross-sectional area at x. Then

$$V = \int_0^2 A(x)\, dx.$$

Now, $A(x) = \int_0^{4-x} (4 - x - y)\, dy$. Therefore,

$$V = \int_0^2 \left[\int_0^{4-x} (4 - x - y)\, dy \right] dx.$$

The expression on the left is a double integral and on the right is an *iterated integral*.

Theorem 9.2 (Fubini)**.** Let $R = [a, b] \times [c, d]$. Let $f : R \to \mathbb{R}$ be a continuous function. Then

$$\iint_R f(x, y)\, dA = \int_a^b \left[\int_c^d f(x, y)\, dy \right] dx = \int_c^d \left[\int_a^b f(x, y)\, dx \right] dy.$$

Example 9.3. Evaluate $\iint_R (1 - 6x^3 y)\, dA$, where $R = [0, 2] \times [-1, 1]$.

$$\iint_R (1 - 6x^3 y)\, dA = \int_0^2 \int_{-1}^1 (1 - 6x^3 y)\, dy\, dx = \int_0^2 [2 - 0]\, dx = 4.$$

Also, reversing the order of integration, we have

$$\iint_R (1 - 6x^3 y)\, dA = \int_{-1}^1 \int_0^2 (1 - 6x^3 y)\, dx\, dy = \int_{-1}^1 [2 - 16y]\, dy = 4.$$

The double integrals can be extended to functions defined on non-rectangular regions. Essentially, the approach is the same as earlier. We partition the region into smaller rectangles, form the Riemann sum, take its limit as the norm of the partition goes to zero. The double integral of f over such a bounded region R can also be evaluated using iterated integrals. Look at R bounded by two continuous functions $g_1(x)$ and $g_2(x)$; or, as a region bounded by two continuous functions $h_1(y)$ and $h_2(y)$.

Theorem 9.3. Let $f(x, y)$ be a continuous real-valued function on a region R.

1. If R is given by $a \leq x \leq b, g_1(x) \leq y \leq g_2(x)$, where $g_1, g_2 : [a, b] \to \mathbb{R}$ are continuous, then
$$\iint_R f(x,y)\,dA = \int_a^b \int_{g_1(x)}^{g_2(x)} f(x,y)\,dy\,dx.$$

2. If R is given by $c \leq y \leq d, h_1(y) \leq x \leq h_2(y)$, where $h_1, h_2 : [c, d] \to \mathbb{R}$ are continuous, then
$$\iint_R f(x,y)\,dA = \int_c^d \int_{h_1(y)}^{h_2(y)} f(x,y)\,dx\,dy.$$

Example 9.4. Evaluate
$$\iint_R \frac{\sin x}{x}\,dA,$$
where R is the triangle in the xy-plane bounded by the lines $y = 0$, $x = 1$, and $y = x$. The triangular region R can be expressed as
$$R = \{(x, y) : 0 \leq y \leq 1,\ y \leq x \leq 1\}.$$
Thus,
$$\iint_R \frac{\sin x}{x}\,dA = \int_0^1 \int_y^1 \frac{\sin x}{x}\,dx\,dy.$$
We are stuck here, since the inner integral cannot be evaluated in elementary form. On the other hand, we can express the same region as $R = \{(x, y) : 0 \leq x \leq 1,\ 0 \leq y \leq x\}$. Then
$$\iint_R \frac{\sin x}{x}\,dA = \int_0^1 \int_0^x \frac{\sin x}{x}\,dy\,dx = \int_0^1 \left(\frac{\sin x}{x} \int_0^x dy\right) dx = \int_0^1 \sin x\,dx = -\cos(1) + 1.$$

Properties of double integrals with respect to addition, multiplication etc. are as follows.

Theorem 9.4. Let $f(x, y)$ and $g(x, y)$ be continuous on a region D, and let c be a constant. Then we have the following:

1. Constant Multiple: $\iint_D c\,f(x,y)\,dA = c \iint_D f(x,y)\,dA$.
2. Sum-Difference: $\iint_D \big(f(x,y) \pm g(x,y)\big)\,dA = \iint_D f(x,y)\,dA \pm \iint_D g(x,y)\,dA$.
3. Additivity: $\iint_{D\cup R} f(x,y)\,dA = \iint_D f(x,y)\,dA + \iint_R f(x,y)\,dA$, provided $f(x, y)$ is continuous on R as well, and D and R are non-overlapping.
4. Domination: If $f(x, y) \leq g(x, y)$ on D, then $\iint_D f(x,y)\,dA \leq \iint_D g(x,y)\,dA$.
5. Area: $\iint_D 1\,dA = \Delta(D) =$ Area of D.
6. Boundedness: If $m \leq f(x, y) \leq M$ on D, then $m\,\Delta(D) \leq \iint_D f(x,y)\,dA \leq M\,\Delta(D)$.

9.3 Triple Integral

Let $f(x, y, z)$ be a real-valued function defined on a bounded region $D \subset \mathbb{R}^3$. As before, we divide the region into smaller cubes enclosed by planes parallel to the coordinate planes. The set of these smaller cubes is called a partition P. The norm of the partition is the maximum volume of any of the smaller cubes. Then form the Riemann sum S and take its limit as the cubes become smaller and smaller. If the limit exists, we say that the limit is the *triple integral* of the function over the region D:

$$\iiint_D f(x, y, z)\, dV = \lim_{\|P\| \to 0} \sum f(x_i^*, y_j^*, z_k^*)(x_i - x_{i-1})(y_j - y_{j-1})(z_k - z_{k-1}),$$

where (x_i^*, y_j^*, z_k^*) is a point in the (i, j, k)-th cube of the partition.

As before, Fubini's theorem states that for continuous functions, if the region D can be written as

$$D = \{(x, y, z) : a \leq x \leq b,\ g_1(x) \leq y \leq g_2(x),\ h_1(x, y) \leq z \leq h_2(x, y)\},$$

then the triple integral can be expressed as an iterated integral:

$$\iiint_D f(x, y, z)\, dV = \int_a^b \int_{g_1(x)}^{g_2(x)} \int_{h_1(x,y)}^{h_2(x,y)} f(x, y, z)\, dz\, dy\, dx.$$

Observe that the volume of D is given by

$$\text{Volume}(D) = \iiint_D 1\, dV.$$

All properties for double integrals hold analogously for triple integrals.

Example 9.5. Evaluate $\int_0^1 \int_0^z \int_0^y e^{(1-x)^3}\, dx\, dy\, dz$ by changing the order of integration. The region is $D = \{(x, y, z) : 0 \leq z \leq 1,\ 0 \leq y \leq z,\ 0 \leq x \leq y\}$.

Its projection on the yz-plane is the triangle bounded by $y = 0$, $z = 1$, and $y = z$, i.e.

$$\{(y, z) : 0 \leq z \leq 1,\ 0 \leq y \leq z\}.$$

Its projection on the xy-plane is the triangle bounded by $x = 0$, $y = 1$, and $y = x$, i.e.

$$\{(x, y) : 0 \leq y \leq 1,\ 0 \leq x \leq y\}.$$

Its projection on the zx-plane is the triangle bounded by $z = 0$, $x = 1$, and $x = z$, i.e.

$$\{(z, x) : 0 \leq x \leq 1,\ x \leq z \leq 1\}.$$

We change the order of integration from $dx\, dy\, dz$ to $dz\, dy\, dx$. Since $0 \leq x \leq y \leq z \leq 1$, the region can be described as

$$D = \{(x, y, z) : 0 \leq x \leq 1,\ x \leq y \leq 1,\ y \leq z \leq 1\}.$$

Thus,

$$\begin{aligned}
\int_0^1\int_0^z\int_0^y e^{(1-x)^3}\,dx\,dy\,dz &= \int_0^1\int_x^1\int_y^1 e^{(1-x)^3}\,dz\,dy\,dx \\
&= \int_0^1\int_x^1 (1-y)e^{(1-x)^3}\,dz\,dx \\
&= \int_0^1 \frac{(1-x)^2}{2}e^{(1-x)^3}\,dx \\
&= \int_0^1 \frac{e^t}{6}\,dt \\
&= \frac{e-1}{6}.
\end{aligned}$$

