UNIT I

Definition and existence of Riemann-Stieltjes

Introduction: The reader will recall from elementary calculus that to find the area of the region under the graph of a positive function \( f \) defined on \([a, b]\), we subdivide the interval \([a, b]\) into a finite number of subintervals, say \( n \), the \( k^{th} \) subinterval having length \( \Delta x_k \), and we consider sums of the form
\[
\sum_{k=1}^{n} f(t_k) \Delta x_k
\]
where \( t_k \) is some point in the \( k^{th} \) subinterval. Such a sum is an approximation to the area by means of rectangles. If \( f \) is sufficiently well behaved in \([a, b]\) continuous. The two concepts, derivative and integral, arise in entirely different ways and it is a remarkable fact indeed that the two are intimately connected. If we consider the definite integral of a continuous function \( f \) as a function of its upper limit, say we write
\[
F(x) = \int_a^x f(t) \, dt.
\]
Then \( F \) has a derivative and \( F'(x) = f(x) \). This important result shows that differentiation and integration are, in a sense, inverse operations.

In this unit we study the process of integration in some detail. Actually we consider a more general concept than that of Riemann namely Riemann-Stieltjes integral, which involves two functions \( f \) and \( \alpha \). The symbol for such an integral is
\[
\int_a^b f(x) \, d\alpha(x)
\]
or something similar, and the usual Riemann integral occurs as the special case in which \( \alpha(x) = x \). When \( \alpha \) has a continuous derivative, the definition is such that the Stieltjes integral becomes the Riemann integral. However, the Stieltjes integral still makes sense when \( \alpha \) is not differentiable or even when \( \alpha \) is discontinuous.

Problems in physics which involve mass distributions that are partly discrete and partly continuous can also be treated by using Stieltjes integrals. In the mathematical theory of probability this integral is a very useful tool that makes possible the simultaneous treatment of continuous and discrete random variables.

Objective: The Riemann-Stieljes Integral is based on the definition of Riemann Integral which we had studied in previous classes for the sake of convenience we are giving the definition and preliminaries of Riemann Integrals. Accordingly, we begin by discussing integration of real-valued functions on intervals.

1.1. The Riemann-Stieljes Integral.

1.1.1 Definitions: Let \([a, b]\) be a given interval. Then a set \( P = \{x_0, x_1, ..., x_{n-1}, x_n\} \) of \([a, b]\) such that \( a = x_0 \leq x_1 \leq ... \leq x_{n-1} \leq x_n = b \) is said to be a Partition of \([a, b]\). The set of all partitions of \([a, b]\) is denoted by \( P([a, b]) \). The intervals \([x_0, x_1], [x_0, x_1], [x_1, x_2], ..., [x_{n-1}, x_n]\) are called the subintervals of \([a, b]\). Write \( \Delta x_i = x_i - x_{i-1} \) is called the length of the interval \([x_i, x_i]\) \((i = 1, ..., n)\) and max \( |\Delta x_i| \) is called the norm of the partition \( P \) and is denoted by \( ||P|| \) or \( \mu(P) \). A partition \( Q \) of \([a, b]\) such that \( P \subset Q \) is called the refinement or finer of the partition \( P \).

Suppose \( f \) is a bounded real valued function defined on \([a, b]\) and
\[ M_i = \sup f(x), \quad m_i = \inf f(x) \quad (x_{i-1} \leq x_i) \text{ for each } P \in P([a, b]). \]

Then \( U(P, f) = \sum_{i=1}^{n} M_i \Delta x_i \) and \( L(P, f) = \sum_{i=1}^{n} m_i \Delta x_i \) are called the **Upper and Lower Riemann sums** or **Upper and Lower Darboux sums** of \( f \) on \([a, b]\) with respect to the partition \( P \). Further write

\[
\int_{a}^{b} f \, dx = \inf U(P, f) \quad \text{and} \quad \int_{a}^{b} f \, dx = \sup L(P, f)
\]

where the inf and the sup are taken over all partitions \( P \) of \([a, b]\) are called the **Upper and Lower Riemann integrals** of \( f \) over \([a, b]\), respectively. If the upper and lower Riemann integrals are equal, we say that \( f \) is **Riemann-integrable** on \([a, b]\) and we write \( f \in R[a, b] \) and we denote the common value of these integrals by

\[
\int_{a}^{b} f \, d(x),
\]

i.e.,

\[
\int_{a}^{b} f \, dx = \int_{a}^{b} f \, dx = \int_{a}^{b} f \, dx.
\]

**Another definition of Riemann Integral**: A bounded function \( f : [a, b] \to R \) is Riemann Integrable if for \( \epsilon > 0 \) there exists \( \delta > 0 \) such that for any partition \( P = \{x_0, x_1, x_2, \ldots x_n\} \) with \( ||P|| < \delta \) and \( \xi_i \in [x_{i-1}, x_i] \)

\[
|\sum f(\xi_i)(x_i-x_{i-1}) - \int f| < \epsilon.
\]

**1.1.1. Lemma.** If \( f : [a, b] \to R \) is a bounded function then the upper and lower Riemann integrals of \( f \) are bounded. Since \( f \) is bounded, there exist two numbers \( m \) and \( M \) such that \( m \leq f(x) \leq M \quad (a \leq x \leq b) \). Hence, for every partition \( P \) of \([a, b]\) we have

\[
m \leq m_i \leq M_i \leq M
\]

\[
\Rightarrow \quad \sum m_i \Delta x_i \leq \sum m_i \Delta x_i \leq \sum M_i \Delta x_i \leq \sum M \Delta x_i, \quad i = 1, 2, 3, \ldots, n.
\]

\[
\Rightarrow \quad m(b-a) \leq L(P, f) \leq U(P, f) \leq M(b-a),
\]

so that the numbers \( L(P, f) \) and \( U(P, f) \) form a bounded set. Therefore by the definition of lower and upper Riemann integrals this shows that the upper and lower integrals are defined for every bounded function \( f \) are bounded also.

The question of their equality, and hence the question of the integrability of \( f \), is a more delicate one. Here we state some other investigations of Riemann integrals.

**1.1.2. Lemma.** If \( f : [a, b] \to R \) is a bounded function, \( P \) is any partition of \([a, b]\) and \( P^* \) is the refinement of \( P \), then

\[
L(P, f) \leq L(P^*, f) \quad \text{and} \quad U(P^*, f) \leq U(P, f).
\]

**1.1.3. Lemma.** If \( f : [a, b] \to R \) is a bounded function and \( P_1, P_2 \) are any two partitions of \([a, b]\) then

\[
L(P_1, f) \leq U(P_2, f) \quad \text{and} \quad L(P_2, f) \leq U(P_1, f).
\]
1.1.4. Lemma. If \( f, g : [a, b] \to \mathbb{R} \) are bounded functions and \( P \) is any partition of \( [a, b] \) then
\[
\begin{align*}
(i) & \quad L(P, f + g) \geq L(P, f) + L(P, g) \\
(ii) & \quad U(P, f + g) \leq U(P, f) + U(P, g).
\end{align*}
\]

1.1.1. Theorem. If \( f : [a, b] \to \mathbb{R} \) is bounded function then
\[
\int_{a}^{b} f \, dx \geq \int_{-a}^{-b} f \, dx.
\]

1.1.2. Theorem (Darboux). If \( f : [a, b] \to \mathbb{R} \) is bounded function then for \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that
\[
U(P, f) < \int_{-a}^{b} f \, dx + \varepsilon \quad \text{and} \quad L(P, f) > \int_{-a}^{b} f \, dx - \varepsilon.
\]

1.1.3. Theorem. If \( f : [a, b] \to \mathbb{R} \) is bounded function is Riemann Integrable if the oscillatory sum < \( \varepsilon \), i.e. \( \omega(P, f) = U(P, f) - L(P, f) < \varepsilon \), for \( \varepsilon > 0 \) and any partition \( P \) of \( [a, b] \).

1.1.4. Theorem. Every continuous function \( f : [a, b] \to \mathbb{R} \) is Riemann Integrable.

1.1.5. Theorem. Every monotone function \( f : [a, b] \to \mathbb{R} \) is Riemann Integrable.

Students you studied the properties given above and other properties of Riemann Integrals in previous classes therefore we are not interested to investigate these here. However we shall immediately consider a more general situation.

1.1.2 Definition. Let \( f : [a, b] \to \mathbb{R} \) is bounded function and \( \alpha \) be a monotonically increasing function on \( [a, b] \). Let \( P = \{x_0, x_1, \ldots, x_{n-1}, x_n\} \) such that \( a = x_0 \leq x_1 \leq \ldots \leq x_{n-1} \leq x_n = b \) be any Partition of \([a, b]\). We write
\[
\Delta \alpha_i = \alpha(x_i) - \alpha(x_{i-1}), \quad i = 1, 2, 3, \ldots, n.
\]
By the definition of monotone function \( \alpha(a) \) and \( \alpha(b) \) are finite therefore \( \alpha \) is bounded on \( [a, b] \), also since \( \alpha \) is monotonically increasing function then clearly \( \Delta \alpha_i \geq 0 \), \( i = 1, 2, 3, \ldots, n \).

Let \( M_i = \sup f(x) \), \( m_i = \inf f(x) \) \( (x_{i-1} \leq x \leq x_i) \) for each \( P \in \mathcal{P}( [a, b] ) \). We define
\[
U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta \alpha_i \quad \text{and} \quad L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta \alpha_i
\]
are called the Upper and Lower Riemann Stieltjes sums respectively. Further we define
\[
\int_{a}^{b} f \, d\alpha = \inf U(P, f, \alpha) \quad \text{and} \quad \int_{-a}^{b} f \, d\alpha = \sup L(P, f, \alpha),
\]
where the inf and the sup are taken over all partitions \( P \) of \([a, b]\), are called the Upper and Lower Riemann Stieltjes integrals of \( f \) over \([a, b]\), respectively. If the upper and lower Riemann Stieltjes integrals are equal, we say that \( f \) is Riemann Stieltjes integrable on \([a, b]\) and we write \( f \in R(\alpha) \) and denote the common value of these integrals by
\[
\int_{a}^{b} f \, d\alpha \quad \text{or} \quad \int_{-a}^{b} f(x) \, d\alpha(x).
\]
1.1.6. Theorem: Hence and refinement of $P$

Proof. If $f: [a, b] \to R$ is a bounded function and $\alpha$ be a monotonically increasing function on $[a, b]$. Let $P$ be any partition of $[a, b]$. Then the upper and lower Riemann-Stieltjes integrals of $f$ with respect to $\alpha$ are bounded.

**Proof.** Since $f$ is bounded, there exist two numbers $m$ and $M$ such that $m \leq f(x) \leq M$ ($a \leq x \leq b$). Hence, for every partition $P$ of $[a, b]$ we have

$$m \leq m_i \leq M_i \leq M$$

$$\Rightarrow \quad \sum m_i \Delta \alpha_i \leq \sum M_i \Delta \alpha_i \leq \sum M \Delta \alpha_i, \quad i = 1, 2, 3, ..., n.$$ 

$$\Rightarrow \quad m[\alpha(b) - \alpha(a)] \leq L(P, f, \alpha) \leq U(P, f, \alpha) \leq M[\alpha(b) - \alpha(a)],$$

so that the numbers $L(P, f, \alpha)$ and $U(P, f, \alpha)$ form a bounded set. Therefore by the definition of lower and upper Riemann-Stieltjes integrals this shows that the upper and lower integrals are defined for every bounded function $f$ are bounded also.

1.1.5. **Lemma.** If $P^*$ is a refinement of the partition $P$ of $[a, b]$, then

$$L(P, f, \alpha) \leq L(P^*, f, \alpha) \quad \text{and} \quad U(P^*, f, \alpha) \leq U(P, f, \alpha).$$

**Proof.** Let $P = \{x_0, x_1, ..., x_n, \} / \text{such that } a = x_0 \leq x_1 \leq ... \leq x_{n-1} = x_n = b$ be any partition of $[a, b]$ and $P^*$ the refinement of $P$ contains just one point $x^*$ more than $P$ such that $x_{i-1} < x^* < x_i$ where $x_{i-1}$ and $x_i$ are two consecutive points of $P$. Let $m_i, m'_i, m''_i$ are the infimum of $f(x)$ in $[x_{i-1}, x_i], [x_{i-1}, x^*]$ and $[x^*, x_i]$ respectively then clearly $m_i \leq m'_i$ and $m_i \leq m''_i$. Therefore

$$L(P^*, f, \alpha) - L(P, f, \alpha) = m_i' \left[ \alpha(x^*) - \alpha(x_{i-1}) \right] + m''_i \left[ \alpha(x_i) - \alpha(x^*) \right] - m_i \left[ \alpha(x_i) - \alpha(x_{i-1}) \right]$$

$$= m_i' \left[ \alpha(x^*) - \alpha(x_{i-1}) \right] + m''_i \left[ \alpha(x_i) - \alpha(x^*) \right] - m_i \left[ \alpha(x_i) - \alpha(x_{i-1}) \right]$$

$$= (m_i' - m_i) \left[ \alpha(x^*) - \alpha(x_{i-1}) \right] + (m''_i - m_i) \left[ \alpha(x_i) - \alpha(x^*) \right] \geq 0.$$ 

Hence $L(P, f, \alpha) \leq L(P^*, f, \alpha)$. If $P^*$ contains $k$ points more than $P$ then by repeating the same process we arrive at the same result. The proof of $U(P^*, f, \alpha) \leq U(P, f, \alpha)$ is analogous.

1.1.7. **Lemma.** If $P_1, P_2$ are any two partitions of $[a, b]$ then $L(P_1, f, \alpha) \leq L(P_2, f, \alpha)$ and

$$L(P_2, f, \alpha) \leq U(P_1, f, \alpha).$$

**Proof.** We have $P_1, P_2$ are any two partitions of $[a, b]$ . Let $P = P_1 \cup P_2$, then $P$ is the common refinement of $P_1, P_2$. Then $L(P_1, f, \alpha) \leq L(P, f, \alpha)$; $U(P, f, \alpha) \leq U(P_1, f, \alpha)$, and

$$L(P_2, f, \alpha) \leq L(P, f, \alpha) \quad \text{and} \quad U(P, f, \alpha) \leq U(P_2, f, \alpha).$$

Hence $L(P_1, f, \alpha) \leq L(P_2, f, \alpha)$ and $L(P_2, f, \alpha) \leq U(P_1, f, \alpha)$.

1.1.6. **Theorem:**

$$\int_a^b f \, da \leq \int_a^b f \, da$$

**Proof.** Since if $P$ is the common refinement of two partitions $P_1$ and $P_2$ of $[a, b]$. Then
\[ L(P_1, f, \alpha) \leq L(P, f, \alpha) \leq U(P, f, \alpha) \leq U(P_2, f, \alpha). \]

Therefore for all partitions \( P_1 \) and \( P_2 \) of \([a, b]\), \[ \sup_{a}^{b} \int f \, d\alpha \leq \inf_{a}^{b} \int f \, d\alpha \]

**1.1.7. Theorem.** \( f \in R(\alpha) \) on \([a, b]\) if and only if for every \( \varepsilon > 0 \) there exists a partition \( P \) such that
\[ U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon \quad \text{(i)} \]

**Proof.** Let for a partition \( P \) of \([a, b]\), (i) holds. Then by the definition of Riemann Stietjes integrals
\[ L(P, f, \alpha) \leq \int f \, d\alpha \leq U(P, f, \alpha). \]

This implies that
\[ 0 \leq \int f \, d\alpha - \int f \, d\alpha \leq U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon \]

Hence
\[ \int f \, d\alpha = \int f \, d\alpha \quad \text{i.e., that is} \quad f \in R(\alpha). \]

Conversely suppose that \( f \in R(\alpha) \), then \[ \int f \, d\alpha = \int f \, d\alpha = \int f \, d\alpha. \] Now if \( \varepsilon > 0 \) be given by definition of Riemann Stietjes integrals there exist partitions \( P_1 \) and \( P_2 \) such that
\[ \int f \, d\alpha < L(P_1, f, \alpha) + \varepsilon/2 \quad \text{and} \quad U(P_2, f, \alpha) - \varepsilon/2 < \int f \, d\alpha. \]

Let \( P = P_1 \cup P_2 \) be the common refinement of \( P_1 \) and \( P_2 \). Then
\[ U(P, f, \alpha) - L(P, f, \alpha) < \int f \, d\alpha + \varepsilon/2 < L(P_1, f, \alpha) + \varepsilon \leq \int f \, d\alpha + \varepsilon. \]

Yields,
\[ U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon. \]

**1.1.8. Theorem.** (a) If for every \( \varepsilon > 0 \) there exists a partition \( P \) such that \( U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon \) holds then it is holds (with the same \( \varepsilon \)) for every refinement of \( P \).

(b) If for every \( \varepsilon > 0 \) there exists a partition \( P = \{x_0, x_1, \ldots, x_n\} \) such that \( U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon \) holds and if \( s_i, t_i \) are arbitrary points in \([x_{i-1}, x_i]\), then
\[ \sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta \alpha_i < \varepsilon. \]

(c) If \( f \in R(\alpha) \) and the hypotheses of (b) hold, then
\[ \sum_{i=1}^{n} |f(t_i)\Delta \alpha_i - \int f \, d\alpha| < \varepsilon. \]

**Proof: (a)** Let \( P^* \) be any refinement of the partition \( P \), then we know that
\[ L(P, f, \alpha) \leq L(P^*, f, \alpha) \quad \text{and} \quad U(P^*, f, \alpha) \leq U(P, f, \alpha), \]
therefore
\[ U(P^*, f, \alpha) - L(P^*, f, \alpha) \leq U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon. \]

Since \( P^* \) is arbitrary, hence the conclusion holds for all refinements of \( P \).

(b) Let \( M_i = \sup \{f(x) \mid x_{i-1} \leq x \leq x_i\} \) then for arbitrary points \( s_i, t_i \) in \([x_{i-1}, x_i] \),
\[ m_i \leq f(s_i) \leq M_i \quad \text{and} \quad m_i \leq f(t_i) \leq M_i, \]
so that \[ |f(s_i) - f(t_i)| \leq M_i - m_i. \]
Thus
\[ \sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta \alpha_i \leq \sum_{i=1}^{n} (M_i - m_i) |\Delta \alpha_i| = U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon. \]
(c) Since if \( f \in R(\alpha) \) then by the obvious inequalities
\[
L(P, f, \alpha) \leq \sum f(t_i) \Delta \alpha_i \leq U(P, f, \alpha) \quad \text{and} \quad L(P, f, \alpha) \leq \int f \, d\alpha \leq U(P, f, \alpha),
\]
implies that
\[
\bigg| \sum_{i=1}^{n} f(t_i) \Delta \alpha_i - \int_{a}^{b} f \, d\alpha \bigg| \leq U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon.
\]

1.1.9. Theorem. If \( f \) is continuous on \([a, b]\) and \( \alpha \) is monotonic increasing on \([a, b]\) then \( f \in R(\alpha) \) on \([a, b]\).

**Proof.** Let \( \varepsilon > 0 \) be given. Choose \( \eta > 0 \) so that \( \alpha(b) - \alpha(a) < \varepsilon / \eta \). Since \( f \) is continuous in the interval \([a, b]\) so it is uniformly continuous on \([a, b]\), there exists a \( \delta > 0 \) such that
\[
| f(x) - f(t) | < \eta \quad ; \quad |x - t| < \delta \forall x, t \in [a, b].
\]
(\(*\))

If \( P = \{x_0, x_1, ..., x_{n-1}, x_n\} \) partition of \([a, b]\) such that \( \|P\| < \delta \) and \( M_i = \sup f(x) \) \( m_i = \inf f(x) \) \( (x_{i-1} \leq x \leq x_i) \) then \( \Delta x_i < \delta \) for all \( i \), therefore by (\(*\)), we have \( M_i - m_i \leq \eta \) \( (i = 1, ..., n) \).
Yields that
\[
U(P, f, \alpha) - L(P, f, \alpha) = \sum_{i=1}^{n} (M_i - m_i) \Delta \alpha_i \leq \eta \sum_{i=1}^{n} \Delta \alpha_i = \eta [\alpha(b) - \alpha(a)] < \varepsilon.
\]

Hence \( f \in R(\alpha) \).

1.1.10. Theorem. If \( f \) is monotonic on \([a, b]\) and \( \alpha \) is continuous on \([a, b]\), then \( f \in R(\alpha) \).
(We still assume, of course, that \( \alpha \) is monotonic.)

**Proof.** Let \( \varepsilon > 0 \) be given. Since \( \alpha \) is continuous on \([a, b]\) and so it takes all values between \( \alpha(a) \) and \( \alpha(b) \) also it is monotonic increasing. Therefore for a positive integer \( n \) there exits a partition \( P = \{x_0, x_1, ..., x_{n-1}, x_n\} \) of \([a, b]\) such that \( \Delta \alpha_i = [\alpha(b) - \alpha(a)] / n \) \( (i = 1, ..., n) \).

We suppose that \( f \) is monotonically increasing then \( M_i = \sup f(x) = f(x_i) \) \( m_i = \inf f(x) = f(x_{i-1}) \) \( (x_{i-1} \leq x \leq x_i) \) so that
\[
U(P, f, \alpha) - L(P, f, \alpha) = \sum_{i=1}^{n} (M_i - m_i) \Delta \alpha_i = [\alpha(b) - \alpha(a)] / n \sum_{i=1}^{n} \{f(x_i) - f(x_{i-1})\}
\]
\[
= [\alpha(b) - \alpha(a)] \{f(b) - f(a)\} / n < \varepsilon \quad \text{as} \quad n \quad \text{is taken large enough.}
\]

Hence \( f \in R(\alpha) \).

1.1.11. Theorem. Suppose \( f \) is bounded on \([a, b]\), \( f \) has only finitely many points of discontinuity on \([a, b]\), and \( \alpha \) is continuous at every point at which \( f \) is discontinuous. Then \( f \in R(\alpha) \).

**Proof.** Let \( \varepsilon > 0 \) be given. Let \( M = \sup |f(x)| \), \( a \leq x \leq b \) and let \( E \) be the set of points at which \( f \) is discontinuous. So \( E \) is finite and \( \alpha \) is continuous at every element of \( E \), we can cover \( E \) by finitely many disjoint intervals \([u_1, v_1] \subset [a, b]\) such that the sum of the corresponding differences \( \alpha(v_j) - \alpha(u_j) \) is less than \( \varepsilon \), i.e.,
\[
\sum_{i=1}^{m} [\alpha(v_j) - \alpha(u_j)] < \varepsilon.
\]
Furthermore without loss generality we may place these intervals in such a way that every point of \( E \cap (a, b) \) lies in the interior of some \([u_j, v_j]\), i.e., \( x \in E \cap [u_j, v_j] \Rightarrow x \) is interior point of \([u_j, v_j]\).

Let \( K = [a, b] - \bigcup_{i=1}^{m} (u_i, v_i) \) then the set \( K \) is compact. Hence \( f \) is uniformly continuous on \( K \) so there exists \( \delta > 0 \) such that

\[ |f(s) - f(t)| < \epsilon \quad \text{if} \quad |s - t| < \delta, \quad \forall \ s, \ t \in K. \]

Consider a partition \( P = \{x_0, x_i, \ldots, x_n\} \) of \([a, b]\) such that

(i) \( \text{each } u_j, v_j \in P \)

(ii) \( \text{all } x \in (u_j, v_j) \Rightarrow x \not\in P \)

(iii) \( \text{if } x_{i-1} \neq u_j \text{ then } \Delta x_i < \delta \text{ for all } i. \)

Then \( M_i - m_i = |M_i - m_i| \leq |M_i| + |m_i| \leq 2|M_i| = 2 \sup |f(x)| \leq 2 \sup |f(x)| |M_i| \leq 2M \ \forall i. \)

Now by (i), we have \( M_i - m_i \leq \epsilon \) if \( x_i \in [a, b] \) and \( x_{i-1} \neq u_j. \) Hence

\[
U(P, f, \alpha) - L(P, f, \alpha) = \sum_{i=1}^{n} (M_i - m_i) \Delta \alpha_i = \sum_{i=1}^{n} (M_i - m_i) \Delta \alpha_i (x_i = u_j) + \sum_{i=1}^{n} (M_i - m_i) \Delta \alpha_i (x_i \neq u_j)
\]

\[ \leq 2M \sum_{i=1}^{n} \Delta \alpha_i + \epsilon \sum_{i=1}^{n} \Delta \alpha_i \leq 2M \epsilon + \epsilon = \epsilon. \] Hence \( f \in R(\alpha). \)

**Question:** Is if \( f \) and \( \alpha \) have a common point of discontinuity, then \( f \in R(\alpha)? \)

**1.1.12. Theorem.** Suppose \( f \in R(\alpha) \) on \([a, b]\), \( m \leq f \leq M, \phi \) is continuous on \([m, M]\), and \( h(x) = \phi(f(x)) \) on \([a, b]\). Then \( h \in R(\alpha) \) on \([a, b]\).

**Proof.** Choose \( \epsilon > 0. \) Since \( \phi \) is uniformly continuous on \([m, M]\), there exists \( \delta > 0 \) such that \( \delta < \epsilon \) and

\[ |\phi(s) - \phi(t)| < \epsilon \quad \text{if} \quad |s - t| < \delta, \quad \forall s, t \in [m, M]. \]

Further since \( f \in R(\alpha), \) there is a partition \( P = \{x_0, x_1, x_2, \ldots, x_n\} \) of \([a, b]\) such that

\[
U(P, f, \alpha) - L(P, f, \alpha) < \delta^2 \quad \text{...} \quad \text{...} \quad (i)
\]

Let \( M_i, m_i \) and \( M_i^*, m_i^* \) are the supremum and infimum of \( f \) and \( h \) respectively in \([x_{i-1}, x_i]\).

Consider two classes : \( i \in A \) if \( M_i - m_i < \delta, \quad M_i^* - m_i^* \leq \epsilon \) and \( i \in B \) if \( M_i - m_i \geq \delta, \quad M_i^* - m_i^* \leq 2K, \) where \( K = \sup |\phi(t)|, m \leq t \leq M. \)

By (i), we have

\[ \delta \sum_{i \in B} \Delta \alpha_i \leq \sum_{i \in B} (M_i - m_i) \Delta \alpha_i < \delta^2 \] implies that \( \sum_{i \in B} \Delta \alpha_i < \delta \). It follows that

\[
U(P, h, \alpha) - L(P, h, \alpha) = \sum_{i \in B} (M_i^* - m_i^*) \Delta \alpha_i + \sum_{i \in B} (M_i - m_i) \Delta \alpha_i
\]

\[ \leq \epsilon [\alpha(b) - \alpha(a)] + 2K \delta < \epsilon [\alpha(b) - \alpha(a) + 2K] = \epsilon'. \] Hence \( h \in R(\alpha). \)

**1.2. Properties of Reimann-Stietjes Integrals**
1.2.1. Theorem (a) If \( f_1, f_2 \in R(\alpha) \) on \([a, b]\), then \( f_1 + f_2 \in R(\alpha) \), \( cf \in R(\alpha) \) for every constant \( c \), and
\[
\int_a^b (f_1 + f_2) \, d\alpha = \int_a^b f_1 \, d\alpha + \int_a^b f_2 \, d\alpha ; \quad \int_a^b cf \, d\alpha = c \int_a^b f \, d\alpha.
\]
(b) If \( f_1(x) \leq f_2(x) \) on \([a,b]\), then
\[
\int_a^b f_1 \, d\alpha \leq \int_a^b f_2 \, d\alpha.
\]
(c) If \( f \in R(\alpha) \) on \([a, b]\) and if \( a < c < b \), then \( f \in R(\alpha) \) on \([a, c]\) and on \([c, b]\), and
\[
\int_a^c f \, d\alpha + \int_c^b f \, d\alpha = \int_a^b f \, d\alpha.
\]
(d) If \( f \in R(\alpha) \) on \([a, b]\) and if \( |f(x)| \leq M \) on \([a, b]\), then
\[
\left| \int_a^b f \, d\alpha \right| \leq M[\alpha(b) - \alpha(a)].
\]
(e) If \( f \in R(\alpha_1) \) and \( f \in R(\alpha_2) \), then \( f \in R(\alpha_1 + \alpha_2) \) and
\[
\int_a^b f \, d(\alpha_1 + \alpha_2) = \int_a^b f \, d\alpha_1 + \int_a^b f \, d\alpha_2.
\]
If \( f \in R(\alpha) \) and \( c \) is a positive constant, then \( f \in R(c\alpha) \) and
\[
\int_a^b f \, d(c\alpha) = c \int_a^b f \, d\alpha.
\]