9.4 Change of Variables

Suppose f maps a region D in $\mathbb{R}^2$ onto a region R in $\mathbb{R}^2$ in a one-one manner. For convenience, we say that D is a region in the uv-plane and R is a region in the xy-plane; and f maps (u, v) to (x, y). Then f can be thought of as a pair of maps: (f_1, f_2). That is, $x = f_1(u, v)$ and $y = f_2(u, v)$. We often show this dependence implicitly by writing

$$x = x(u, v), y = y(u, v).$$

Suppose a transformation $(u, v) \mapsto (x, y)$ is given. How does the area of a small rectangle change under this map? A typical small rectangle in the uv-plane with sides Δu and Δv has corners

$$A_1 = (a, b),\ A_2 = (a + \Delta u,\ b),\ A_3 = (a,\ b + \Delta v),\ A_4 = (a + \Delta u,\ b + \Delta v).$$

Let the images of A_k under $(u, v) \to (x, y)$ be $B_k = (a_k, b_k)$ for $k = 1, \ldots, 4$. Then

$$a_1 = x(a, b),\ a_2 = x(a + \Delta u, b) \approx x(a, b) + x_u\,\Delta u,$$

$$a_3 = x(a, b + \Delta v) \approx x(a, b) + x_v\,\Delta v,$$

$$a_4 = x(a + \Delta u, b + \Delta v) \approx x(a, b) + x_u\,\Delta u + x_v\,\Delta v,$$

where $x_u = x_u(a, b)$ and $x_v = x_v(a, b)$. Similar approximations hold for b_1, b_2, b_3, b_4 using y_u and y_v. The area of the image of the rectangle $A_1A_2A_3A_4$ is approximately the area of the parallelogram $B_1B_2B_3B_4$, which equals twice the area of the triangle $B_1B_2B_4$. This area equals

$$\big|(a_4 - a_1)(b_4 - b_2) - (a_4 - a_2)(b_4 - b_1)\big| = \left|\det\begin{pmatrix} x_u & x_v \\ y_u & y_v \end{pmatrix}\right|_{(a,b)} \Delta u\,\Delta v.$$

This determinant is called the *Jacobian* of the map $(u, v) \mapsto (x, y)$ and is denoted by

$$J(x(u, v), y(u, v)) = \frac{\partial(x, y)}{\partial(u, v)}.$$

Thus, the area of the image of a rectangle with corner at (a, b) and side lengths Δu, Δv is approximately

$$|J(x(u, v), y(u, v))|\,\Delta u\,\Delta v,$$

where J is evaluated at (a, b). This approximation assumes that x_u, x_v, y_u, y_v are continuous.

Assume $x = x(u, v)$ and $y = y(u, v)$ have continuous partial derivatives with respect to u and v. Assume also that a region D in the uv-plane corresponds one-to-one with a region R in the xy-plane under the transformation. Let $f(x, y)$ be a continuous real-valued function on R and define

$$\tilde{f}(u, v) = f(x(u, v), y(u, v)).$$

Divide D into small rectangles. The areas of rectangles and their images satisfy

$$\text{Area of } R = |J| \cdot \text{Area of } D = \left|\frac{\partial(x, y)}{\partial(u, v)}\right| \text{Area of } D.$$

Forming Riemann sums and taking limits, we obtain the change–of–variables formula:

$$\iint_R f(x, y)\, dA = \iint_D \tilde{f}(u, v)\, |J(x(u, v), y(u, v))|\, dA = \iint_D \tilde{f}(u, v) \left|\frac{\partial(x, y)}{\partial(u, v)}\right| dA.$$

Example 9.6. Evaluate the double integral $\iint_R (y - x)\, dA$, where R is the region bounded by the lines $y - x = 1, y - x = -3, 3y + x = 7, 3y + x = 15$. Take

$$u = y - x, \qquad v = 3y + x.$$

Then solving for x and y,

$$x = \frac{1}{4}(v - 3u), \qquad y = \frac{1}{4}(u + v).$$

Thus the region becomes

$$D = \{(u, v) : -3 \le u \le 1,\ 7 \le v \le 15\}.$$

The Jacobian is

$$J = x_u y_v - x_v y_u = \left(-\tfrac{3}{4}\right)\left(\tfrac{1}{4}\right) - \left(\tfrac{1}{4}\right)\left(\tfrac{1}{4}\right) = -\frac{1}{4}.$$

Therefore,

$$\iint_R (y - x)\, dA = \iint_D u\, |J|\, dA = \iint_D \frac{1}{4}\, u\, dA = \int_{-3}^{1} \int_7^{15} \frac{1}{4}\, u\, dv\, du = \int_{-3}^{1} \frac{1}{8}(15^2 - 7^2)\, dv = 88.$$

Example 9.7. Evaluate $\int_0^4 \int_{y/2}^{1+y/2} \frac{2x-y}{2}\, dx\, dy$ using the transformation $u = x - \frac{y}{2}, v = \frac{y}{2}$.

Note that $x = u + v, y = 2v$. The regions become

$$R = \{(x, y) : 0 \le y \le 4,\ y/2 \le x \le 1 + y/2\},\ G = \{(u, v) : 0 \le u \le 1,\ 0 \le v \le 2\}.$$

Since $2x - y = 2(u + v) - 2v = 2u$, then the Jacobian is

$$|J| = |x_u y_v - x_v y_u| = |(1)(2) - (1)(0)| = 2.$$

Thus, we obtain

$$\iint_R \frac{2x - y}{2}\, dA = \iint_G 2u\, dA = \int_0^2 \int_0^1 2u\, du\, dv = \int_0^2 \left[2u^2\right]_0^1 dv = \int_0^2 dv = 2.$$

10 Vector Integrals

10.1 Line Integral

Line integrals are single integrals which are obtained by integrating a function over a curve instead of integrating over an interval.

Let $f(x, y, z)$ be a real-valued function defined on a region D. Let C be a curve lying in D, given in parametric form

$$r(t) = x(t)\,\hat{i} + y(t)\,\hat{j} + z(t)\,\hat{k},\ a \le t \le b.$$

The values of f along the curve C are given by the composite function $f(x(t), y(t), z(t))$. Partition the curve C into n sub-arcs. Choose a point (x_k, y_k, z_k) on the kth sub-arc. If the kth sub-arc has length Δs_k, form the Riemann sum

$$S_n = \sum_{k=1}^{n} f(x_k, y_k, z_k)\,\Delta s_k.$$

As $n \to \infty$, the lengths $\Delta s_k \to 0$. If the limit

$$\lim_{n\to\infty} S_n$$

exists, then this limit is called the *line integral* of f *over the curve* C:

$$\int_C f(x, y, z)\,ds = \lim_{n\to\infty} S_n.$$

In practice, the line integral is computed by parameterizing the curve C.

Theorem 10.1. Let $C : r(t) = x(t)\,\hat{i} + y(t)\,\hat{j} + z(t)\,\hat{k}$ be a parametrization of a curve C lying in a region $D \subset \mathbb{R}^3$. Let $f : D \to \mathbb{R}$ be continuous, and let the component functions $x(t), y(t), z(t)$ have continuous first-order derivatives. Then the line integral of f over C exists and is given by

$$\int_C f(x, y, z)\,ds = \int_a^b f(x(t), y(t), z(t))\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\,dt.$$

We also write

$$ds = |r'(t)|\,dt = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\,dt.$$

Example 10.1. Integrate $f(x, y, z) = x - 3y^2 + z$ over the line segment from $(0, 0, 0)$ to $(1, 1, 1)$.