**Proof.** (a) Since \( f_1, f_2 \in R(\alpha) \) then for given \( \varepsilon > 0 \) there are partitions \( P_j(j = 1, 2) \) such that
\[
U(P_j, f_j, \alpha) - L(P_j, f_j, \alpha) < \varepsilon/2, \quad j = 1, 2. \quad \ldots \quad \ldots \quad (i)
\]
If \( P = P_1 \cup P_2 \) then (i) remains true for \( P \),
\[
U(P, f_j, \alpha) - L(P, f_j, \alpha) < \varepsilon/2, \quad j = 1, 2. \quad \ldots \quad \ldots \quad (ii)
\]
Now if \( f = f_1 + f_2 \), \( M_i, M'_i, M''_i, m_i, m'_i, m''_i \) are the supremum and infimum in \( x \in [x_{i-1}, x_i] \) of \( f, f_1, f_2 \) respectively then \( M_i \leq M_1 + M'_1, m_i \geq m_1 + m'_1 \) so that
\[
U(P, f, \alpha) = \sum_{i=1}^n M_i \Delta \alpha_i \leq \sum_{i=1}^n (M'_i + M''_i) \Delta \alpha_i = \sum_{i=1}^n M'_i \Delta \alpha_i + \sum_{i=1}^n M''_i \Delta \alpha_i \leq U(P, f_1, \alpha) + U(P, f_2, \alpha).
\]
Similarly, \( L(P, f, \alpha) \geq L(P, f_1, \alpha) + L(P, f_2, \alpha) \). Therefore
\[
U(P, f, \alpha) - L(P, f, \alpha) \leq [U(P, f_1, \alpha) + U(P, f_2, \alpha) - \{L(P, f_1, \alpha) + L(P, f_2, \alpha)\}] \leq [U(P, f_1, \alpha) - L(P, f_1, \alpha)] + [U(P, f_2, \alpha) - L(P, f_2, \alpha)] < \varepsilon, \quad \text{by (ii)}
\]
which proves that \( f = f_1 + f_2 \in R(\alpha) \) so that
\[
\int_a^b f \, d\alpha = \int_a^b f \, d\alpha = \int_a^b f \, d\alpha.
\]
Further for any \( \varepsilon > 0 \) and a partition \( P \) of \([a, b]\)
\[
U(P, f_1, \alpha) < \int_a^b f \, d\alpha + \varepsilon/2 \quad \text{and} \quad U(P, f_2, \alpha) < \int_a^b f \, d\alpha + \varepsilon/2
\]
then on adding these, we have
\[
U(P, f_1, \alpha) + U(P, f_2, \alpha) < \int_a^b f \, d\alpha + \varepsilon.
\]
Let \( x \in \mathbb{R} \) be any partition of \([a, b]\) and \( f(x) \). Then for given \( \varepsilon > 0 \) there is a partition \( P \) of \([a, b]\) such that
\[
U(P, f, \alpha) - L(P, f, \alpha) = \varepsilon.
\]
Moreover, \( f(x) \) is a monotonically increasing function on \([a, b]\). Then for each \( \alpha \), \( \alpha(x_i) - \alpha(x_{i-1}) \geq 0 \) also if \( m_i = \inf f(x) \leq x_i \leq x_{i-1} \leq x \leq b \), then \( L(P, f, \alpha) \geq 0 \). Consequently \( |\Delta \alpha| = \sum \Delta \alpha_i \leq \sum \alpha(x_i) - \alpha(x_{i-1}) \geq 0 \) also if \( m_i = \inf f(x) \leq x_i \leq x_{i-1} \leq x \leq b \), then \( L(P, f, \alpha) \geq 0 \). Therefore if \( f_1(x) \leq f_2(x) \) on \([a, b]\) then \( f_2(x) - f_1(x) \geq 0 \) implies that
\[
0 \leq \int_{a}^{b} (f_2 - f_1) dx = \int_{a}^{b} f_2 dx - \int_{a}^{b} f_1 dx.
\]
(c) Since \( f \in \mathbb{R} \) then for given \( \varepsilon > 0 \) there exist a partition \( P = \{ x_0, x_1, \ldots, x_n \} \) of \([a, b]\) such that
\[
U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon.
\]
where \( U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta \alpha_i \) and \( L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta \alpha_i \). Let \( x_i = c \), then
\[
U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta \alpha_i + \sum_{i=1}^{n} \Delta \alpha_i \quad \text{and} \quad L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta \alpha_i + \sum_{i=1}^{n} \Delta \alpha_i.
\]
Therefore \( U(P, f, \alpha) - L(P, f, \alpha) = \sum_{i=1}^{n} (M_i - m_i) \Delta \alpha_i \leq \varepsilon \), by (vi)

which implies that \( \sum_{i=1}^{r} (M_i - m_i) \Delta \alpha_i < \varepsilon \) and \( \sum_{i=r+1}^{n} (M_i - m_i) \Delta \alpha_i < \varepsilon \).

Hence \( f \in R(\alpha) \) on \([a, c]\) and on \([c, b]\).

Further since \( f \in R(\alpha) \) on \([a, c]\) and on \([c, b]\) therefore for \( \varepsilon > 0 \), we have

\[
\left| \sum_{i=a}^{c} f(t_i) \Delta \alpha_i - \int_a^c f(t) \, dt \right| < \varepsilon/3, \quad \left| \sum_{i=c}^{b} f(t_i) \Delta \alpha_i - \int_c^b f(t) \, dt \right| < \varepsilon/3 \text{ and } \left| \sum_{i=b}^{a} f(t_i) \Delta \alpha_i - \int_b^a f(t) \, dt \right| < \varepsilon/3.
\]

Therefore,

\[
\left| \int_a^b f(t) \, dt - \int_a^c f(t) \, dt - \int_c^b f(t) \, dt \right| = \left| \sum_{i=a}^{c} f(t_i) \Delta \alpha_i - \int_a^c f(t) \, dt \right| + \left| \sum_{i=c}^{b} f(t_i) \Delta \alpha_i - \int_c^b f(t) \, dt \right| + \left| \sum_{i=b}^{a} f(t_i) \Delta \alpha_i - \int_b^a f(t) \, dt \right|
\]

\[
\leq \left| \sum_{i=a}^{c} f(t_i) \Delta \alpha_i \right| + \left| \sum_{i=c}^{b} f(t_i) \Delta \alpha_i \right| + \left| \sum_{i=b}^{a} f(t_i) \Delta \alpha_i \right| < \varepsilon.
\]

Letting \( \varepsilon \to 0 \), we have \( \int_a^b f(t) \, dt = \int_a^c f(t) \, dt + \int_c^b f(t) \, dt \).

(d) Since \( \int_a^b f(t) \, dt \leq |\int_a^b f(t) \, dt| \) (we will prove this result in next theorem). If \( |f(x)| \leq M \) on \([a, b]\), then

\[
\left| \int_a^b f(t) \, dt \right| \leq \int_a^b |f(t)| \, dt \leq \int_a^b M \, dt = M(b-a).
\]

(e) We have given that \( f \in R(\alpha_1) \) and \( f \in R(\alpha_2) \) then for given \( \varepsilon > 0 \) there exists partitions \( P_1 \) and \( P_2 \) of \([a, b]\) such that

\[
U(P_i, f, \alpha_i) - L(P_i, f, \alpha_i) < \varepsilon/2, \quad j = 1, 2 \quad \ldots \quad \ldots \quad \text{(i)}
\]

If \( P = P_1 \cup P_2 \) then \( U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon/2, \quad j = 1, 2 \quad \ldots \quad \ldots \quad \text{(ii)} \)

Suppose that \( \alpha = \alpha_1 + \alpha_2 \) then, we can write

\[
\sum_{i=1}^{n} M_i[\alpha(x_i) - \alpha(x_{i-1})] = \sum_{i=1}^{n} M_i[(\alpha_1 + \alpha_2)(x_i) - (\alpha_1 + \alpha_2)(x_{i-1})]
\]

\[
= \sum_{i=1}^{n} M_i[\alpha_1(x_i) - \alpha_1(x_{i-1})] + \sum_{i=1}^{n} M_i[\alpha_2(x_i) - \alpha_2(x_{i-1})]
\]

\[i.e., \quad U(P, f, \alpha) = U(P, f, \alpha_1) + U(P, f, \alpha_2), \quad \text{similarly} \quad L(P, f, \alpha) = L(P, f, \alpha_1) + L(P, f, \alpha_2).
\]

So, \( U(P, f, \alpha) - L(P, f, \alpha) = [U(P, f, \alpha_1) - L(P, f, \alpha_1)] + [U(P, f, \alpha_2) - L(P, f, \alpha_2)] < \varepsilon \) by (ii).

Hence \( f \in R(\alpha_1+\alpha_2) \). Further if \( f \in R(\alpha_1+\alpha_2) \) and \( c \) is any positive constant then for a partition \( P \) of \([a, b]\), we have \( U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon \)

\[
\Rightarrow \quad cU(P, f, \alpha) - L(P, f, \alpha) < c\varepsilon
\]

\[
\Rightarrow \quad U(P, f, c\alpha) - L(P, f, c\alpha) < c\varepsilon
\]

\[
\Rightarrow \quad f \in R(c\alpha)
\]

Further

\[
\int_a^b f(c\alpha) = \inf_{P} U(P, f, c\alpha) = \inf_{P} U(P, f, \alpha) = c \int_a^b f \, da.
\]
1.2.2. Theorem. If \( f \in \text{R(}\alpha\) \) and \( g \in \text{R(}\alpha\) \) on \([a, b]\) then

(a) \( fg \in \text{R(}\alpha\) \);
(b) \( |f| \in \text{R(}\alpha\) \) and \( \int_a^b |f| \, d\alpha \leq \int_a^b |f| \, d\alpha \).

**Proof.** (a) Since \( f \in \text{R(}\alpha\) \) and \( g \in \text{R(}\alpha\) \) on \([a, b]\) then \( f + g, f - g \in \text{R(}\alpha\) \) on \([a, b]\). Also if \( \phi(t) = t^2 \) then \( h(x) = \phi(f(x)) \) \( [f(x)]^2 \in \text{R(}\alpha\) \) on \([a, b]\). Therefore by identity \( 4fg = (f + g)^2 - (f - g)^2 \in \text{R(}\alpha\) \) or \( fg \in \text{R(}\alpha\) \).

(b) Since \( f \in \text{R(}\alpha\) \) then for given \( \varepsilon > 0 \) there is a partitions \( P \) such that

\[
\int f \, d\alpha = \inf U(P, f, \alpha) = f(s) \quad \text{and} \quad \int f \, d\alpha = \sup L(P, f, \alpha) < \varepsilon.
\]

Also since \( M_i, m_i \) are the supremum and infimum of \( f \) in \( x \in [x_{i-1}, x_i] \).

Now if \( M'_i, m'_i \) are the supremum and infimum of \( |f| \) in \( x \in [x_{i-1}, x_i] \). Then for \( x_{i-1}, x_i \in [x_{i-1}, x_i] \) we have \( ||f(x_i) - |f(x_i)|| \leq |f(x_i) - f(x_s)| \) implies that \( M'_i - m'_i \leq M_i - m_i \)

\[
\Rightarrow \quad \sum_{i=1}^n (M_i - m_i) \Delta \alpha_i < \varepsilon, \quad \text{where} \quad M_i, m_i \text{ are the supremum and infimum of } f \text{ in } x \in [x_{i-1}, x_i].
\]

Therefore by theorem 1.2.2(b) is not true always.

1.2.1. **Definition. (step function)** The unit step function \( I \) is defined by

\[
I(x) = \begin{cases} 
0, & \text{if } x \leq 0, \\
1, & \text{if } x > 0.
\end{cases}
\]

1.2.3. **Theorem.** If \( a < s < b \), \( f \) is bounded on \([a, b]\), \( f \) is continuous at \( s \) and \( \alpha(x) = I(x - s) \), then

\[
\int_a^b f \, d\alpha = f(s).
\]

**Proof.** Let \( P = \{x_0, x_1, x_2, x_3\} \), where \( x_0 = a \) and \( x_1 = s < x_2 < x_3 = b \) be any partition of \([a, b]\). Then \( M_i = \sup f(x) = f(s) \) and \( m_i = \inf f(x) = f(s), x \in [x_{i-1}, x_i], i = 1, 2, 3 \) as \( f \) is continuous at \( s \). Therefore

\[
U(P, f, \alpha) = \sum_{i=1}^3 M_i \Delta \alpha_i = \sum_{i=1}^3 M_i[\alpha(x_i) - \alpha(x_{i-1})] = M_1[\alpha(x_1) - \alpha(x_0)] + M_2[\alpha(x_2) - \alpha(x_1)]
\]

\[
+ M_3[\alpha(x_3) - \alpha(x_2)] = M_1[\int(x_1 - s) - I(x_0 - s)] + M_2[\int(x_2 - s) - I(x_1 - s)] + M_3[\int(x_3 - s) - I(x_2 - s)]
\]

\[
= M_2. \quad \text{Similarly} \quad L(P, f, \alpha) = \sum_{i=1}^3 m_i \Delta \alpha_i = m_2. \quad \text{Since} \ f \ \text{is continuous at} \ s \ \text{so that} \ M_2 \ \text{and} \ m_2 \ \text{converge to} \ f(s) \ \text{as} \ x_2 \to s. \ \text{Therefore}
\]

\[
\int_a^b f \, d\alpha = \inf U(P, f, \alpha) = f(s) \quad \text{and} \quad \int_a^b f \, d\alpha = \sup L(P, f, \alpha) = f(s). \quad \text{Hence}
\]
1.2.4. **Theorem.** Suppose \( c_n \geq 0 \) for \( n = 1, 2, 3, \ldots, \Sigma c_n \) converges, \( \{s_n\} \) is a sequence of distinct points in \((a, b)\), and

\[
\int_a^b f \, d\alpha = \int_a^b f \, d\alpha = \int_a^b f \, d\alpha = f(s).
\]

\[\alpha(x) = \sum_{n=1}^{\infty} c_n I(x - s_n).\]

Let \( f \) be continuous on \([a, b]\). Then

\[
\int_a^b f \, d\alpha = \sum_{n=1}^{\infty} c_n f(s_n).
\]

**Proof.** Since \( c_n I(x - s_n) \leq c_n \), then by the comparison test the series (*) is converges for every \( x \). Also its sum \( \alpha(x) \) is evidently monotonic and \( \alpha(a) = 0, \alpha(b) = \Sigma_{n=1}^{\infty} c_n \).

Also since \( \Sigma c_n \) converges then for given \( \varepsilon > 0 \) choose \( N \) so that \( \Sigma_{n=1}^{N} c_n < \varepsilon \).

Put \( \alpha_1(x) = \sum_{n=1}^{N} c_n I(x - s_n) \) and \( \alpha_2(x) = \sum_{n=1}^{\infty} c_n I(x - s_n) \).