Lets first parametrize the curve C: $r(t) = t\,\hat{i} + t\,\hat{j} + t\,\hat{k},\ 0 \le t \le 1$. Then

$$x(t) = y(t) = z(t) = t,\ r'(t)| = \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}.$$

Thus,

$$\int_C f\,ds = \int_0^1 \big(x(t) - 3y^2(t) + z(t)\big)|\mathbf{r}'(t)|\,dt = \int_0^1 \big(t - 3t^2 + t\big)\sqrt{3}\,dt = 0.$$

10.2 Line Integral of Vector Fields

We now generalize line integrals to vector fields.

A *vector field* is a function defined on a region D in the plane or in space that assigns a vector to each point in D. If D is a region in space, a vector field on D may be written as

$$F(x, y, z) = M(x, y, z)\,\hat{\imath} + N(x, y, z)\,\hat{\jmath} + P(x, y, z)\,\hat{k}.$$

For example: Vectors in a gravitational field point toward the center of mass that produces the field. The velocity vectors of a projectile define a vector field along its trajectory. Let $F(x, y, z)$ be a continuous vector field defined over a curve C given parametrically by

$$r(t) = x(t)\,\hat{\imath} + y(t)\,\hat{\jmath} + z(t)\,\hat{k},\;\; a \le t \le b.$$

The *line integral* of F along C, also called the *work done by moving a particle along C under the force field F*, is defined by

$$\int_C F \cdot dr = \int_C F(r(t)) \cdot r'(t)\,dt = \int_C F \cdot T\,ds,$$

where

$$T(t) = \frac{r'(t)}{|r'(t)|}$$

is the unit tangent vector at a point on C.

Example 10.2. Evaluate the line integral of the vector field

$$F(x, y, z) = x^2\,\hat{\imath} - xy\,\hat{\jmath}$$

along the first-quarter unit circle in the first quadrant.

The curve C is given by $r(t) = \cos t\,\hat{\imath} + \sin t\,\hat{\jmath},\; 0 \le t \le \frac{\pi}{2}$. Then the work done is

$$\begin{aligned}\int_C F \cdot dr &= \int_0^{\pi/2} F(r(t)) \cdot r'(t)\,dt \\ &= \int_0^{\pi/2} \big(\cos^2 t(-\sin t) + (-\cos t \sin t)(\cos t)\big)\,dt \\ &= -\frac{2}{3}.\end{aligned}$$

Let the vector field be $F(x, y, z) = M(x, y, z)\,\hat{\imath} + N(x, y, z)\,\hat{\jmath} + P(x, y, z)\,\hat{k}$. Let C be the curve given by

$$r(t) = x(t)\,\hat{\imath} + y(t)\,\hat{\jmath} + z(t)\,\hat{k},\;\; a \le t \le b.$$

Then the line integral of $\mathbf{F}$ along C is

$$\int_C F \cdot dr = \int_a^b F(r(t)) \cdot r'(t)\,dt = \int_a^b \Big[M\,\frac{dx}{dt} + N\,\frac{dy}{dt} + P\,\frac{dz}{dt}\Big]dt = \int_C M\,dx + N\,dy + P\,dz.$$

10.3 Conservative Fields

Let $f(x, y, z)$ be a function defined on a region in $\mathbb{R}^3$. If the partial derivatives f_x, f_y, f_z exist, then the *gradient field* of $f(x, y, z)$ is the field of gradient vectors

$$\operatorname{grad} f = \nabla f = \frac{\partial f}{\partial x}\,\hat{\imath} + \frac{\partial f}{\partial y}\,\hat{\jmath} + \frac{\partial f}{\partial z}\,\hat{k}.$$

The gradient field of the surface $f(x, y, z) = c$ can be visualized as follows: At each point on the surface, we draw the gradient vector, which is normal (perpendicular) to the surface, at that point.

Example 10.3. For $f(x, y, z) = xyz$, the gradient field is $\nabla f = yz\,\hat{\imath} + zx\,\hat{\jmath} + xy\,\hat{k}$.Notice that $f(x, y, z)$ has a continuous gradient if and only if f_x, f_y, f_z are continuous on the domain of f.

A vector field F is called *conservative* if there exists a scalar function f such that $F = \nabla f$. We say that a line integral

$$\int_C F \cdot fr$$

is *independent of path* if for any curve C' lying in the domain of F, and having the same initial and terminal points as C, we have

$$\int_C F \cdot dr = \int_{C'} F \cdot dr.$$

Thus, if F is conservative, then the line integral $\int_C F \cdot dr$ is path independent.

A *closed curve* is a curve having the same initial and end points. When C is a closed curve, the line integral over C is written as

$$\oint_C F \cdot dr.$$

A *simple curve* is a curve which does not intersect itself. A connected region D is said to be a *simply connected region* if and only if every simple closed curve lying in D encloses only points from D.

10.4 Green's Theorem

Let C be a simple closed curve in the plane. The positive orientation of C refers to a single counter-clockwise traversal of C. If C is given by $r(t), a \le t \le b$, then its positive orientation refers to a traversal of C keeping the region D bounded by the curve to the left.

Theorem 10.2 (Green's Theorem)**.** Let C be a positively oriented, simple closed, piecewise smooth curve in the plane, and let D be the region bounded by C (that is, $C = \partial D$). If $M(x, y)$ and $N(x, y)$ have continuous partial derivatives on an open region containing D, then

1. $\oint_C M\,dx + N\,dy = \iint_D \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right) dA.$

2. $\oint_C M\,dy - N\,dx = \iint_D \left(\frac{\partial M}{\partial x} + \frac{\partial N}{\partial y}\right) dA.$

Proof. We only prove for a special kind of region to give an idea of how it is proved. Consider the region

$$D = \{(x, y) : a \le x \le b,\ f(x) \le y \le g(x)\}$$

Assume that f and g are continuous functions. Then

$$\iint_D M_y \, dA = \int_a^b \int_{f(x)}^{g(x)} M_y \, dy \, dx = \int_a^b [M(x, g(x)) - M(x, f(x))] \, dx.$$

Now we compute the line integral $\oint_C M \, dx$ by breaking C into four parts C_1, C_2, C_3, and C_4. The curve C_1 is given by $x = x$, $y = f(x)$, $a \le x \le b$. Thus

$$\int_{C_1} M \, dx = \int_a^b M(x, f(x)) \, dx.$$

On C_2 and also on C_4 the variable x is a single point. So,

$$\int_{C_2} M \, dx = 0, \qquad \int_{C_4} M \, dx = 0.$$

As x increases, C_3 is traversed backward. That is, C_3 is given by $x = x$, $y = g(x)$, $a \le x \le b$. So,

$$\int_{C_3} M \, dx = -\int_a^b M(x, g(x)) \, dx.$$

Therefore,

$$\iint_D M_y \, dA = \oint_C M \, dx.$$

Similarly, express D using y as the variable of integration. Then we have

$$\iint_D N_x \, dA = \oint_C N \, dy.$$

Next, add the two results obtained to get

$$\oint_C (M \, dx + N \, dy) = \iint_D (N_x - M_y) \, dA.$$

The second form follows similarly. $\square$

Example 10.4. Verify Green's theorem for the field

$$F = (x - y)\,\hat{\imath} + x\,\hat{\jmath}$$

where C is the unit circle oriented positively.

The curve C is parametrized by

$$r(t) = (\cos t)\,\hat{\imath} + (\sin t)\,\hat{\jmath},\ 0 \le t \le 2\pi.$$

The region D is the unit disk. We have

$$M = x - y = \cos t - \sin t,\ N = x = \cos t,$$

$$dx = -\sin t\, dt,\;\; dy = \cos t\, dt,$$

and the partial derivatives

$$M_x = 1, \quad M_y = -1, \qquad N_x = 1, \quad N_y = 0.$$

Compute the line integral:

$$\oint_C (M\, dx + N\, dy) = \int_0^{2\pi} \big[(\cos t - \sin t)(-\sin t) + \cos^2 t\big]\, dt = 2\pi.$$

Compute the corresponding double integral:

$$\iint_D (N_x - M_y)\, dA = \iint_D (1 - (-1))\, dA = 2 \cdot \text{Area}(D) = 2\pi.$$

Thus, both forms agree with their corresponding double integrals, and *Green's theorem is verified.*

Example 10.5. Evaluate the integral $I = \oint_C \big(3y - e^{\sin x}\big)\, dx + \big(7x + \sqrt{1+y^4}\big)\, dy$, where C is the positively oriented circle $x^2 + y^2 = 9$.