By theorem 1.2.1 & 1.2.3, we have

\[
\int_a^b f \, d\alpha = \sum_{n=1}^{\infty} c_n f(s_n) \quad \text{and} \quad \left| \int_a^b f \, d\alpha \right| \leq M\varepsilon.
\]

where \( M = \sup |f(x)| \). Since \( \alpha = \alpha_1 + \alpha_2 \) then \( \alpha_2 = \alpha - \alpha_1 \), it follows that

\[
\left| \int_a^b f \, d\alpha - \Sigma_{n=1}^{N} c_n f(s_n) \right| \leq M\varepsilon.
\]

Letting \( N \to \infty \), we have

\[
\int_a^b f \, d\alpha = \sum_{n=1}^{\infty} c_n f(s_n).
\]

1.2.5. **Theorem.** *(Relation between Riemann integral and Riemann-Stieljes integral)* Assume \( \alpha \) increases monotonically and \( \alpha' \in \mathbb{R}[a, b] \). Let \( f \) be a bounded real function on \([a, b]\). Then \( f \in R(\alpha) \) if and only if \( f \alpha' \in R[a, b] \). Moreover

\[
\int_a^b f \, d\alpha = \int_a^b f(x) \alpha'(x) \, dx \quad \ldots \quad \ldots \quad (i)
\]

**Proof.** Let \( \varepsilon > 0 \) be given, since \( \alpha' \in \mathbb{R}[a, b] \) therefore there is a partition \( P = \{x_0, \ldots, x_n\} \) of \([a, b]\) such that

\[
U(P, \alpha') - L(P, \alpha') < \varepsilon \quad \ldots \quad \ldots \quad (ii)
\]

Now by mean value theorem for real valued function for \( t_i \in [x_{i-1}, x_i] \) such that

\[
\alpha(x_i) - \alpha(x_{i-1}) = \alpha'(t_i)(x_i - x_{i-1}) \quad \text{or} \quad \Delta\alpha(x_i) = \alpha'(t_i) \Delta x_i, \quad \text{for} \ i = 1, \ldots, n. \quad \ldots \quad \ldots \quad (iii)
\]

Again since equation (ii) is true for \( s_i \in [x_{i-1}, x_i] \), then

\[
\sum_{i=1}^{n} | \alpha'(s_i) - \alpha'(t_i) | \Delta x_i < \varepsilon \quad \ldots \quad \ldots \quad (iv)
\]

If \( M = \sup |f(x)| \), then by (iii), we have

\[
\sum_{i=1}^{n} f(s_i) \Delta \alpha_i = \sum_{i=1}^{n} f(s_i) \alpha'(t_i) \Delta x_i \quad \text{it follows that}
\]
\begin{align*}
| \sum_{i=1}^{n} f(s_i) \Delta a_i - \sum_{i=1}^{n} f(s_i) \alpha'(t_i) \Delta x_i | &= | \sum_{i=1}^{n} f(s_i) \alpha'(t_i) \Delta x_i - \sum_{i=1}^{n} f(s_i) \alpha'(t_i) \Delta x_i | \\
&= | \sum_{i=1}^{n} f(s_i) [\alpha'(t_i) - \alpha'(t_i)] \Delta x_i | \\
&= \sum_{i=1}^{n} |f(s_i)| |\alpha'(t_i) - \alpha'(t_i)| \Delta x_i \\
&\leq \sum_{i=1}^{n} \sup |f(x)| |\alpha'(t_i) - \alpha'(t_i)| \Delta x_i = M \sum_{i=1}^{n} |\alpha'(t_i) - \alpha'(t_i)| \Delta x_i < M \varepsilon \quad \text{[by (iv)]}
\end{align*}

In particular, \( \sum_{i=1}^{n} f(s_i) \Delta a_i \leq U(P, f, \alpha') + M \varepsilon \quad \forall \, s_i \in [x_{i-1}, x_i] \). Then
\[
U(P, f, \alpha) \leq U(P, f, \alpha') + M \varepsilon \quad \text{or} \quad U(P, f, \alpha) - U(P, f, \alpha') \leq M \varepsilon.
\]
Similarly by the same argument we have
\[
U(P, f, \alpha') \leq U(P, f, \alpha) + M \varepsilon \quad \text{or} \quad U(P, f, \alpha') - U(P, f, \alpha) \leq M \varepsilon.
\]
Thus
\[
|U(P, f, \alpha) - U(P, f, \alpha')| \leq M \varepsilon \quad \text{...} \quad \text{...} \quad (v)
\]
Here we note that (ii) and (v) remains true if \( P \) is replaced by any refinement. We conclude that
\[
| \int_{a}^{b} f d\alpha - \int_{a}^{b} f(x) \alpha'(x) dx | \leq M \varepsilon.
\]
Since \( \varepsilon \) is arbitrary. Hence
\[
\int_{a}^{b} f d\alpha = \int_{a}^{b} f(x) \alpha'(x) dx \quad \text{...} \quad \text{...} \quad (vi)
\]
Similarly,
\[
\int_{a}^{b} g d\alpha = \int_{a}^{b} f(x) \alpha'(x) dx \quad \text{...} \quad \text{...} \quad (vii)
\]
From (vi) and (vii) we have
\[
\int_{a}^{b} g d\alpha = \int_{a}^{b} f(x) \alpha'(x) dx.
\]

1.2.6. Theorem. (change of variable). Suppose \( \phi \) is a strictly increasing continuous function that maps an interval \([A, B]\) onto \([a, b]\). Suppose \( \alpha \) is monotonically increasing on \([a, b]\) and \( f \in R(\alpha) \) on \([a, b]\). Define \( \beta \) and \( g \) on \([A, B]\) by
\[
\beta(y) = \alpha(\phi(y)), \quad g(y) = f(\phi(y)) \quad \text{...} \quad \text{...} \quad (i)
\]
Then \( g \in R(\beta) \) and
\[
\int_{A}^{B} g d\beta = \int_{a}^{b} f d\alpha.
\]

Proof. Let to each partition \( P = \{x_0, ..., x_n\} \) of \([a, b]\) corresponds a partition \( Q = \{y_0, ..., y_n\} \) of \([A, B]\), then \( x_i = \phi(y_i) \). All partitions of \([A, B]\) are obtained in this way. Since the values taken by \( f \) on \([x_{i-1}, x_i]\) are exactly the same as those taken by \( g \) on \([y_{i-1}, y_i]\), we see that
\[
U(Q, g, \beta) = U(P, f, \alpha) \quad \text{and} \quad L(Q, g, \beta) = L(P, f, \alpha).
\]
Since \( f \in R(\alpha) \) then for given \( \varepsilon > 0 \), \( |U(P, f, \alpha) - L(P, f, \alpha)| < \varepsilon \).
Hence for the partition \( Q \)
\[
|U(Q, g, \beta) - L(Q, g, \beta)| < \varepsilon. \quad \text{Which shows that} \quad g \in R(\beta).
\]
Furthermore
\[
\int_{A}^{B} g d\beta = \int_{a}^{b} f d\alpha.
\]
Note: Consider the following special case: Take $\alpha(x) = x$, then $\beta = \phi$. Assume $\phi' \in R [A, B]$. Then by theorem 1.2.5, we have

$$\int_{a}^{b} f(x)dx = \int_{A}^{B} f(\phi(y)) \phi'(y) dy.$$  

1.3. Integration and Differentiation. In this section we discuss the famous the fundamental theorem of calculus, which stated that integration and differentiation are, in a certain sense, inverse operations. We shall made this study for Riemann integrals.

1.3.1. Definition. Integral Function. If $f$ is Riemann integrable function on $[a, b]$ then a function

$$F(x) = \int_{a}^{x} f(t) dt$$ is called an integral function. Further if $f(x)$ is differentiable on $[a, b]$ and $F'(x) = f(x)$, then $F(x)$ is called the primitive or anti-derivative of $f$ on $[a, b]$.

Here we note that the primitive of $f(x)$ is not unique also an integrable function is not necessarily continuous but the function associated to $f$ is always continuous as shown in the following theorem

1.3.1. Theorem. Let $f \in R[a, b]$. For $a \leq x \leq b$, put $F(x) = \int_{a}^{x} f(t) dt$. Then $F$ is continuous on $[a, b]$ and if $f(x)$ is continuous at a point $x_0$ of $[a, b]$, then $F$ is differentiable at $x_0$, and $F'(x_0) = f(x_0)$. In other word “The integral of a Riemann integrable function is continuous and is differentiable if $f$ is continuous”.

Proof. Since $f \in R[a, b]$ so by definition $f$ is bounded therefore there exists a positive constant $M$ such that $|f(t)| \leq M$ for $a \leq t \leq b$. Now for $a \leq x \leq y \leq b$, we have

$$|F(y) - F(x)| = |\int_{a}^{x} f(t) dt - \int_{a}^{y} f(t) dt| = |\int_{a}^{y} f(t) dt + \int_{y}^{x} f(t) dt| = |\int_{y}^{x} f(t) dt| \leq M(y - x).$$

Suppose that for given $\varepsilon > 0$ such that $|y - x| < \varepsilon/M$. Then we have

$$|F(y) - F(x)| < \varepsilon.$$

This proves continuity (and, in fact, uniform continuity) of $F$.

Next suppose $f$ is continuous at $x_0$. Then for given $\varepsilon > 0$, choose $\delta > 0$ such that $|f(t) - f(x_0)| < \varepsilon$; if $|t - x_0| < \delta$, where $a \leq t \leq b$. Hence, if $x_0 - \delta < s \leq x_0 \leq t < x_0 + \delta$ and $a \leq s < t \leq b$, we observe that

$$\left\{1/(t - s)\right\} \int_{s}^{t} f(t) dt = 1.$$ Therefore

$$||[F(t) - F(s)]/(t - s)) - f(x_0)| = |\left\{1/(t - s)\right\} \int_{s}^{t} f(u)du - \left\{1/(t - s)\right\} \int_{s}^{t} f(x_0)du|$$
\[
\leq \left\{ 1/|t - s| \right\} \int_s^t |f(u) - f(x_0)| du < \varepsilon
\]

It follows that \( F \) is differentiable at \( x_0 \) and \( F'(x_0) = f(x_0) \).

Since here \( x_0 \) is arbitrary so that here we observe that the continuity of \( f \) implies the differentiability of \( F \). Some time this theorem is called the \textbf{first fundamental theorem of integral calculus}.

\textbf{1.3.2. Theorem (The fundamental theorem of calculus).} If \( f \in \mathbb{R}[a, b] \) and if there is a differentiable function \( F \) on \([a, b]\) such that \( F' = f \), then

\[
\int_a^b f(x) \, dx = F(b) - F(a).
\]

\textbf{Proof.} Let \( \varepsilon > 0 \) be given, since \( F' = f \in \mathbb{R}[a, b] \) therefore there exists a \( P = \{x_0, \ldots, x_n\} \) of \([a, b]\) with \( ||P|| < \delta \) such that

\[
|\sum F'(t_i) \Delta x_i - \int F'(x) dx| < \varepsilon \quad \text{where} \quad t_i \in [x_{i-1}, x_i] \quad \cdots \quad \cdots \quad (i)
\]

Now by Lagrange’s mean value theorem for point \( t_i \in [x_{i-1}, x_i] \) such that

\[
F(x_i) - F(x_{i-1}) = F'(t_i) \Delta x_i \quad \text{for} \quad i = 1, \ldots, n.
\]

Therefore

\[
\sum F'(t_i) \Delta x_i = \sum [F(x_i) - F(x_{i-1})] = F(b) - F(a).
\]

Therefore by equation \((i)\), we have

\[
|F(b) - F(a) - \int F'(x) dx| < \varepsilon. \quad \text{Which implies that}
\]

\[
F(b) - F(a) = \int_a^b f(x) dx = \int f(x) dx \quad \text{or} \quad \int_a^b f(x) dx = F(b) - F(a).
\]

\textbf{1.3.3. Theorem (integration by parts).} Suppose \( F \) and \( G \) are differentiable functions on \([a, b]\), \( F' = f \in \mathbb{R}[a, b] \) and \( G' = g \in \mathbb{R}[a, b] \). Then

\[
\int_a^b F(x)g(x) dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x) dx.
\]

\textbf{Proof.} Put \( H(x) = F(x)G(x) \), since \( F \) and \( G \) are differentiable so \( H(x) \) is also differentiable and \( H'(x) = F(x)G'(x) + G(x)F'(x) \in \mathbb{R}[a, b] \). Then using fundamental theorem of integral calculus, we have

\[
\int_a^b H'(x) \, dx = H(b) - H(a) = F(b)G(b) - F(a)G(a).
\]

\[
\Rightarrow \int_a^b F(x)G'(x) + \int_a^b G(x)F'(x) \, dx = H(b) - H(a) = F(b)G(b) - F(a)G(a).
\]

\[
\Rightarrow \int_a^b F(x)g(x) = H(b) - H(a) = F(b)G(b) - F(a)G(a) - \int_a^b G(x)f(x) \, dx.
\]

\textbf{1.4. Integration of vector-valued functions.}
1.4.1. Definition  Let \( f_1, \ldots, f_k \) be real functions on \([a, b]\), and let \( f = (f_1, \ldots, f_k) \) be the corresponding mapping of \([a, b]\) into \( \mathbb{R}^k \). If \( \alpha \) increases monotonically on \([a, b]\), we say that \( f \in \mathbb{R}(\alpha) \) if \( f_j \in \mathbb{R}(\alpha) \) for \( j = 1, \ldots, k \). In this case, we have

\[
\int_a^b f \, d\alpha = (\int_a^b f_1 \, d\alpha, \ldots, \int_a^b f_k \, d\alpha).
\]

i.e., \( \int f \, d\alpha \) is the point in \( \mathbb{R}^k \) whose \( j \)-th co-ordinate is \( \int f_j \, d\alpha \).

1.4.1. Theorem (a) If \( f, F \in \mathbb{R}(\alpha) \) on \([a, b]\), then \( f + F \in \mathbb{R}(\alpha) \), \( cf \in \mathbb{R}(\alpha) \) for every constant \( c \), and

\[
\int_a^b (f + F) \, d\alpha = \int_a^b f \, d\alpha + \int_a^b F \, d\alpha ; \quad \int_a^b cf \, d\alpha = c \int_a^b f \, d\alpha.
\]

(b) If \( f \in \mathbb{R}(\alpha) \) on \([a, b]\) and if \( a < c < b \), then \( f \in \mathbb{R}(\alpha) \) on \([a, c]\) and on \([c, b]\), and 

\[
\int_a^c f \, d\alpha + \int_c^b f \, d\alpha = \int_a^b f \, d\alpha.
\]

(c) If \( f \in \mathbb{R}(\alpha_1) \) and \( f \in \mathbb{R}(\alpha_2) \), then \( f \in \mathbb{R}(\alpha_1 + \alpha_2) \) and 

\[
\int_a^b f \, d(\alpha_1 + \alpha_2) = \int_a^b f \, d\alpha_1 + \int_a^b f \, d\alpha_2 ;
\]

If \( f \in \mathbb{R}(\alpha) \) and \( c \) is positive constant, then \( f \in \mathbb{R}(c\alpha) \) and 

\[
\int_a^b f \, d(c\alpha) = c \int_a^b f \, d\alpha.
\]

Proof. Since by the definition of vector–valued function these properties holds good. We simply apply the results (a), (c), and (e) of Theorem 1.2.1 to each co-ordinate of \( f \).

1.4.2. Theorem. (a) Let \( \alpha \) increases monotonically and \( \alpha' \in \mathbb{R}[a, b] \). Let \( f \) be a bounded real function on \([a, b]\). Then \( f \in \mathbb{R}(\alpha) \) if and only if \( f \alpha' \in \mathbb{R}[a, b] \). Moreover 

\[
\int_a^b f \, d\alpha = \int_a^b f(x) \, \alpha'(x) \, dx
\]

(b) Let \( f \in \mathbb{R}[a, b] \). For \( a \leq x \leq b \), put \( F(x) = \int_a^x f(t) \, dt \). Then \( F \) is continuous on \([a, b]\).

Furthermore, if \( f \) is continuous at a point \( x_0 \) of \([a, b]\), then \( F \) is differentiable at \( x_0 \), and 

\[
F'(x_0) = f(x_0).
\]

Proof. If we apply the theorems 1.2.5 and 1.3.1 on each co-ordinate of \( f \), then the results are valid. Therefore by the definition of vector–valued function these results holds good.
1.4.3. If \( f \) and \( F \) are as in theorem 1.4.2(b) \( f \in R(\alpha) \) and if \( F \) is differentiate function \( [a, b] \) such that \( F' = f \), then

\[
\int_a^b f(t) \, dt = F(b) - F(a).
\]

**Proof.** Let \( \varepsilon > 0 \) be given, since \( F' = f \in R[a, b] \) therefore there exists a partition \( P = \{x_0, \ldots, x_n\} \) of \( [a, b] \) with \( \|P\| < \delta \) such that

\[
|\sum F'_j(t_i) \Delta x_i - \int_a^b F'(t) \, dt| < \varepsilon \quad \text{where} \quad t_i \in [x_{i-1}, x_i] \quad \text{for each} \quad j.
\]

yields that

\[
|\sum F'(t_i) \Delta x_i - \int_a^b F'(t) \, dt| < \varepsilon \quad \text{where} \quad t_i \in [x_{i-1}, x_i] \quad \ldots \quad \ldots \quad (i)
\]

Now by Lagrange’s mean value theorem for point \( t_i \in [x_{i-1}, x_i] \) such that

\[
F_j(x_i) - F_j(x_{i-1}) = F'_j(t_i)\Delta x_i \quad \text{for} \quad i = 1, \ldots, n \quad \text{and each} \quad j
\]

yields that

\[
F(x_i) - F(x_{i-1}) = F'(t_i)\Delta x_i \quad \text{for} \quad i = 1, \ldots, n.
\]

Therefore

\[
\sum_{i=1}^{n} F'(t_i)\Delta x_i = \sum F(x_i) - F(x_{i-1}) = F(b) - F(a).
\]

Therefore by equation (i), we have

\[
|F(b) - F(a) - \int_a^b F'(t) \, dt| < \varepsilon. \quad \text{Which implies that}
\]

\[
F(b) - F(a) = \int_a^b F'(t) \, dt = \int_a^b f(t) \, dt \quad \text{or} \quad \int_a^b \, f(t) \, dt = F(b) - F(a).
\]

1.4.4. **Theorem** If \( f \) maps \([a, b]\) into \( R^k \) and if \( f \in R(\alpha) \) for some monotonically increasing function \( \alpha \) on \([a, b]\), then \( \|f\| \in R(\alpha) \), and

\[
\int_a^b \|f\| \, d\alpha \leq \int_a^b \|f\| \, d\alpha.
\]

**Proof.** If \( f_1, \ldots, f_k \) are the components of \( f \), then \( \|f\| = (f_1^2 + f_2^2 + \ldots + f_k^2)^{1/2} \). Since \( f \in R(\alpha) \), implies that \( f_j \in R(\alpha) \) for each \( j \) so their sum is also belongs to \( R(\alpha) \). Since \( x^2 \) is a continuous function of \( x \), then the square-root function is continuous also continuous in it domain. Hence by theorem 1.1.12 \( \|f\| \in R(\alpha) \).

Let we put \( y = (y_1, \ldots, y_k) \), where \( y_j = \int_a^b f_j \, d\alpha \), then we have \( y = \int_a^b f \, d\alpha \) and

\[
|y|^2 = \sum y_j^2 = \sum y_j \|f_j\| \, d\alpha = \int (\sum y_j f_j) \, d\alpha.
\]

By the Schwarz inequality

\[
\sum y_j f_j(t) \leq \|y\| \|f(t)\| \quad (a \leq t \leq b).
\]

Hence, we have

\[
|y|^2 \leq \|y\| \|f\| \, d\alpha \quad \text{or} \quad |y| \leq \int_a^b \|f\| \, d\alpha \quad \text{or} \quad \int_a^b \|f\| \, d\alpha \leq \int_a^b \|f\| \, d\alpha.
\]

1.5. **Rectifiable Curves.**
1.5.1. Definition. A continuous mapping \( \gamma \) of an interval \([a, b]\) into \( \mathbb{R}^k \) is said to be a curve in \( \mathbb{R}^k \). We may also say that \( \gamma \) is a curve on \([a, b]\). If

1. \( \gamma \) is one-to-one then it called an arc.
2. If \( \gamma(a) = \gamma(b) \) then it is said to be a closed curve.

3. Here we note that with each curve \( \gamma \) in \( \mathbb{R}^k \) there is associated a subset of \( \mathbb{R}^k \), called the range of \( \gamma \), but different curves may have the same range.

Now we associate to each partition \( P = \{x_0, \ldots, x_n\} \) of \([a, b]\) and to each curve \( \gamma \) on \([a, b]\) the points \( \gamma(x_0), \gamma(x_1), \gamma(x_2), \ldots, \gamma(x_n) \) are the vertices of an inscribed polygon. We denote the length of the polygon by \( \Lambda(P, \gamma) \) and define by

\[
\Lambda(P, \gamma) = \sum_{i=1}^{n} |\gamma(x_i) - \gamma(x_{i-1})|,
\]

where the \( i \)th term in this sum is the distance (in \( \mathbb{R}^k \)) between the points \( \gamma(x_{i-1}) \) and \( \gamma(x_i) \). As our partition becomes finer and finer, this polygon approaches the range of \( \gamma \) more and more closely and the polygonal length tends to the length of \( \gamma \) which is denoted by \( \Lambda(\gamma) \) and defined by \( \Lambda(\gamma) = \sup \Lambda(P, \gamma) \), where the supremum is taken over all partitions of \([a, b]\). If \( \Lambda(\gamma) < \infty \) then \( \gamma \) is said to be rectifiable.

Here \( \Lambda(\gamma) \) is given by a Riemann integral in certain cases. Our next theorem devote to this for continuously differentiable curves.

1.5.1. Theorem. If \( \gamma' \) is continuous on \([a, b]\), then \( \gamma \) is rectifiable, and \( \Lambda(\gamma) = \int_{a}^{b} |\gamma'(t)| \, dt \).

Proof. If \( a \leq x_{i-1} < x_i \leq b \), then

\[
|\gamma(x_i) - \gamma(x_{i-1})| = \int_{x_{i-1}}^{x_i} |\gamma'(t)| \, dt \leq \int_{x_{i-1}}^{x_i} |\gamma'(t)| \, dt.
\]

Hence

\[
\Lambda(P, \gamma) \leq \int_{a}^{b} |\gamma'(t)| \, dt
\]

for every partition \( P \) of \([a, b]\).

Consequently

\[
\Lambda(P, \gamma) \leq \int_{a}^{b} |\gamma'(t)| \, dt.
\]

(i)

Now for arbitrary \( \varepsilon > 0 \), since \( \gamma' \) is uniformly continuous on \([a, b]\), there exists \( \delta > 0 \) such that

\[
|\gamma'(s) - \gamma'(t)| < \varepsilon \quad \text{for} \quad |s - t| < \delta, \forall s, t \in [a, b].
\]

(ii)

Let \( P = \{x_0, \ldots, x_n\} \) be any partition of \([a, b]\), with \( \Delta x_i < \delta \), for all \( i \), then for \( x_{i-1} \leq t \leq x_i \), we have

\[
|\gamma'(t)| \leq |\gamma'(x_i)| + \varepsilon.
\]

Hence

\[
\int_{x_{i-1}}^{x_i} |\gamma'(t)| \, dt \leq |\gamma'(x_i)| \Delta x_i + \varepsilon \Delta x_i.
\]

\[
= \int_{x_{i-1}}^{x_i} [\gamma'(t) + \gamma'(x_i) - \gamma'(t)] \, dt + \varepsilon \Delta x_i.
\]

\[
\leq |\int_{x_{i-1}}^{x_i} \gamma'(t) \, dt| + |\int_{x_{i-1}}^{x_i} [\gamma'(x_i) - \gamma'(t)] \, dt| + \varepsilon \Delta x_i.
\]
\[ \gamma(x_i) - \gamma(x_{i-1}) \]

\[ \leq \left| \gamma(x_i) - \gamma(x_{i-1}) \right| + 2\varepsilon \Delta x_i. \]

On adding these inequalities, we obtain

\[ \int_a^b \left| \gamma'(t) \right| \, dt \leq \Lambda(P, \gamma) + 2\varepsilon (b - a) \leq \Lambda(\gamma) + 2\varepsilon (b - a). \]

Since \( \varepsilon \) was arbitrary therefore,

\[ \int_a^b \left| \gamma'(t) \right| \, dt \leq \Lambda(\gamma). \quad \ldots \quad \ldots \quad (iii) \]

From (i) and (iii), we have

\[ \int_a^b \left| \gamma'(t) \right| \, dt = \Lambda(\gamma). \]

**1.6. Rearrangements of terms of a Series, Riemann theorem.**

As we have studied the concept of convergence and absolute convergence of series and other related properties in earlier classes. In this section we study the nature of a convergent series, when its terms are rearranged in different manner. For this first we know what is the rearrangement of terms of a given series.

**1.6.1. Definition.** Let \( \{k_n\} \), \( n = 1, 2, 3, \ldots \) be a sequence in which every positive integer is non-repeating (In other words \( \{k_n\} \) is an injective map on the set of positive integer into itself). We put \( a'_n = a_{k_n} \), \( n = 1, 2, 3, \ldots \), then \( \sum a'_n \) is called the rearrangement of the a given \( \sum a_n \).

If we suppose that \( \{s_n\} \) and \( \{s'_n\} \) are the partial sums of \( \sum a_n \), \( \sum a'_n \) respectively, then it can be easily seen that the terms of thee sequences are entirely different in numbers and \( \{s_n\} \) is convergent whenever the series \( \sum a_n \) is convergent. Then there arise natural question about the nature of the sequence \( \{s'_n\} \) (as well as the series \( \sum a'_n \) also), i.e., it is convergent or divergent. Furthermore, thus led to the problem of determining under what condition all rearrangements of a convergent series will convergent and where the sums are necessarily the same. We can understand this by the following example:

**1.6.2 Example.** Consider the convergent series \( 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots \) and its rearrangement

\( 1 + \frac{1}{3} - \frac{1}{2} + \frac{1}{5} + \frac{1}{7} - \frac{1}{4} + \frac{1}{9} + \frac{1}{11} - \frac{1}{6} \ldots \) and if \( \{s_n\} \) and \( \{s'_n\} \) are their partial sums respectively, then \( s = \lim s_n < 1 - \frac{1}{2} + \frac{1}{3} = 5/6. \)

Now since fore \( k \geq 1, \) we se that \( \{1/(4k - 3)\} +\{1/(4k - 1)\} - (1/2k) > 0 \), we se that \( s'_3 < s'_6 < s'_9 < \ldots \), implies that \( \lim s'_n > s'_3 = 5/6. \) Hence \( s'_n \) does not converge to \( s. \)

This example illustrate the next theorem due to Riemann.

**1.6.3. Theorem(Riemann theorem)** Let \( \sum a_n \) be a series of real numbers which is converges, but not absolutely. Suppose there are numbers \( \alpha, \beta \) such that \( -\infty \leq \alpha \leq \beta \leq \infty. \) Then there exists a rearrangement \( \sum a'_n \) of \( \sum a_n \) with partial sum \( s'_n \) such that
\[ \liminf_{n \to \infty} s'_n = \alpha \quad \text{and} \quad \limsup_{n \to \infty} s'_n = \beta. \quad \ldots \quad \ldots \quad (i) \]

In other words “the terms of a conditionally convergent series can be rearranged such that the new series may be convergent, divergent or oscillatory”.

**Proof.** Consider \( p_n = (|a_n| + a_n)/2 \) and \( q_n = (|a_n| - a_n)/2 \) \((n = 1, 2, 3, \ldots)\). Then \( p_n - q_n = a_n \) and \( p_n + q_n = |a_n| \) and since \( p_n \geq 0, q_n \geq 0 \) also if the series \( \Sigma p_n \) and \( \Sigma q_n \) are convergent then the series \( \Sigma(p_n + q_n) = \Sigma |a_n| \) is convergent. Hence the series \( \Sigma p_n \) and \( \Sigma q_n \) must be divergent.

Again since \( \Sigma a_n = \Sigma(p_n - q_n) = \Sigma p_n - \Sigma q_n \) then the divergence of \( \Sigma p_n \) and convergence of \( \Sigma q_n \) (or vice versa) implies divergence of \( \Sigma a_n \) again contradicts the fact that the series \( \Sigma a_n \) is convergent.

Now we construct sequences \( \{m_n\} \) and \( \{k_n\} \) such that the series
\[
P_1 + \ldots + P_m - Q_1 - \ldots - Q_k + P_{m+1} + \ldots + P_{m+q} - Q_{k+1} - \ldots - Q_k + \ldots, \quad \ldots \quad (ii)
\]
is a rearrangement of the series \( \Sigma a_n \) and satisfies \( \liminf_{n \to \infty} s'_n = \alpha \quad \text{and} \quad \limsup_{n \to \infty} s'_n = \beta. \)

Further choose sequences \( \{\alpha_n\} \) and \( \{\beta_n\} \) such that \( \alpha_n \to \alpha \) and \( \beta_n \to \beta \), \( \alpha_n < \beta_n \), \( \beta_1 > 0 \). In such a way that, let \( m_1 \) and \( k_1 \) be the smallest integers for which
\[
P_1 + \ldots + P_{m_1} - Q_1 - \ldots - Q_k + P_{m_1+1} + \ldots + P_{m_2} - Q_{k+1} - \ldots - Q_k + \ldots < \alpha_1;
\]
and continue in this way, since it is possible because of \( \Sigma P_n, \Sigma Q_n \) are diverge.

If \( s_n \) and \( t_n \) are the partial sums of (ii) whose last terms are \( P_{m_n} \) and \( -Q_{k_n} \), then
\[
|s_n - \beta_n| \leq P_{m_n} \quad \text{and} \quad |t_n - \alpha_n| \leq Q_{k_n}.
\]

Since \( \Sigma a_n \) is convergent then \( P_n, Q_n \to 0 \) as \( n \to \infty \), we have \( s_n \to \beta \) and \( t_n \to \alpha \).

Finally we see that no number less than \( \alpha \) or greater than \( \beta \) can be a sub-sequential limit of the partial sums of (ii). Hence
\[
\liminf_{n \to \infty} s'_n = \alpha \quad \text{and} \quad \limsup_{n \to \infty} s'_n = \beta.
\]

**1.6.4. Theorem.** Let \( \Sigma a_n \) be a series of complex numbers which is absolutely convergent then every rearrangements of \( \Sigma a_n \) are converges and all converges to same sum..

Or, “The sum of every rearrangements of a absolutely convergent series remains unchanged.”

**Proof.** Let \( \Sigma a'_n \) be any rearrangement of \( \Sigma a_n \) and \( s_n, s'_n \) be the partial sums of \( \Sigma a_n, \Sigma a'_n \) respectively. Since \( \Sigma a_n \) convergence absolutely therefore by Cauchy’s criterion of convergent of series for given \( \varepsilon > 0 \) there exists a positive integer \( N \) such that
Now we choose \( p \) such that the integers 1, 2, 3, …, \( N \) lies in the set \( \{ k_1, k_2, \ldots, k_p \} \). Then for \( n > p \) the numbers \( a_1, a_2, \ldots,a_N \) will cancel in the difference \( s_n - s'_n \), therefore \( |s_n - s'_n| < \varepsilon \) by (i). Consequently \( \{s'_n\} \) converges to the same sum as \( \{s_n\} \).

### 1.7. Unity Summary

#### 1.7.1 Definitions:

Let \([a, b]\) be a given interval. Then a set \( P = \{x_0, \ x_1, \ldots, x_{n-1}, \ x_n\} \) of \([a, b]\) such that \( a = x_0 \leq x_1 \leq \ldots \leq x_{n-1} \leq x_n = b \) is said to be a **Partition** of \([a, b]\). The set of all partitions of \([a, b]\) is denoted by \( \mathcal{P}(a, b) \). The intervals \([x_0, x_1]\), \([x_1, x_2]\), \ldots, \([x_{n-1}, x_n]\) are called the subintervals of \([a, b]\). Write \( \Delta x_i = x_i - x_{i-1} \) is called the length of the interval \([x_{i-1}, x_i]\) \((i = 1, \ldots, n)\) and \( \max \{\Delta x_i\} \) is called the norm of the partition \( P \) and is denoted by \( ||P|| \) or \( \mu(P) \). A partition \( Q \) of \([a, b]\) such that \( P \subset Q \) is called the refinement or finer of the partition \( P \).