Take D as the disk $x^2 + y^2 \leq 9$. Then by Green's theorem

$$\begin{aligned} I &= \iint_D \left[\frac{\partial}{\partial x}\big(7x + \sqrt{1+y^4}\big) - \frac{\partial}{\partial y}\big(3y - e^{\sin x}\big)\right] dA \\ &= \iint_D (7-3)\, dA \\ &= 4 \cdot 9\pi = 36\pi. \end{aligned}$$

10.5 Curl and Divergence of a Vector Field

If $F = M\,\hat{\imath} + N\,\hat{\jmath} + P\,\hat{k}$ is a vector field in $\mathbb{R}^3$, where the partial derivatives of the component functions exist, then the *curl* of F is the vector field

$$\operatorname{curl} F = \left(\frac{\partial P}{\partial y} - \frac{\partial N}{\partial z}\right)\hat{\imath} + \left(\frac{\partial M}{\partial z} - \frac{\partial P}{\partial x}\right)\hat{\jmath} + \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right)\hat{k}.$$

Writing in operator notation, recall that

$$\nabla = \hat{\imath}\frac{\partial}{\partial x} + \hat{\jmath}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}.$$

Then

$$\operatorname{curl} F = \nabla \times F = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ M & N & P \end{vmatrix}.$$

Example 10.6. If $F = zx\,\hat{\imath} + xyz\,\hat{\jmath} - y^2\,\hat{k}$, then

$$\operatorname{curl} F = -y(2+x)\,\hat{\imath} + x\,\hat{\jmath} + yz\,\hat{k}.$$

Theorem 10.3. Let F be a vector field defined over a simply connected region D whose component functions have continuous second-order partial derivatives. Then F is conservative if and only if $\nabla \times F = 0$.

It is important to note that, $\nabla \times \nabla f = 0$.

Example 10.7. Is the vector field $F = zx\,\hat{\imath} + xyz\,\hat{\jmath} - y^2\hat{k}$ conservative? We compute the curl.

$$\operatorname{curl} F = (-y(2+x))\,\hat{\imath} + x\,\hat{\jmath} + yz\,\hat{k} \neq 0.$$

Thus, F is not conservative.

Example 10.8. Is the vector field $F = y^2z^3\,\hat{\imath} + 2xyz^3\,\hat{\jmath} + 3xy^2z^2\,\hat{k}$ conservative? Compute the curl.

$$\operatorname{curl} F = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ y^2z^3 & 2xyz^3 & 3xy^2z^2 \end{vmatrix} = \left(6xy^2z^2 - 6xy^2z^2\right)\hat{\imath} - \left(3y^2z^2 - 3y^2z^2\right)\hat{\jmath} + \left(2yz^3 - 2yz^3\right)\hat{k} = 0.$$

Hence, F is conservative. Indeed, $F = \nabla f$, where $f(x, y, z) = xy^2z^3$.

If $F = M\,\hat{\imath} + N\,\hat{\jmath} + P\,\hat{k}$ is a vector field defined on a region, where its component functions have first–order partial derivatives, then the *divergence* of F is

$$\operatorname{div} F = \nabla \cdot F = \frac{\partial M}{\partial x} + \frac{\partial N}{\partial y} + \frac{\partial P}{\partial z}.$$

The divergence is also called the *flux* or *flux density.*

Theorem 10.4. Let $F = M\,\hat{\imath} + N\,\hat{\jmath} + P\,\hat{k}$ be a vector field defined on a simply connected region $D \subset \mathbb{R}^3$, where M, N, P have continuous second–order partial derivatives. Then $\operatorname{div}\operatorname{curl} F = 0$.

Proof. Note that

$$\begin{aligned}\operatorname{div}\operatorname{curl} F &= \nabla \cdot (\nabla \times F) \\ &= \frac{\partial}{\partial x}\left(\frac{\partial P}{\partial y} - \frac{\partial N}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial M}{\partial z} - \frac{\partial P}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right).\end{aligned}$$

Using equality of mixed partial derivatives (Clairaut's theorem), each pair of terms cancels:

$$\frac{\partial^2 P}{\partial x \partial y} - \frac{\partial^2 N}{\partial x \partial z} + \frac{\partial^2 M}{\partial y \partial z} - \frac{\partial^2 P}{\partial y \partial x} + \frac{\partial^2 N}{\partial z \partial x} - \frac{\partial^2 M}{\partial z \partial y} = 0.$$

This completes the proof. □

The divergence of the gradient of a scalar function f is called the *Laplacian* of f, since

$$\operatorname{div}\operatorname{grad} f = \nabla \cdot \nabla f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} := \nabla^2 f.$$

The operator $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is called the *Laplacian.*

10.6 Surface Area

As we know, smooth surfaces can be given by a function such as $z = f(x, y)$. More generally, a smooth surface is given parametrically by $x = x(u, v),\ y = y(u, v),\ z = z(u, v)$, where (u, v) varies over a given parameter region in the uv-plane. The parametric equations can also be written in vector form as $r(u, v) = x(u, v)\,\hat{i} + y(u, v)\,\hat{j} + z(u, v)\,\hat{k}$. Let S be a smooth surface given parametrically by $x = x(u, v),\ y = y(u, v),\ z = z(u, v)$, where (u, v) ranges over a parameter region D in the uv-plane. Suppose that S is covered exactly once as (u, v) varies over D. For simplicity, assume that D is a rectangle. We write S in vector form

$$r(u, v) = x(u, v)\,\hat{i} + y(u, v)\,\hat{j} + z(u, v)\,\hat{k}.$$

Divide D into smaller rectangles R_{ij} with the lower–left corner point $P_{ij} = (u_i, v_j)$. For simplicity, let the partition be uniform with u-lengths Δu and v-lengths Δv. The part S_{ij} of S that corresponds to R_{ij} has the corner P_{ij} with position vector

$$r(u_i, v_j).$$

The tangent vectors to S at P_{ij} are

$$r_u := r_u(u_i, v_j) = x_u(u_i, v_j)\,\hat{i} + y_u(u_i, v_j)\,\hat{j} + z_u(u_i, v_j)\,\hat{k},$$

$$r_v := r_v(u_i, v_j) = x_v(u_i, v_j)\,\hat{i} + y_v(u_i, v_j)\,\hat{j} + z_v(u_i, v_j)\,\hat{k}.$$

The tangent plane to S is the plane that contains the two tangent vectors $r_u(u_i, v_j)$ and $r_v(u_i, v_j)$. The normal to S at P_{ij} is the vector

$$r_u(u_i, v_j) \times r_v(u_i, v_j).$$

Since S is assumed to be smooth, this normal vector is nonzero. The part S_{ij} is a curved parallelogram on S whose sides can be approximated by the vectors $r_u\,\Delta u$ and $r_v\,\Delta v$. Thus the area of S_{ij} can be approximated by

$$\text{Area}(S_{ij}) \approx \|r_u(u_i, v_j) \times r_v(u_i, v_j)\|\ \Delta u\,\Delta v.$$

An approximation to the area of S is then obtained by summing over all i and j:

$$\text{Area}(S) \approx \sum_i \sum_j \|r_u(u_i, v_j) \times r_v(u_i, v_j)\|\ \Delta u\,\Delta v.$$

We define the surface area by taking the limit as the partition is refined. Let S be a smooth surface given parametrically by

$$r(u, v) = x(u, v)\,\hat{i} + y(u, v)\,\hat{j} + z(u, v)\,\hat{k},$$

where $(u, v) \in D$, a region in the uv-plane, and suppose S is covered exactly once as (u, v) varies over D. Then the surface area of S is

$$\text{Area}(S) = \iint_D \|r_u \times r_v\|\ dA,$$

where

$$r_u = x_u\,\hat{i} + y_u\,\hat{j} + z_u\,\hat{k},\ \ r_v = x_v\,\hat{i} + y_v\,\hat{j} + z_v\,\hat{k}.$$