Let \( f: [a, b] \rightarrow \mathbb{R} \) is bounded function and \( \alpha \) be a monotonically increasing function on \([a, b]\) and \( P \) be any **Partition** of \([a, b]\). We write \( \Delta \alpha_i = \alpha(x_i) - \alpha(x_{i-1}), \ i = 1, 2, 3, \ldots,n \).

Let \( M_i = \sup f(x) \), \( m_i = \inf f(x) \) \((x_{i-1} \leq x \leq x_i)\) for each \( P \in \mathcal{P}(a, b) \). We define

\[
U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta \alpha_i \quad \text{and} \quad L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta \alpha_i
\]

are called the **Upper and Lower Riemann Stieltjes sums** respectively. Further we define

\[
\int_a^b f \, d\alpha = \inf_{P \in \mathcal{P}(a, b)} U(P, f, \alpha) \quad \text{and} \quad \int_a^b f \, d\alpha = \sup_{P \in \mathcal{P}(a, b)} L(P, f, \alpha),
\]

where the inf and the sup are taken over all partitions \( P \) of \([a, b]\), are called the **Upper and Lower Riemann Stieltjes integrals** of \( f \) over \([a, b]\), respectively. If the upper and lower Riemann Stieltjes integrals are equal, we say that \( f \) is **Riemann Stieltjes integrable** on \([a, b]\).

#### 1.7.2.

(i) If \( f: [a, b] \rightarrow \mathbb{R} \) is bounded function and \( \alpha \) be a monotonically increasing function on \([a, b]\). Let \( P \) be any Partition of \([a, b]\). Then the upper and lower Riemann-Stieltjes integrals of \( f \) with respect to \( \alpha \) are bounded.

(ii) If \( P^* \) is a refinement of the partition \( P \) of \([a, b]\), then

\[
L(P, f, \alpha) \leq L(P^*, f, \alpha) \quad \text{and} \quad U(P^*, f, \alpha) \leq U(P, f, \alpha).
\]

(iii) If \( P_1, P_2 \) are any two partitions of \([a, b]\) then \( L(P_1, f, \alpha) \leq U(P_2, f, \alpha) \) and

\[
L(P_2, f, \alpha) \leq U(P_1, f, \alpha).
\]

(iv)

\[
\int_a^b f \, d\alpha \leq \int_a^b f \, d\alpha
\]

(v) \( f \in R(\alpha) \) on \([a, b]\) if and only if for every \( \varepsilon > 0 \) there exists a partition \( P \) such that

\[
U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon
\]

(vi) If for every \( \varepsilon > 0 \) there exists a partition \( P = \{x_0, x_1, \ldots, x_n\} \) such that \( U(P, f, \alpha) - L(P, f, \alpha) < \varepsilon \) holds and if \( s_i, t_i \) are arbitrary points in \([x_{i-1}, x_i]\), then

\[
\sum_{r=1}^{n} |s_r - t_r| < \varepsilon.
\]
\[ \sum_{i=1}^{n} |f(s_i) - f(t_i)| \Delta \alpha_i < \varepsilon. \]

Further if \( f \in R(\alpha) \), then
\[ \left| \sum_{i=1}^{n} f(t_i) \Delta \alpha_i - \int_a^b f \, d\alpha \right| < \varepsilon. \]

(vii) If \( f \) is continuous on \([a, b]\) and \( \alpha \) is monotonic increasing on \([a, b]\), then \( f \in R(\alpha) \) on \([a, b]\).

(viii) If \( f \) is monotonic on \([a, b]\) and if \( \alpha \) is continuous on \([a, b]\), then \( f \in R(\alpha) \).

(ix) Suppose \( f \) is bounded on \([a, b]\), \( f \) has only finitely many points of discontinuity on \([a, b]\), and \( \alpha \) is continuous at every point at which \( f \) is discontinuous. Then \( f \in R(\alpha) \).

(x) Suppose \( f \in R(\alpha) \) on \([a, b]\), \( m \leq f \leq M \), \( \phi \) is continuous on \([m, M]\), and \( h(x) = \phi(f(x)) \) on \([a, b]\). Then \( h \in R(\alpha) \) on \([a, b]\).

1.7.3. Properties of Reimann-Stietjes

(a) If \( f_1, f_2 \in R(\alpha) \) on \([a, b]\), then \( f_1 + f_2 \in R(\alpha) \), \( cf \in R(\alpha) \) for every constant \( c \), and
\[ \int_a^b (f_1 + f_2) \, d\alpha = \int_a^b f_1 \, d\alpha + \int_a^b f_2 \, d\alpha ; \quad \int_a^b cf \, d\alpha = c \int_a^b f \, d\alpha . \]

(b) If \( f_1(x) \leq f_2(x) \) on \([a, b]\), then
\[ \int_a^b f_1 \, d\alpha \leq \int_a^b f_2 \, d\alpha . \]

(c) If \( f \in R(\alpha) \) on \([a, b]\) and if \( a < c < b \), then \( f \in R(\alpha) \) on \([a, c]\) and on \([c, b]\), and
\[ \int_a^c f \, d\alpha + \int_c^b f \, d\alpha = \int_a^b f \, d\alpha . \]

(d) If \( f \in R(\alpha) \) on \([a, b]\) and if \( |f(x)| \leq M \) on \([a, b]\), then
\[ \left| \int_a^b f \, d\alpha \right| \leq M[\alpha(b) - \alpha(a)]. \]

(e) If \( f \in R(\alpha_1) \) and \( f \in R(\alpha_2) \), then \( f \in R(\alpha_1 + \alpha_2) \) and
\[ \int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f \, d\alpha_1 + \int_a^b f \, d\alpha_2 ; \]

If \( f \in R(\alpha) \) and \( c \) is positive constant, then \( f \in R(c\alpha) \) and
\[ \int_a^b f \, d(c\alpha) = c \int_a^b f \, d\alpha . \]

(f) If \( f \in R(\alpha) \) and \( g \in R(\alpha) \) on \([a, b]\) then \( fg \in R(\alpha) \), \( |f| \in R(\alpha) \) and
\[ \left| \int_a^b f \, d\alpha \right| \leq \int_a^b |f| \, d\alpha . \]
(g) If \( a < s < b \), \( f \) is bounded on \([a, b]\), \( f \) is continuous at \( s \) and \( \alpha(x) = I(x-s) \),

then \( \int_a^b f \, d\alpha = f(s) \).

(h) Suppose \( c_n \geq 0 \) for \( n = 1, 2, 3, \ldots \), \( \Sigma c_n \) converges, \( \{s_n\} \) is a sequence of distinct points in \((a, b)\), and

\[
\alpha(x) = \Sigma_{n=1}^{\infty} c_n I(x-s_n).
\]

Then \( \int_a^b f \, d\alpha = \Sigma_{n=1}^{\infty} c_n f(s_n) \).

(i) Assume \( \alpha \) increases monotonically and \( \alpha' \in R[a, b] \). Let \( f \) be a bounded real function on \([a, b]\). Then \( f \in R(\alpha) \) if and only if \( f \alpha' \in R[a, b] \). Moreover

\[
\int_a^b f \, d\alpha = \int_a^b f(x) \alpha'(x) \, dx.
\]

(j) Suppose \( \phi \) is a strictly increasing continuous function that maps an interval \([A, B]\) onto \([a, b]\). Suppose \( \alpha \) is monotonically increasing on \([a, b]\) and \( f \in R(\alpha) \) on \([a, b]\). Define \( \beta \) and \( g \) on \([A, B]\) by \( \beta(y) = \alpha(\phi(y)) \), \( g(y) = f(\phi(y)) \).

Then \( g \in R(\beta) \) and \( \int_A^B g \, d\beta = \int_a^b f \, d\alpha \).

(k) Let \( f \in R[a, b] \). For \( a \leq x \leq b \), put \( F(x) = \int_a^x f(t) \, dt \). Then \( F \) is continuous on \([a, b]\).

Furthermore, if \( f \) is continuous at a point \( x_0 \) of \([a, b]\), then \( F \) is differentiable at \( x_0 \), and \( F'(x_0) = f(x_0) \). In other word “The integral of a Riemann integrable function is continuous and is differentiable if \( f \) is continuous”.

(l) If \( f \in R[a, b] \) and if there is a differentiate function \( F \) on \([a, b]\) such that \( F' = f \), then

\[
\int_a^b f(x) \, dx = F(b) - F(a).
\]

(m) Suppose \( F \) and \( G \) are differentiate functions on \([a, b]\), \( F' = f \in R[a, b] \) and \( G' = g \in R[a, b] \). Then

\[
\int_a^b F(x)g(x) \, dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x) \, dx.
\]

1.7.4. Integration of vector-valued functions. Let \( f_1, \ldots, f_k \) be real functions on \([a, b]\), and let \( \mathbf{f} = (f_1, \ldots, f_k) \) be the corresponding mapping of \([a, b]\) into \( R^k \). If \( \alpha \) increases monotonically on \([a, b]\), we say that \( \mathbf{f} \in R(\alpha) \) if \( f_j \in R(\alpha) \) for \( j = 1, \ldots, k \). In this case, we have

\[
\int_a^b \mathbf{f} \, d\alpha = (\int_a^b f_1 \, d\alpha, \ldots, \int_a^b f_k \, d\alpha).
\]

i.e., \( \int f \, d\alpha \) is the point in \( R^k \) whose \( j^{th} \) co-ordinate is \( \int_a^b f_j \, d\alpha \).
1.7.5. Rectifiable Curves. A continuous mapping $\gamma$ of an interval $[a, b]$ into $\mathbb{R}^k$ is said to be a curve
in $\mathbb{R}^k$. We may also say that $\gamma$ is a curve on $[a, b]$. If
(1) $\gamma$ is one-to-one then it called an arc.
(2) If $\gamma(a) = \gamma(b)$ then it is said to be a closed curve.
(3) The length of $\gamma$ which is denoted by $\mathcal{L}(\gamma)$ and defined by and if $\mathcal{L}(\gamma) < \infty$ then $\gamma$ is said
to be rectifiable.

(4) If $\gamma'$ is continuous on $[a, b]$, then $\gamma$ is rectifiable, and $\mathcal{L}(\gamma) = \int_a^b |\gamma'(t)| \, dt$.

1.7.6. Rearrangements of terms of a Series: Let $\{k_n\}$, $n = 1, 2, 3, \ldots$ be a sequence in which
every positive integer is non-repeating (In other words $\{k_n\}$ is an injective map on the set of positive integer into itself). We put $a_n = a_{k_n}$, $n = 1, 2, 3, \ldots$, then $\sum a'_n$ is called the
rearrangement of the a given $\sum a_n$.

(a) The terms of a conditionally convergent series can be rearrange such that the new series may
be convergent, divergent or oscillatory.

(b) The sum of every rearrangements of a absolutely convergent series is remains unchanged.

1.8. Assignments/Check your Progress.

1. Suppose $\alpha$ increases on $[a, b]$, $f(x) = k \forall x \in [a, b]$. Then prove that $f \in \mathcal{R}(\alpha)$ and
$$\int_a^b f \, d\alpha = \alpha(b) - \alpha(a)$$

2. If $f(x) = x$ and $\alpha(x) = x^2$ on $[0, 1]$. Is $f \in \mathcal{R}(\alpha)$, if yes then find its value of $\int_0^1 f \, d\alpha$. Ans 2/3.

3. If $f(x) = \alpha(x) = x^2$ on $[0, 1]$. Is $f \in \mathcal{R}(\alpha)$, if yes then find its value of $\int_0^1 f \, d\alpha$. Ans 1/2.

4. Suppose $\alpha$ increases on $[a, b]$ and is continuous at $x_0$, $a \leq x_0 \leq b$, $f(x_0) = 1$, and $f(x) = 0$ if $x \neq x_0$.
Prove that $f \in \mathcal{R}(\alpha)$ and that $\int f \, d\alpha = 0$.

5. Suppose $f \geq 0,f$ is continuous on $[a, b]$, and $\int_a^b f(x) \, dx = 0$. Prove that $f(x) = 0$ for all $x \in [a, b]$.

6. (i) If $f, g \in \mathcal{R}(\alpha)$, $g \geq 0$ and $\int_a^b f \, d\alpha = 1 = \int_a^b g \, d\alpha$, then $\int_a^b f \, g \, d\alpha \leq 1$.

   where $1/p + 1/q = 1$.

   (ii) If $f$ and $g$ are complex functions in $\mathcal{R}(\alpha)$, then
$$\left| \int_a^b f \, g \, d\alpha \right| \leq \left( \int_a^b |f|^p \, d\alpha \right)^{1/p} \left( \int_a^b |g|^q \, d\alpha \right)^{1/q}.$$  

7. Suppose $\alpha$ increases monotonically on $[a, b]$, $g$ is continuous, and $g(x) = G'(x)$ for $a \leq x \leq b$.
Prove that
$$\int_a^b \alpha(x) g(x) \, dx = G(b) \alpha(b) - G(a) \alpha(a) - \int_a^b G \, d\alpha.$$

8. Let $\gamma_1, \gamma_2, \gamma_3$ be curves in the complex plane, defined on $[0, 2\pi]$ by
$$\gamma_1(t) = e^{it}, \gamma_2(t) = e^{2it}, \gamma_3(t) = e^{2\sin(1/\gamma)}$$
Show that these three curves have the same range, that $\gamma_1$ and $\gamma_2$ are rectifiable, that the length
of $\gamma_1$ is $2\pi$, that the length of $\gamma_2$ is $4\pi$, and that $\gamma_3$ is not rectifiable.
9. Let $\alpha$ be continuous function of bounded variation on $[a, b]$, assume $g \in R(\alpha)$ on $[a, b]$ and define $\beta(x) = \int_a^x g \, d\alpha$ if $x \in [a, b]$ then prove that if $f$ is increasing on $[a, b]$ there exits a point $x_0$ in $[a, b]$ such that

$$\int_a^b f \, d\beta = f(a)\int_a^{x_0} g \, d\alpha + f(b)\int_{x_0}^b g \, d\alpha.$$

Furthermore if $f$ is continuous on $[a, b]$ then

$$\int_a^b f g \, d\alpha = f(a)\int_a^{x_0} g \, d\alpha + f(b)\int_{x_0}^b g \, d\alpha.$$

1.9 . Points for discussion/Clarification.

1.10 Suggested Study material

5. H.K. Pathak - Real Analysis
6. Lecture Notes on Real analysis Richard F. Bass
Sequences and Series of Functions

Introduction: The sequences \( \{f_n\} \) whose terms are real or complex-valued functions having a common domain on the real line \( \mathbb{R} \) or in the complex plane \( \mathbb{C} \). For each \( x \) in the domain we can form another sequence \( \{f_n(x)\} \) whose terms are the corresponding function values. Let \( S \) denote the set of \( x \) for which \( \{f_n(x)\} \) converges. The function \( f \) defined by the equation 
\[ f(x) = \lim_{n \to \infty} f_n(x), \] if \( x \in S \) is called the limit function of the sequence \( \{f_n\} \), and we say that \( \{f_n\} \) converges pointwise to \( f \) on the set \( S \). If each function of a sequence \( \{f_n\} \) has a certain property, such as continuity, differentiability, or integrability, to what extent is this property transferred to the limit function? For example, if each function \( f_n \) is continuous at \( c \), is the limit function \( f \) also continuous at \( c \)? We shall see that, in general, it is not. In fact, we shall find that pointwise convergence is usually not strong enough to transfer any of the properties mentioned above from the individual terms \( f_n \) to the limit function \( f \). Therefore we are led to study stronger methods of convergence that do preserve these properties. The most important of these is the notion of uniform convergence.