In the case where the surface S is given as the graph of a function $z = f(x, y)$ over a region $(x, y) \in D$, we take the parameters $u = x$, $v = y$, and $z = z(u, v) = f(x, y)$. Thus S is given by

$$r(u, v) = u\,\hat{\imath} + v\,\hat{\jmath} + f(u, v)\,\hat{k}.$$

We see that

$$r_u = \hat{\imath} + f_x\,\hat{k},\ r_v = \hat{\jmath} + f_y\,\hat{k}.$$

and

$$r_u \times r_v = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix} = f_x\,\hat{\imath} - f_y\,\hat{\jmath} + \hat{k}.$$

Therefore,

$$\text{Area}(S) = \iint_D \|r_u \times r_v\|\; dA = \iint_D \sqrt{f_x^2 + f_y^2 + 1}\, dA.$$

Example 10.9. Find the surface area of the part of the surface

$$z = x^2 + 2y$$

that lies above the triangular region in the xy-plane with vertices $(0, 0)$, $(1, 0)$, and $(1, 1)$. The triangular region is

$$T = \{(x, y) : 0 \leq x \leq 1,\ 0 \leq y \leq x\}.$$

The required surface area is

$$\iint_T \sqrt{(2x)^2 + 2^2 + 1}\, dA = \int_0^1 \int_0^x \sqrt{4x^2 + 5}\, dy\, dx.$$

Evaluating,

$$\int_0^1 \int_0^x \sqrt{4x^2 + 5}\, dy\, dx = \int_0^1 x\sqrt{4x^2 + 5}\, dx = \frac{1}{12}\left((27)^{3/2} - 5^{3/2}\right).$$

Surface Area - A Generalized Form: Recall that for a surface S which is given by $z = f(x, y)$, the surface area is

$$\iint_D \sqrt{f_x^2 + f_y^2 + 1}\, dA.$$

Here, D is the region in the xy-plane obtained by projecting S onto the plane. Look at this surface as $f(x, y, z) = z - f(x, y) = 0$. Then

$$\nabla f = f_x\,\hat{\imath} + f_y\,\hat{\jmath} - \hat{k}.$$

If p is the unit normal to the projected rectangle, then

$$p = \hat{k}.$$

Thus,

$$\frac{\|\nabla f\|}{|\nabla f \cdot p|} = \frac{\sqrt{f_x^2 + f_y^2 + 1}}{1},$$

which is the integrand in the surface area formula.
Warning: $\nabla f \cdot \mathbf{p}$ must not be zero.
A derivation similar to the surface area formula gives the following:

Theorem 10.5. Let the surface S be given implicitly by

$$f(x, y, z) = c.$$

Let R be a closed bounded region obtained by projecting the surface onto a plane whose unit normal is p. Suppose that f is continuous on R and that $\nabla f \cdot p \neq 0$ on R. Then the surface area of S is

$$\text{Area}(S) = \iint_R \frac{\|\nabla f\|}{|\nabla f \cdot p|}\, dA.$$

Of course, whenever possible, we project onto the coordinate planes.

Example 10.10. Find the area of the surface cut from the bottom of the paraboloid $x^2 + y^2 = z$ by the plane $z = 4$.

The surface S is given implicitly by $f(x, y, z) = x^2 + y^2 - z = 0$. Projecting onto the xy-plane gives the region $R = \{(x, y) : x^2 + y^2 \leq 4\}$. We compute

$$\nabla f = \langle 2x,\ 2y,\ -1\rangle, \qquad \|\nabla f\| = \sqrt{1 + 4x^2 + 4y^2}.$$

Let the plane of projection be the xy-plane, whose unit normal is

$$p = \hat{k}.$$

Thus

$$\nabla f \cdot p = -1, \qquad |\nabla f \cdot p| = 1.$$

So the surface area is

$$\text{Area}(S) = \iint_R \frac{\|\nabla f\|}{|\nabla f \cdot p|}\, dA = \iint_R \sqrt{1 + 4x^2 + 4y^2}\, dA.$$

Switch to polar coordinates:

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad 0 \leq r \leq 2, \qquad 0 \leq \theta \leq 2\pi.$$

Thus,

$$\text{Area}(S) = \int_0^{2\pi} \int_0^2 \sqrt{1 + 4r^2}\, r\, dr\, d\theta.$$

Compute the radial integral:

$$\int_0^2 r\sqrt{1 + 4r^2}\, dr = \frac{1}{12}\left((1 + 16)^{3/2} - 1\right) = \frac{1}{12}\left(17^{3/2} - 1\right).$$

So the surface area is

$$\text{Area}(S) = 2\pi \cdot \frac{1}{12}\left(17^{3/2} - 1\right) = \frac{\pi}{6}\left(17^{3/2} - 1\right).$$

10.7 Integrating Over a Surface

Suppose a function $g(x, y, z)$ is defined over a surface S given by $f(x, y, z) = c$. To compute the integral of g over the surface (with area elements taken on the surface), we look at the region R on which this surface is defined as a function.

Divide the region R into smaller rectangles A_k. Consider the corresponding surface pieces S_k. Then

$$\text{Area}(S_k) \approx \left. \frac{\|\nabla f\|}{|\nabla f \cdot p|} \right|_{(x_k, y_k, z_k)} A_k.$$

Assuming that g is nearly constant on the smaller surface fragment S_k, we form the sum

$$\sum_k g(x_k, y_k, z_k) \left. \frac{\|\nabla f\|}{|\nabla f \cdot \mathbf{p}|} \right|_{(x_k, y_k, z_k)} A_k.$$

If this sum converges to a limit, then we define that limit as the integral of g over the surface S. Let S be a surface given by $f(x, y, z) = c$. Let the projection of S onto a plane with unit normal p be the region R. Let $g(x, y, z)$ be defined over S. Then the surface integral of g over S is

$$\iint_S g\, dS = \iint_R g(x, y, z) \frac{\|\nabla f\|}{|\nabla f \cdot p|}\, dA.$$

We also write the surface differential as

$$dS = \frac{\|\nabla f\|}{|\nabla f \cdot p|}\, dA.$$

Warning: $\nabla f \cdot p$ must *not* be zero.
If the surface S can be represented as a union of non-overlapping smooth surfaces $S_1, \ldots, S_n$, then

$$\iint_S g\, dS = \iint_{S_1} g\, dS + \cdots + \iint_{S_n} g\, dS.$$

If

$$g(x, y, z) = g_1(x, y, z) + \cdots + g_m(x, y, z)$$

over the surface S, then

$$\iint_S g\, dS = \iint_S g_1\, dS + \cdots + \iint_S g_m\, dS.$$

Similarly, if $g(x, y, z) = k\, h(x, y, z)$ for a constant k over S, then

$$\iint_S g(x, y, z)\, dS = k \iint_S h(x, y, z)\, dS.$$

Example 10.11. Integrate $g(x, y, z) = xyz$ over the surface of the cube cut from the first octant by the planes $x = 1$, $y = 1$, and $z = 1$.

We integrate g over the six surfaces and add the results. Since $g = xyz = 0$ on the coordinate planes, we only need integrals over sides A, B, C. Side A is the surface defined on the region

$$R_A = \{(x, y) : 0 \le x \le 1,\ 0 \le y \le 1\}$$

on the xy-plane. For this surface,

$$p = \hat{k},\ \nabla f = \hat{k},\ \|\nabla f\| = 1,\ \nabla f \cdot p = 1.$$

Here $g(x, y, z) = xyz$ and on this face $z = 1$, so $g = xy$. Therefore,

$$\iint_A g(x, y, z)\, dS = \iint_{R_A} xy \frac{\|\nabla f\|}{|\nabla f \cdot p|}\, dA = \int_0^1 \int_0^1 xy\, dx\, dy = \int_0^1 \frac{y}{2}\, dy = \frac{1}{4}.$$

Similarly,

$$\iint_B g(x, y, z)\, dS = \frac{1}{4},\ \iint_C g(x, y, z)\, dS = \frac{1}{4}.$$

Thus,

$$\iint_S g\, dS = \frac{3}{4}.$$

Example 10.12. Evaluate the surface integral of $g(x, y, z) = x^2$ over the unit sphere.