Objective: As in previous classes we have been studied the concept and properties of sequences. In this unit we deal with sequences \( \{f_n\} \) of real or complex valued functions having domain in \( \mathbb{R} \) or in \( \mathbb{C} \). For every \( x \) in the domain we can form a sequence \( \{f_n(x)\} \). For example if \( f_n : [0, 1] \to \mathbb{R} \) defined by \( f_n(x) = x^n \), \( \forall \ n \in \mathbb{N} \), then we can form different sequences as \( \{0,0,0,\ldots\}, \{1/2,1/2^2,1/2^3,\ldots\}, \ldots \) etc., and the properties like convergence of these sequences depend on \( n \) as well as \( x \) in the domain. Here we shall discuss the pointwise and uniform convergence and other related concepts of sequences and series of functions.

2.1. Sequence and Series of Functions:

Definition 2.1.1: Let \( E \) be any non-empty set of \( \mathbb{R} \) and \( \forall \ n \in \mathbb{N} \) and \( f_n : E \to \mathbb{R} \) be any function then \( \{f_n\} \) is said to be a sequence of functions from \( E \) to \( \mathbb{R} \) and \( \Sigma f_n \) is said to be a series of functions from \( E \) to \( \mathbb{R} \). Clearly, for each \( x \in E \), we have a sequence \( \{f_n(x)\} \) and a series \( \Sigma f_n(x) \). For some values of \( x \) this sequence (or series) is convergent and divergent on the other values of \( x \), for example if \( f_n : \mathbb{N} \to \mathbb{R} \) defined by \( f_n(x) = x^n \), \( \forall \ n \in \mathbb{N} \), then \( \{f_n(x)\} \) is convergent only when \( x = 1 \) and divergent for other values of \( x \in \mathbb{N} \). However the series \( \Sigma x^n \) is divergent for all values of \( x \in \mathbb{N} \).

2.2. Pointwise and Uniform Convergence:

Pointwise convergence:

Definition 2.2.1: Let \( \{f_n\} \) be a sequence of functions defined on a set \( E \) and suppose that for each \( x \in E \) the sequence of numbers \( \{f_n(x)\} \) converges to a function \( f(x) \), i.e., 
\[ f(x) = \lim_{n \to \infty} f_n(x), \] for each \( x \in E \) ... (i) 
Then we say that \( \{f_n\} \) converges on \( E \) and \( f \) is the limit or limit function of \( \{f_n\} \) and the convergence of \( \{f_n\} \) is called pointwise convergence of \( \{f_n\} \) on \( E \). In other words the sequence \( \{f_n\} \) is said to be pointwise converges to \( f \) if for \( \varepsilon > 0 \) and each \( x \in E \), \( \exists \ n_0 = n_0(\varepsilon, x) \in \mathbb{N} \)
such that \( |f_n(x) - f(x)| < \varepsilon, \ \forall \ n \geq n_0 \). Remember that here \( n_0 \) is depend on \( x \) and \( \varepsilon \) both. Similarly if the series \( \Sigma f_n(x) \) converges for each \( x \in E \) and \( \Sigma_{n=1}^\infty f_n(x) = f(x) \) then the function \( f \) is called the sum (or pointwise sum) function of the series \( \Sigma f_n \). For example if \( f_0 : [0, 1] \to \mathbb{R} \) defined by \( f_n(x) = x^n, \ \forall \ n \in \mathbb{N} \) and if \( f(x) \) is the limit function of \( f_n(x) \) then \( f(x) = 0 \) if \( x \in (0, 1) \) and \( f(x) = 1 \) if \( x = 1 \). Therefore \( \{f_n\} \) converges pointwise. Similarly for all real \( x \) consider the geometric series \( \Sigma_{0}^{\infty} \{x^n/(1 + x^2)^n\} \) and if \( f(x) \) is the sum function of the series then \( f(x) = 0 \) if \( x = 0 \) and \( f(x) = 1 + x^2 \) if \( x \neq 0 \), therefore the series is pointwise converges to \( f(x) \).

Here a question arises if each function of the sequence \( \{f_n\} \) has a certain property like continuity, differentiability or integrability etc.; is the same property is true for the limit (or sum) function \( f \) ?

Consider if each function of the sequence \( \{f_n\} \) is continuous at a point \( c \), i.e., \( \lim_{n \to \infty} f_n(x) = f_n(c) \), then

\[ \lim_{n \to \infty} f(x) = f(c) \oplus \lim_{x \to c} f_n(x) = \lim_{n \to \infty} \lim_{x \to c} f_n(x) \text{ is true ?} \]

Therefore at this stage the question arises about continuity is that can we interchange the order of limits in (ii) ? We shall see that, in general that we can not i.e. either the limit does not exists or if it dose exist, it need not be equal to \( f(c) \). We shall discuss this situation by means of some examples that limit processes can not be interchanged in general without affecting the result.

**Example 2.2.1:** Consider the double sequence \( s_{m,n} = m/(m + n); \ m, n = 1, 2, 3, \ldots \).

Clearly it is the pointwise convergent sequence. Keeping \( n \) fixed and letting \( m \to \infty \), \( s_{m,n} \to 1 \) implies that \( \lim_{n \to \infty} \lim_{m \to \infty} s_{m,n} = 1 \). On the other hand, keeping \( m \) fixed and letting \( n \to \infty \), \( s_{m,n} \to 0 \) implies that \( \lim_{m \to \infty} \lim_{n \to \infty} s_{m,n} = 0 \).

Hence \( \lim_{n \to \infty} \lim_{m \to \infty} s_{m,n} \neq \lim_{m \to \infty} \lim_{n \to \infty} s_{m,n} \).

**Example 2.2.2:** Consider the sequence \( f_m(x) = \lim_{n \to \infty} (\cos m!\pi x)^{2n}; \ m = 1, 2, 3, \ldots \). Clearly it is the pointwise convergent sequence of continuous functions. When \( m!x \) is an integer then \( f_m(x) = 1 \) and for all other values of \( x \) \( f_m(x) = 0 \). If we suppose \( f(x) = \lim_{m \to \infty} f_m(x) \), then \( f(x) = 0 \) if \( x \) is irrational and \( f(x) = 1 \) if \( x \) is rational. Thus we have obtained a discontinuous limit function.

**Example 2.2.3:** Consider the sequence \( f_n(x) = \sin nx/\sqrt{n} \) (\( x \) is real and \( n = 1, 2, 3, \ldots \)). Clearly it is the pointwise convergent sequence. Since \( f(x) = \lim_{n \to \infty} f_n(x) = 0 \) for all \( x \). Then \( \lim_{n \to \infty} f_n'(0) = \lim_{n \to \infty} \sqrt{n} \cos nx = \infty \neq \lim_{n \to \infty} f'(0) = 0 \).

**Example 2.2.4:** Consider the sequence \( f_n(x) = n^2x(1 - x^2)^n \) (\( 0 \leq x \leq 1 \), \( n = 1, 2, 3, \ldots \)). Clearly it is the pointwise convergent sequence. If \( f(x) = \lim_{n \to \infty} f_n(x) \), then \( f(x) = 0 \) for all \( x \in [0, 1] \). Now by simple calculation we have \( \int_{0}^{1} n^2x(1 - x^2)^n \ dx = n^2/(2n + 2) \to \infty \) as \( n \to \infty \), whereas \( \int_{0}^{1} [\lim_{n \to \infty} f_n(x)]\ dx = 0 \). Hence \( \lim_{n \to \infty} \int_{0}^{1} f_n(x)\ dx \neq \int_{0}^{1} [\lim_{n \to \infty} f_n(x)]\ dx \).

From the above examples we conclude that the pointwise convergence is not sufficient to allow us to interchange the order of limits, limit and differentiation, limit and integration. Therefore pointwise convergence is usually not strong enough to transfer any of the properties mentioned above from the individual term \( f_n \) to the limit function \( f \). Hence we are led to the study stronger
Uniform Convergence:

**Definition 2.2.2:** A sequence \( \{f_n\} \) of functions defined on a set \( E \) is said to be **converges uniformly** on \( E \) to a function \( f \) if for \( \varepsilon > 0 \) \( \exists \ n_0 = n_0(\varepsilon) \in \mathbb{N} \) such that \( |f_n(x) - f(x)| < \varepsilon, \forall n \geq n_0 \), for all \( x \in E \).

Symbolically it is written \( f_n \to f \) uniformly on \( E \) or \( \lim_{n \to \infty} f_n(x) = f(x) \) uniformly on \( E \) or \( f_n \to f \) uniformly on \( E \). Clearly uniform convergence implies pointwise convergence but not conversely. The difference between these concepts is this: If \( \{f_n\} \) converges pointwise on \( E \), then there exists a function \( f \) such that, for every \( \varepsilon > 0 \) and each \( x \in E \) \( \exists n_0 = n_0(\varepsilon, x) \in \mathbb{N} \) depending on \( \varepsilon \) and \( x \) such that \( |f_n(x) - f(x)| < \varepsilon \forall n \geq n_0 \).

But if \( \{f_n\} \) converges uniformly on \( E \) it is possible, for every \( \varepsilon > 0 \) \( \exists n_0 = n_0(\varepsilon) \in \mathbb{N} \) depending only on \( \varepsilon \) such that \( |f_n(x) - f(x)| < \varepsilon \forall n \geq n_0 \). Similarly a series \( \sum f_n(x) \) converges uniformly on \( E \) if the sequence \( \{s_n\} \) of partial sums defined by \( \sum_{i=1}^{n} f_i(x) = s_n(x) \) converges uniformly on \( E \). Suppose \( s_n \to s \) uniformly on \( E \), then we write \( \sum_{n}^{\infty} f_i(x) = s(x) \) uniformly on \( E \). In other words the series \( \sum f_n(x) \) is said to be uniformly convergent to \( s(x) \) if for \( \varepsilon > 0 \) \( \exists n_0 = n_0(\varepsilon) \in \mathbb{N} \) independent of \( x \), such that \( |s_n(x) - s(x)| < \varepsilon \forall n \geq n_0 \) for all \( x \in E \).

Let each term of the uniformly convergent sequence \( \{f_n\} \) is real-valued then by \( |f_n(x) - f(x)| < \varepsilon \forall n \geq n_0 \) we mean that \( f(x) - \varepsilon < f_n(x) < f(x) + \varepsilon, \forall n \geq n_0 \) implies that the entire graph of \( f_n \) (i.e. the set \( \{(x, y) : y = f_n(x), x \in E\} \) lies within a “band” of height \( 2\varepsilon \) situated symmetrically about the graph of \( f \) as shown in fig (i). This is a geometrical interpretation of uniform convergence.

A sequence \( \{f_n\} \) is said to be **uniformly bounded** on the set \( E \) if there exists a constant \( M > 0 \) such that \( |f_n(x)| \leq M, \forall x \in E \) and all \( n \) and the number \( M \) is called the uniform bound for \( \{f_n\} \). Clearly if each individual function is bounded and \( f_n \to f \) uniformly on \( E \), then the sequence \( \{f_n\} \) is uniformly bounded on \( E \). If there is a point \( x \in E \) is such that a sequence \( \{f_n\} \) is not convergent uniformly in any neighborhood of \( x \) then \( x \) is called the **point of non-uniform convergence** of the sequence \( \{f_n\} \).

Now we illustrate the notion of uniform convergence through some examples:

**Example 2.2.6:** Show that the sequence \( \{f_n\} \) where \( f_n : \mathbb{R} \to \mathbb{R} \) defined by \( f_n(x) = x/n \) \( \forall x \in \mathbb{R}, n \in \mathbb{N} \), convergent pointwise but not uniformly.
Solution: Let for arbitrary \( \varepsilon > 0 \) there exists \( m \in \mathbb{N} \) such that \( |f_n(x) - f(x)| < \varepsilon \) \( \forall n \geq m \). Since here \( f_n(x) = x/n \rightarrow 0 = f(x) \ \forall x \in \mathbb{R} \), therefore \( |x/n - 0| < \varepsilon \) \( \forall n \geq m \). In particular \( |x|/m < \varepsilon \) or, \( m > |x|/\varepsilon \). Clearly \( m \) depends on \( x \) and \( \varepsilon \) both. Hence the sequence \( \{f_n\} \) convergent pointwise but not uniformly. But it convergent uniformly on \([0, 1]\), since here \( m > |x|/\varepsilon > 1/\varepsilon \), clearly \( m \) depends only on \( \varepsilon \).

Example 2.2.7: Show that the sequence \( \{f_n\} \) where \( f_n : [0, 1] \rightarrow \mathbb{R} \) defined by \( f_n(x) = x^n \ \forall x \in [0, 1], \ n \in \mathbb{N} \), convergent pointwise but not uniformly.

Solution: Clearly this sequence is pointwise convergent, since \( f(x) = \lim_{n \rightarrow \infty} f_n(x) = \begin{cases} 0 & \text{if } x \in [0, 1) \\ 1 & \text{if } x = 1. \end{cases} \)

Let for arbitrary \( \varepsilon > 0 \) there exists \( m \in \mathbb{N} \) such that \( |f_n(x) - f(x)| < \varepsilon \) \( \forall n \geq m \). Or, \( |x^n - 0| < \varepsilon \ \forall n \geq m \) and \( x \in [0, 1) \). In particular \( x^m < \varepsilon \) or \( 1/x^m > 1/\varepsilon \) \( \Rightarrow m \log(1/x) > \log(1/\varepsilon) \Rightarrow m > [\log(1/\varepsilon) / \log(1/x)] \). Clearly \( m \) depends on \( x \) and \( \varepsilon \) both and it is not possible to determine \( m \in \mathbb{N} \) such that \( |f_n(x) - f(x)| < \varepsilon \) \( \forall n \geq m \) \( \forall x \in [0, 1] \).

Hence the sequence \( \{f_n\} \) convergent pointwise but not uniformly.

Example 2.2.8: Show that the sequence \( \{f_n\} \) where \( f_n : (0, \infty) \rightarrow \mathbb{R} \) defined by \( f_n(x) = 1 + \{x / (1 + nx)\} \ \forall x \in (0, \infty), n \in \mathbb{N} \), converges uniformly.

Solution: Since \( f(x) = \lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} [1 + \{x / (1 + nx)\}] = 1 \ \forall x \in (0, \infty) \) and \( n \in \mathbb{N} \).

Let for arbitrary \( \varepsilon > 0 \) there exists \( m \in \mathbb{N} \) such that \( |f_n(x) - f(x)| < \varepsilon \) \( \forall n \geq m \). Or, \( |1 + \{x / (1 + nx)\} - 1| < \varepsilon \ \forall n \geq m \). In particular \( \{x / (1 + mx)\} < \varepsilon \)\( \Rightarrow (1 + mx)/x > 1/\varepsilon \) or, \( 1 > 1/x > (1/\varepsilon) - m \) implies that \( m > (1/\varepsilon) - 1 \). Here \( m \) is depend only on \( \varepsilon \). Hence the sequence \( \{f_n\} \) convergent uniformly.