The sphere can be divided into the upper and lower hemispheres. Let S be the upper hemisphere:

$$f(x, y, z) := x^2 + y^2 + z^2 = 1,\ z \geq 0.$$

Its projection onto the xy-plane is the disk

$$R = \{(x, y) : x = r\cos\theta,\ y = r\sin\theta,\ 0 \leq r \leq 1,\ 0 \leq \theta \leq 2\pi\}.$$

Here,

$$p = \hat{k},\ \nabla f = \langle 2x, 2y, 2z\rangle,\ \|\nabla f\| = 2,$$
$$|\nabla f \cdot p| = |2z| = 2\sqrt{1 - r^2}.$$

Thus,

$$\iint_S x^2\, dS = \iint_R x^2 \frac{\|\nabla f\|}{|\nabla f \cdot p|}\, dA = \iint_R x^2 \frac{2}{2\sqrt{1 - r^2}}\, dA = \iint_R \frac{x^2}{\sqrt{1 - r^2}}\, dA.$$

Switch to polar coordinates:

$$x = r\cos\theta.$$

Therefore,

$$\iint_S x^2\, dS = \int_0^{2\pi} \int_0^1 \frac{r^2 \cos^2\theta}{\sqrt{1 - r^2}}\, r\, dr\, d\theta = \int_0^{2\pi} \cos^2\theta\, d\theta \int_0^1 \frac{r^3}{\sqrt{1 - r^2}}\, dr.$$

Now,

$$\int_0^{2\pi} \cos^2\theta\, d\theta = \pi,$$

and with $u = 1 - r^2$, $du = -2r\, dr$,

$$\int_0^1 \frac{r^3}{\sqrt{1 - r^2}}\, dr = \frac{2}{3}.$$

Thus,

$$\iint_S x^2\, dS = \pi \cdot \frac{2}{3} = \frac{2\pi}{3}.$$

10.8 Surface Integral of a Vector Field

A smooth surface is called *orientable* if it is possible to define a vector field of unit normal vectors $\hat{n}$ to the surface which varies continuously with position. Once such normal vectors are chosen, the surface is considered an *oriented surface*. If the surface S is given by $z = f(x, y)$ then we take its orientation by considering the unit normal

$$\hat{n} = \frac{\langle -f_x, -f_y, 1\rangle}{\sqrt{1 + f_x^2 + f_y^2}}.$$

If S is part of a level surface $g(x, y, z) = c$ then we may take

$$\hat{n} = \frac{\nabla g}{\|\nabla g\|}.$$

If S is given parametrically as

$$\vec{r}(u, v) = x(u, v)\,\hat{\imath} + y(u, v)\,\hat{\jmath} + z(u, v)\,\hat{k},$$

then

$$\hat{n} = \frac{\vec{r}_u \times \vec{r}_v}{\|\vec{r}_u \times \vec{r}_v\|}.$$

Sometimes we may take the negative sign if preferred. Conventionally, the outward direction is taken as the positive direction. Let $\vec{F}$ be a continuous vector field defined over an oriented surface S with unit normal $\hat{n}$. The surface integral of F over S, also called the *flux* of F across S, is

$$\iint_S F \cdot \hat{n}\, dS.$$

The flux is the integral of the scalar component of F along the unit normal to the surface. Thus in a flow, the flux is the net rate at which the fluid is crossing the surface S in the chosen positive direction. If S is part of a level surface $g(x, y, z) = c$ which is defined over the region D then

$$dS = \frac{\|\nabla g\|}{|\nabla g \cdot \hat{p}|}\, dA,$$

so the flux across S is

$$\iint_S F \cdot \hat{n}\, dS = \iint_S F \cdot \frac{\nabla g}{\|\nabla g\|}\, dS = \iint_D F \cdot \frac{\nabla g}{\nabla g \cdot \hat{p}}\, dA.$$

If S is parametrized by $\vec{r}(u, v)$ where D is the region in the uv-plane, then

$$dS = \|\vec{r}_u \times r_v\|\, dudv,$$

so the flux across S is

$$\iint_S F \cdot \hat{n}\, dS = \iint_D F(\vec{r}(u, v)) \cdot (r_u \times \vec{r}_v)\, dudv.$$

Example 10.13. Find the flux of $F = yz\,\hat{\imath} + z^2\,\hat{\jmath} + x\,\hat{k}$ outward through the surface S which is cut from the cylinder $y^2 + z^2 = 1$, $z \geq 0$, by the planes $x = 0$ and $x = 1$. S is given by

$g(x, y, z) = y^2 + z^2 - 1 = 0$ defined over the rectangle $R = R_{xy}$ as in the figure. The outward unit normal is

$$\hat{n} = \frac{\nabla g}{\|\nabla g\|} = \frac{\langle 0,\ 2y,\ 2z\rangle}{2\sqrt{y^2+z^2}} = \langle 0,\ y,\ z\rangle.$$

Here $\hat{p} = \hat{k}$, so

$$dS = \frac{\|\nabla g\|}{|\nabla g \cdot \hat{k}|}\, dA = \frac{\sqrt{(2y)^2+(2z)^2}}{|2z|}\, dA = \frac{1}{z}\, dA.$$

Now

$$F \cdot \hat{n} = yz^2 + z^3 = z(y^2+z^2) = z.$$

Therefore the outward flux through S is

$$\iint_S F \cdot \hat{n}\, dS = \iint_R z\left(\frac{1}{z}\right) dA = \iint_R 1\, dA = \text{Area}(R) = 2.$$

Example 10.14. Find the flux of the vector field $F = \langle z,\ y,\ x\rangle$ across the unit sphere. If no direction of the normal vector is given and the surface is a closed surface, we take $\hat{n}$ in the positive (outward) direction. Using spherical coordinates, the unit sphere S is parametrized by

$$\vec{r}(\theta, \phi) = \sin\phi\cos\theta\,\hat{\imath} + \sin\phi\sin\theta\,\hat{\jmath} + \cos\phi\,\hat{k},$$

where $0 \le \theta \le 2\pi$ and $0 \le \phi \le \pi$, giving the region D. Then

$$\vec{F}(\vec{r}) = \langle \cos\phi,\ \sin\phi\sin\theta,\ \sin\phi\cos\theta\rangle,$$

and

$$r_\theta \times r_\phi = \sin^2\phi\cos\theta\,\hat{\imath} + \sin^2\phi\sin\theta\,\hat{\jmath} + \sin\phi\cos\phi\,\hat{k}.$$

Consequently,

$$\iint_S F \cdot \hat{n}\, dS = \iint_D F(r) \cdot (r_\theta \times r_\phi)\, d\theta\, d\phi,$$

$$= \int_0^{2\pi}\int_0^{\pi} \left(2\sin^2\phi\cos\theta\cos\phi + \sin^2\phi\sin^2\phi\right) d\phi\, d\theta = \frac{4}{3}.$$

10.9 Stokes' Theorem

Consider an oriented surface with a unit normal vector $\hat{n}$. Call the boundary curve of S as C. The orientation of S induces an orientation on C. We say that C is *positively oriented* if and only if whenever you walk in the positive direction of C keeping your head pointing towards $\hat{n}$, the surface S will be to your left. Recall that Green's theorem relates a double integral in the plane to a line integral over its boundary. We will have a generalization of this to three dimensions. Write the boundary curve of a given smooth surface as C. The boundary is assumed to be a closed curve, positively oriented unless specified otherwise.

Theorem 10.6. Let S be an oriented piecewise-smooth surface that is bounded by a simple, closed, piecewise-smooth boundary curve C with positive orientation. Let $F = M\,\hat{\imath} + N\,\hat{\jmath} + P\,\hat{k}$ be a vector field with M, N, P having continuous partial derivatives on an open region in space that contains S. Then

$$\oint_C F \cdot dr = \iint_S (\nabla \times F) \cdot \hat{n}\, dS.$$

If S lies in the xy-plane, then $\hat{n} = \hat{k}$ and $dS = dA$, so we obtain

$$\iint_S (\nabla \times F) \cdot \hat{n}\, dS = \iint_D (\nabla \times F) \cdot \hat{k}\, dA = \iint_D (N_x - M_y)\, dx\, dy,$$

as Green's theorem states. In fact, we can use Green's theorem to prove Stokes' theorem in the case where S is the graph of a smooth function $z = f(x, y)$ with a smooth boundary, and the vector field F is smooth.

Proof. Let $F = M\,\hat{i} + N\,\hat{j} + P\,\hat{k}$. We see that

$$\iint_S F \cdot dr = \iint_S (M\,dx + N\,dy + P\,dz).$$

And

$$\iint_S (\nabla \times F) \cdot \hat{n}\, dS = \iint_S (\nabla \times (M\hat{i})) \cdot \hat{n}\, dS + \iint_S (\nabla \times (N\hat{j})) \cdot \hat{n}\, dS + \iint_S (\nabla \times (P\hat{k})) \cdot \hat{n}\, dS.$$

We show that the M-, N-, and P-components in both are equal. Assume S is given by $z = f(x, y)$ for $(x, y) \in D$. Orient D positively, i.e., counterclockwise. Choose a parameterization of the boundary of D:

$$r(t) = x(t)\,\hat{i} + y(t)\,\hat{j},\ \ a \le t \le b.$$

Then the surface S has the parameterization

$$r(t) = x(t)\,\hat{i} + y(t)\,\hat{j} + f(x(t), y(t))\,\hat{k},\ \ a \le t \le b.$$

Thus

$$\int_S M(x, y, z)\, dx = \int_a^b M\big(x(t), y(t), f(x(t), y(t))\big)\, x'(t)\, dt.$$

Or equivalently,

$$\int_S M(x, y, z)\, dx = \int_{\partial D} M(x, y, f(x, y))\, dx.$$

Next, we apply Green's theorem on the integral on the right to obtain:

$$\int_S M(x, y, z)\, dx = \iint_D M_y(x, y, f(x, y))\, dA.$$

Applying the chain rule to the right-hand integrand gives

$$\iint_D M_y(x, y, f(x, y))\, dA = \iint_D \big(M_y(x, y, f(x, y)) + M_z(x, y, f(x, y))\, f_y\big)\, dA.$$

Now we compute $\displaystyle\iint_S (\nabla \times (M\hat{i})) \cdot \hat{n}\, dS$. Note that S has the parameterization

$$r(x, y) = x\,\hat{i} + y\,\hat{j} + f(x, y)\,\hat{k}.$$

So its (unnormalized) normal vector is $n = \langle -f_x,\, -f_y,\, 1\rangle$. Then $\nabla \times (M\hat{i}) = \langle 0,\ M_z,\ M_y\rangle$, hence

$$(\nabla \times (M\hat{i})) \cdot n = M_z\,(-f_y) + M_y(1) = M_y - M_z f_y.$$

Therefore,
$$\iint_S (\nabla \times (M\hat{\imath})) \cdot \hat{n}\, dS = \iint_D (M_y + M_z f_y)\, dA.$$
Thus,
$$\iint_S (\nabla \times (M\hat{\imath})) \cdot \hat{n}\, dS = \int_S M(x,y,z)\, dx.$$
Similarly, the N- and P-components become respectively equal. □

Example 10.15. Consider S as the hemisphere $x^2 + y^2 + z^2 = 9,\ z \geq 0$. Let $F(r) = y\,\hat{\imath} - x\,\hat{\jmath}$. The bounding curve for S in the xy-plane is

$$C:\ x^2 + y^2 = 9,\ z = 0.$$

A parameterization of C is

$$r(\theta) = 3\cos\theta\,\hat{\imath} + 3\sin\theta\,\hat{\jmath},\ 0 \leq \theta \leq 2\pi.$$

Then
$$\begin{aligned}\oint_C F \cdot dr &= \int_0^{2\pi} [(3\sin\theta)(-3\sin\theta) - (3\cos\theta)(3\cos\theta)]\, d\theta \\ &= \int_0^{2\pi} \big(-9\sin^2\theta - 9\cos^2\theta \big)\, d\theta \\ &= -18\pi.\end{aligned}$$
This is the line integral in Stokes' Theorem. For the surface integral, compute the curl:

$$\nabla \times F = (P_y - N_z)\,\hat{\imath} + (M_z - P_x)\,\hat{\jmath} + (N_x - M_y)\,\hat{k} = -2\,\hat{k}.$$

On the surface $g(x,y,z) = x^2 + y^2 + z^2 - 9 = 0$,

$$\hat{n} = \frac{\nabla g}{\|\nabla g\|} = \frac{1}{3}(x\,\hat{\imath} + y\,\hat{\jmath} + z\,\hat{k}).$$

Here the projection plane is the xy-plane, so

$$\hat{p} = \hat{k}.$$

Thus the surface differential is
$$dS = \frac{\|\nabla g\|}{\nabla g \cdot \hat{p}}\, dA = \frac{3}{z}\, dA.$$
Therefore,
$$(\nabla \times F) \cdot \hat{n} = 2z/3.$$
Hence
$$\iint_S (\nabla \times F) \cdot \hat{n}\, dS = \iint_D \frac{2z}{3} \cdot \frac{3}{z}\, dA = -\iint_D 2\, dA = -18\pi.$$

Example 10.16. Evaluate $\oint_C \big((x^2 - y)\,\hat{\imath} + 4z\,\hat{\jmath} + x^2\,\hat{k} \big) \cdot d\vec{r}$ where C is the intersection of the plane $z = 2$ and the cone $z = x^2 + y^2$.

Parameterize the cone (instead of the usual r, use θ) as

$$r(r,\theta) = r\cos\theta\,\hat{\imath} + r\sin\theta\,\hat{\jmath} + r^2\,\hat{k},\ 0 \leq r \leq 2,\ 0 \leq \theta \leq 2\pi.$$

Then

$$F = (x^2 - y)\,\hat{\imath} + 4z\,\hat{\jmath} + x^2\,\hat{k}.$$

The oriented normal is

$$n = \frac{r_r \times r_\theta}{\|r_r \times r_\theta\|} = \frac{1}{2}\big(- \cos\theta\,\hat{\imath} - \sin\theta\,\hat{\jmath} + \hat{k}\big).$$

Compute the curl:

$$\nabla \times F = (P_y - N_z)\,\hat{\imath} + (M_z - P_x)\,\hat{\jmath} + (N_x - M_y)\,\hat{k} = -4\,\hat{\imath} - 2r\cos\theta\,\hat{\jmath} + \hat{k}.$$

Then

$$(\nabla \times F) \cdot n = \frac{1}{\sqrt{2}}\big(4\cos\theta + r\sin(2\theta) + 1\big).$$

The surface element is

$$dS = r\,dr\,d\theta \;\Rightarrow \quad dS = 2\,dr\,d\theta.$$

By Stokes' theorem,

$$\oint_C F \cdot dr = \iint_S (\nabla \times F) \cdot n\,dS = \int_0^{2\pi} \int_0^2 \big(4\cos\theta + \sin(2\theta) + 1\big)\,2\,dr\,d\theta = 4\pi.$$

Recall that a conservative field is one which can be expressed as the gradient of another scalar field. In such a case,

$$\nabla \times F = 0.$$

Then, from Stokes' theorem, it follows that $\oint_C F \cdot dr = 0$.

Theorem 10.7. If $\nabla \times F = 0$ at each point of an open simply connected region D in space, then on any piecewise smooth closed path C lying in D, $\oint_C F \cdot dr = 0$.

10.10 Gauss' Divergence Theorem

We have seen how to relate an integral of a function over a region with the integral of (possibly) some other related function over the boundary of the region. For definite integrals on intervals:

$$\int_a^b f(t)\,dt = f(b) - f(a).$$

For a path from a point P to a point Q in $\mathbb{R}^3$:

$$\int_C \nabla f \cdot dr = f(Q) - f(P).$$

For a region D in $\mathbb{R}^2$:

$$\iint_D (N_x - M_y)\,dA = \oint_{\partial D} F \cdot dr.$$

For a surface S in $\mathbb{R}^3$:

$$\iint_S (\nabla \times F) \cdot \hat{n}\,dS = \oint_{\partial S} F \cdot dr.$$

This suggests a generalization to three dimensions, and we use the divergence of a vector field for this purpose. Recall that $\operatorname{div} F = \nabla \cdot F$. That is, for a vector field $F = M(x,y,z)\,\hat{\imath} + N(x,y,z)\,\hat{\jmath} + P(x,y,z)\,\hat{k}$, the divergence is the scalar function $\operatorname{div} F = M_x + N_y + P_z$. Our generalization is

$$\iiint_D (\operatorname{div} F)\,dV = \iint_S F \cdot \hat{n}\,dS.$$

Theorem 10.8. Let S be a piecewise smooth, simple, closed, bounded surface that encloses a solid region D in $\mathbb{R}^3$. Suppose S has been oriented positively by its outward normals. Let F be a vector field whose component functions have continuous partial derivatives on an open region that contains D. Then

$$\iint_S F \cdot \hat{n}\, dS = \iiint_D (\operatorname{div} F)\, dV.$$

Proof. We prove the divergence theorem in the special case that D is a box in $\mathbb{R}^3$ given by

$$D = [a, b] \times [c, d] \times [e, f].$$

Let

$$F = M\,\hat{\imath} + N\,\hat{\jmath} + P\,\hat{k}.$$

Then

$$\begin{aligned}
\iint_S F \cdot \hat{n}\, dS &= \int_e^f \int_c^d M(b, y, z)\, dy\, dz - \int_e^f \int_c^d M(a, y, z)\, dy\, dz \\
&= \int_e^f \int_c^d \big[\, M(b, y, z) - M(a, y, z) \,\big]\, dy\, dz \\
&= \int_e^f \int_c^d \int_a^b M_x(x, y, z)\, dx\, dy\, dz \\
&= \iiint_D \operatorname{div} F\, dV.
\end{aligned}$$

We prove that the respective components are equal. Thus, we consider only the $\hat{\imath}$-component, i.e. we take $F = M\,\hat{\imath}$, and prove the divergence theorem in this case. The solid D has six faces. The surface integral over S is the sum of integrals over these faces. A simplification occurs: since $F = M\hat{\imath}$, $F \cdot \hat{k} = 0$, so F is orthogonal to the normals of the top, bottom, and the two side faces. Let the remaining faces be S_f (front face) and S_b (back face). Thus,

$$\iint_S F \cdot \hat{n}\, dS = \iint_{S_f} F \cdot \hat{n}\, dS + \iint_{S_b} F \cdot \hat{n}\, dS.$$

A parameterization of the faces is:

$$S_f : r(y, z) = (b,\ y,\ z),\ \ S_b : r(y, z) = (a,\ y,\ z),$$

for $c \le y \le d$ and $e \le z \le f$. The outward normal to S_f is $\hat{\imath}$ and to S_b is $-\hat{\imath}$. Thus,

$$\iint_S F \cdot \hat{n}\, dS = \int_e^f \int_c^d M(b, y, z)\, dy\, dz - \int_e^f \int_c^d M(a, y, z)\, dy\, dz.$$

Therefore,

$$\iint_S F \cdot \hat{n}\, dS = \int_e^f \int_c^d \big[M(b, y, z) - M(a, y, z)\big]\, dy\, dz.$$

Now integrating with respect to x:

$$\int_e^f \int_c^d \int_a^b M_x(x, y, z)\, dx\, dy\, dz = \iiint_D M_x\, dV = \iiint_D \operatorname{div} F\, dV,$$

since for $F = M\hat{\imath}$ we have

$$\operatorname{div} F = M_x.$$

Thus, the divergence theorem holds in this special case. ☐

Example 10.17. Consider the field $F = x\,\hat{\imath} + y\,\hat{\jmath} + z\,\hat{k}$ over the sphere

$$S : x^2 + y^2 + z^2 = a^2.$$

The outer unit normal to S, computed from $f = x^2 + y^2 + z^2 - a^2$, is

$$\hat{n} = \frac{\nabla f}{\|\nabla f\|} = \frac{2(x\,\hat{\imath} + y\,\hat{\jmath} + z\,\hat{k})}{2\sqrt{x^2 + y^2 + z^2}} = \frac{1}{a}(x\,\hat{\imath} + y\,\hat{\jmath} + z\,\hat{k}).$$

Hence, on the given surface,

$$F \cdot \hat{n}\,dS = \frac{1}{a}(x^2 + y^2 + z^2)\,dS = a\,dS.$$

Therefore,

$$\iint_S F \cdot \hat{n}\,dS = \iint_S a\,dS = a \cdot \text{Area}(S) = a(4\pi a^2) = 4\pi a^3.$$

Now consider the triple integral. Since

$$\text{div}\,F = M_x + N_y + P_z = x_x + y_y + z_z = 3,$$

we have, with D the ball bounded by S,

$$\iiint_D \text{div}\,F\,dV = \iiint_D 3\,dV = 3 \cdot \text{Volume}(D) = 3\left(\frac{4\pi a^3}{3}\right) = 4\pi a^3.$$

Example 10.18. Find the outward flux of the vector field $F = xy\,\hat{\imath} + yz\,\hat{\jmath} + zx\,\hat{k}$ through the surface cut from the first octant by the planes $x = 1$, $y = 1$, and $z = 1$.

The solid D is a cube having six faces. Call the surface of the cube S. Instead of computing the surface integral, we use the Divergence Theorem. With $F = xy\,\hat{\imath} + yz\,\hat{\jmath} + zx\,\hat{k}$, we have

$$\nabla \cdot F = \frac{\partial(xy)}{\partial x} + \frac{\partial(yz)}{\partial y} + \frac{\partial(zx)}{\partial z} = y + z + x.$$

Therefore, the required flux is

$$\iint_S F \cdot \hat{n}\,dS = \iiint_D (\nabla \cdot F)\,dV = \int_0^1 \int_0^1 \int_0^1 (x + y + z)\,dx\,dy\,dz = \frac{3}{2}.$$

References

[1] T. M. Apostol, *Calculus Volume I*, Second Edition, Wiley 1991.

[2] R. G. Bartle, Donald R. Sherbert, *Introduction to Real Analysis*, Third Edition, John Wiley and Sons, 1999.

[3] E. Kreyszig, *Advanced Engineering Mathematics*, 10th ed., John Wiley & Sons, 2010.

[4] A. Kumar, S. Kumaresan, *A Basic Course in Real Analysis*, CRC Press, Taylor and Francis Group, A Chapman and Hall Book, 2014.

[5] S. K. Mapa, *Higher Algebra Abstract and Linear*, Levant Books, Eleventh Edition, 2012.

[6] S. K. Mapa, *Introduction to Real Analysis*, Levant Books, Revised 6th Edition, 2012.

[7] J. E. Marsden, A. J. Tromba, and A. Weinstein, *Basic Multivariable Calculus*, Springer-Verlag, 1993.

[8] N. Piskunov, *Differential and Integral Calculus*, Vols. I and II, Mir Publishers, 1974.

[9] A. Singh, *Classnotes - MA1101, Functions of Several Variables*, Indian Institute of Technology Madras.

[10] G. Strang, *Calculus*, Wellesley–Cambridge Press, 2010.

[11] G.B. Thomas Jr., M.D. Weir, and J.R. Hass, *Thomas Calculus*, Pearson, 2009.