Example 2.2.9: Show that the sequence \( \{f_n\} \) where \( f_n : \mathbb{R} \rightarrow \mathbb{R} \) defined by \( f_n(x) = 1/(1 + nx^2) \), \( n \in \mathbb{N} \), not convergent uniformly if \( \mathbb{X} = \mathbb{R} \) but convergent uniformly if \( \mathbb{X} = (0, \infty) \).

Solution: Since \( f(x) = \lim_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} \{1 / (1 + nx^2)\} = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{if } x \neq 0. \end{cases} \)

Hence as examples 2.2.8 and 2.2.9 the sequence \( \{f_n\} \) not convergent uniformly if \( \mathbb{X} = \mathbb{R} \) but convergent uniformly if \( \mathbb{X} = (0, \infty) \).

2.3. Cauchy Criterion for uniform convergence: Following is the Cauchy criterion for uniform convergence.

Theorem 2.3.1. (Cauchy Criterion for uniform convergence for sequence) Any sequence \( \{f_n\} \) of functions defined on a set \( \mathbb{E} \) is converges uniformly on \( \mathbb{E} \) to a function \( f \) iff the following condition (called the Cauchy condition) hold:

For every \( \varepsilon > 0 \), \( \exists p \in \mathbb{N} \) such that \( |f_n(x) - f_m(x)| < \varepsilon \) \( \forall n, m \geq p \) and \( \forall x \in \mathbb{E} \) … … (i)

Proof. Suppose that \( f_n \rightarrow f \) uniformly on \( \mathbb{E} \), then for every \( \varepsilon > 0 \), \( \exists p = p(\varepsilon) \in \mathbb{N} \) such that \( |f_n(x) - f(x)| < \varepsilon/2 \) \( \forall n \geq p \) and \( \forall x \in \mathbb{E} \).

Therefore,
\[
|f_n(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f(x) - f_m(x)| < \varepsilon/2 + \varepsilon/2 = \varepsilon, \quad \forall \ m, n \geq p \text{ and } \forall \ x \in E.
\]

Conversely, suppose that the condition (1) satisfied for all \( x \in E \). Then by Cauchy criterion of convergence of sequences for fixed \( x \in E \) the sequence \( \{f_n(x)\} \) is convergent and say converges to a function \( f(x) \). Now for \( \varepsilon > 0 \) using the condition (i), we have
\[
|f_n(x) - f_{n+k}(x)| < \varepsilon \quad \forall n \geq p, k > 0 \text{ and } \forall \ x \in E
\] 

... (ii)

Letting \( k \to \infty \), by (ii), we have
\[
|f_n(x) - f(x)| < \varepsilon \quad \forall n \geq p, \forall \ x \in E, \text{ yields } f_n \to f \text{ uniformly on } E.
\]

**Note:** Sometime as an the immediate consequence of definition of uniform convergence if \( \lim_{k \to \infty} f_n(x) = f(x) \ (x \in E) \) and put \( M_n = \sup_{x \in E}|f_n(x) - f(x)| \). Then \( f_n \to f \) uniformly on \( E \) iff \( M_n \to 0 \) as \( n \to \infty \).

**Theorem 2.3.2. (Cauchy Criterion for uniform convergence for series)** A series \( \sum f_n \) of functions defined on a set \( E \) converges uniformly on \( E \) to a function \( f \) iff the following condition (called the Cauchy condition) hold:

For every \( \varepsilon > 0 \), \( \exists m \in \mathbb{N} \) such that \( \sum_{k=n+1}^{n+p} f_k(x) < \varepsilon \quad \forall n \geq m, p = 1, 2, 3, \ldots \) and \( \forall x \in E \)

**Proof.** Consider the sequence \( \{s_n(x)\} \) of partial sum of the \( \sum f_n \), where
\[
s_n(x) = f_1(x) + f_2(x) + f_3(x) + \ldots + f_n(x).
\]

Then
\[
|s_{n+p}(x) - s_n(x)| = |f_{n+1}(x) + f_{n+2}(x) + f_{n+3}(x) + \ldots + f_{n+p}(x)| = \sum_{k=n+1}^{n+p} f_k(x)
\]

Suppose that the series \( \sum f_n \) converges uniformly to \( f \) on \( E \) iff \( s_n(x) \to f(x) \) uniformly. Now by Cauchy criterion for uniformly convergence of sequence of functions sequence \( \{s_n(x)\} \) converges uniformly to \( f(x) \)

iff for every \( \varepsilon > 0 \), \( \exists m \in \mathbb{N} \) such that \( |s_{n+p}(x) - s_n(x)| < \varepsilon \quad \forall n \geq m \text{ and } \forall x \in E \)

iff \( \sum_{k=n+1}^{n+p} f_k(x) < \varepsilon \quad \forall n \geq m \text{ and } \forall x \in E \).

**2.4. Test for uniform convergence:** Here we give some test for uniform convergence.

**Theorem 2.4.1. (M_n-test for sequence):** A sequence \( \{f_n\} \) of functions defined on a set \( E \) and if
\[
\lim_{k \to \infty} f_n(x) = f(x) \ (x \in E) \text{ and } M_n = \sup_{x \in E}|f_n(x) - f(x)| \). Then \( f_n \to f \) uniformly on \( E \) iff \( M_n \to 0 \) as \( n \to \infty \).

**Proof.** Suppose that \( f_n \to f \) uniformly on \( E \), then for every \( \varepsilon > 0 \), \( \exists p = p(\varepsilon) \in \mathbb{N} \) such that
\[
|f_n(x) - f(x)| < \varepsilon, \quad \forall n \geq p \text{ and } \forall x \in E.
\]

\[
M_n = \sup_{x \in E}|f_n(x) - f(x)| < \varepsilon, \quad \forall n \geq p.
\]

\[
M_n \to 0 \text{ as } n \to \infty.
\]

Conversely, suppose that \( M_n \to 0 \) as \( n \to \infty \), then for \( \varepsilon > 0 \), \( \exists p \in \mathbb{N} \) such that
\[
|f_n(x) - f(x)| < \varepsilon, \quad \forall n \geq p \text{ and } \forall x \in E \implies M_n < \varepsilon, \quad \forall n \geq p \text{ and } \forall x \in E,
\]

\[
\sup_{x \in E}|f_n(x) - f(x)| < \varepsilon, \quad \forall n \geq p
\]

\[
|f_n(x) - f(x)| < \varepsilon, \quad \forall n \geq p \text{ and } \forall x \in E, \text{ yields } f_n \to f \text{ uniformly on } E.
\]
Theorem 2.4.2. (Weierstrass M-test for series): A sequence \( \{f_n\} \) of functions defined on a set \( E \) and suppose that \( |f_n(x)| \leq M_n \) (\( \forall x \in E, n = 1, 2, 3, \ldots \)). Then series \( \Sigma f_n \) converges uniformly (and absolutely) on \( E \) if \( \Sigma M_n \) converges.

**Proof.** Suppose that if \( \Sigma M_n \) converges, then by Cauchy’s criterion for convergence of series for every \( \varepsilon > 0 \), \( \exists \ m \in \mathbb{N} \) such that
\[
\sum_{k=n+1}^{n+p} |M_k| < \varepsilon, \quad \forall n \geq m, \ p = 1, 2, 3, \ldots
\]
Since \( |f_n(x)| \leq M_n \) (\( \forall x \in E, n = 1, 2, 3, \ldots \)). Therefore,
\[
\sum_{k=n+1}^{n+p} |f_k(x)| = |f_{n+1}(x) + f_{n+2}(x) + f_{n+3}(x) + \ldots + f_{n+p}(x)| \\
\leq |f_{n+1}(x)| + |f_{n+2}(x)| + |f_{n+3}(x)| + \ldots + |f_{n+p}(x)| \\
\leq M_{n+1} + M_{n+2} + M_{n+3} + \ldots + M_{n+p} = \sum_{k=n+1}^{n+p} M_k < \varepsilon \ \forall n \geq m, \ p = 1, 2, 3, \ldots \forall x \in E,
\]
which implies that \( \Sigma f_n \) converges uniformly on \( E \).

**Example 2.4.1.** Show that the sequence \( \{f_n\} \) where \( f_n(x) = nx(1 - x)^n \) does not converge uniformly on \([0, 1]\).

**Solution.** Here \( f(x) = \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} [nx/(1 - x)^n] \ [\infty/\infty \text{ form}] \)
\[
= - \lim_{n \to \infty} [x/(1 - x)^n \log(1 - x)] \quad \text{[by L’ Hospital Law]} \\
= - \lim_{n \to \infty} [x (1 - x)^n / \log(1 - x)] \to 0, \text{ since } (1 - x)^n \to 0 \text{ as } n \to \infty.
\]
Hence \( f(x) = 0, \ \forall x \in [0, 1] \).
Now \( M_n = \sup\{|f_n(x) - f(x)| : x \in [0, 1]\} = \sup\{|nx(1 - x)^n| : x \in [0, 1]\} \)
\[
\geq n(1/n)[1 - (1/n)]^n \ [\text{for } x = 1/n \in [0, 1]] = [1 - (1/n)]^n \to e \neq 0 \text{ as } n \to \infty.
\]
Hence by theorem 2.4.1 the given sequence is not converges uniformly.

**Example 2.4.2.** Show that the series \( \Sigma f_n \) where \( f_n(x) = \cos nx / n^p \), \( n = 1, 2, 3, \ldots \) converge uniformly on \( R \) for \( p > 1 \).

**Solution.** Here since \( |f_n(x)| = |\cos nx / n^p| \leq 1/n^p, \ \forall x \in R \). Now the series on right hand side convergent for \( p > 1 \). Hence by Weierstrass’s M-test the given series convergence uniformly on \( R \), for \( p > 1 \).

**Note:** Weierstrass’s M-test is only sufficient but not necessary i.e., non-convergence of \( \Sigma M_n \) does not imply the nature of \( \Sigma f_n \). Although Weiestrass’s M-Test is applicable to a restricted class of series, i.e., series which are absolutely convergent as well, yet it is of great practical importance as this test can be applied to nearly all series which are frequently used in the literature. In all other cases we have to make use of more delicate tests, which we construct by analogy with those for series of arbitrary terms- Abel’s and Dirichlet’s test.

**Theorem 2.4.3. (Abel’s test for series):** A series \( \Sigma u_n(x)v_n(x) \) converges uniformly on \([a,b]\) if

(i) \( \Sigma u_n(x) \) is uniformly convergent in \([a, b]\)

(ii) the sequence \( \{v_n(x)\} \) is monotonic for every \( x \) in \([a, b]\)
(iii) the sequence \( \{v_n(x)\} \) is uniformly bounded in \([a, b]\).

**Proof.** Let \( R_{n, p}(x) \) and \( r_{n, p}(x) \) are the partial remainder of the series \( \sum u_n(x)v_n(x) \) and \( \sum u_n(x) \) respectively. Then

\[
R_{n, p}(x) = u_{n+1}(x)v_{n+1}(x) + u_{n+2}(x)v_{n+2}(x) + u_{n+3}(x)v_{n+3}(x) + \ldots + u_{n+p}(x)v_{n+p}(x),
\]

\[
r_{n, p}(x) = r_{n, 1}(x)v_{n+1}(x) + [r_{n, 2}(x) - r_{n, 1}(x)]v_{n+2}(x) + [r_{n, 3}(x) - r_{n, 2}(x)]v_{n+3}(x) + \ldots
\]

\[
+ [r_{n, p}(x) - r_{n, p-1}(x)]v_{n+p}(x),
\]

\[
= r_{n, 1}(x)(v_{n+1}(x) - v_{n+2}(x)) + r_{n, 2}(x)(v_{n+2}(x) - v_{n+3}(x)) + \ldots
\]

\[
+ r_{n, p-1}(x)(v_{n+p-1}(x) - v_{n+p}(x)) + r_{n, p}(x)v_{n+p}(x),
\]

\[\text{...(i)}\]

Now since the sequence \( \{v_n(x)\} \) is monotonic for every \( x \) in \([a, b]\) and so the sign of terms \( [v_{n+1}(x) - v_{n+2}(x)], [v_{n+2}(x) - v_{n+3}(x)], \ldots [v_{n+p-1}(x) - v_{n+p}(x)] \) are same for fixed value of \( x \) in \([a, b]\).

Further since the sequence \( \{v_n(x)\} \) is uniformly bounded in \([a, b]\), we have a positive number \( k \) such that \( |v_n(x)| \leq k \) for all \( x \) in \([a, b]\) and all \( n \in \mathbb{N} \). \[\text{...(ii)}\]

Again since the series \( \sum u_n(x) \) is uniformly convergent in \([a, b]\) the for a given \( \varepsilon > 0 \) there exists a positive integer \( m \) such that

\[
|\sum_{n=1}^{m} u_n(x)v_n(x)| < \frac{\varepsilon}{3k}, \quad \forall \ n \geq m \quad \text{...(iii)}
\]

for all \( x \) in \([a, b]\) and all positive integral value of \( p \). Hence from (i) and (iii), we have

\[
|R_{n, p}(x)| < (\frac{\varepsilon}{3k}) |v_{n+1}(x) - v_{n+2}(x)| + (\frac{\varepsilon}{3k}) |v_{n+2}(x) - v_{n+3}(x)| + \ldots
\]

\[
+ (\frac{\varepsilon}{3k}) |v_{n+p-1}(x) - v_{n+p}(x)| + (\frac{\varepsilon}{3k}) |v_{n+p}(x)|,
\]

\[
\leq (\frac{\varepsilon}{3k}) [v_{n+1}(x)] + [v_{n+2}(x)] + (\frac{\varepsilon}{3k}) [v_{n+p}(x)], \quad \text{[using (*)]}
\]

\[
< (\frac{\varepsilon}{3k}) 2k = \frac{\varepsilon}{3k} k = \varepsilon, \quad \forall \ n \geq m \text{ and all } x \text{ in } [a, b] \quad \text{[by (ii)]}
\]

Hence the series \( \sum u_n(x)v_n(x) \) is converges uniformly on \([a, b]\).

**Theorem 2.4.4. (Dirichlet’s test for series):** A series \( \sum u_n(x)v_n(x) \) is converges uniformly on \([a, b]\) if

(i) the sequence \( \{v_n(x)\} \) is a positive monotonic decreasing sequence converging uniformly to zero for all \( x \) in \([a, b]\)

(ii) \( s_n(x) = \sum_{n=1}^{i} u_i(x) \) is uniformly bounded in \([a, b]\).

**Proof.** Let \( R_{n, p}(x) \) is the partial remainder of the series \( \sum u_n(x)v_n(x) \). Then

\[
R_{n, p}(x) = u_{n+1}(x)v_{n+1}(x) + [u_{n+2}(x)v_{n+2}(x) + u_{n+3}(x)v_{n+3}(x) + \ldots + u_{n+p}(x)v_{n+p}(x),
\]

\[
= [s_{n+1}(x) - s_n(x)]v_{n+1}(x) + [s_{n+2}(x) - s_{n+1}(x)]v_{n+2}(x) + [s_{n+3}(x) - s_{n+2}(x)]v_{n+3}(x)
\]

\[
+ \ldots + [s_{n+p}(x) - s_{n+p-1}(x)]v_{n+p}(x),
\]

\[
= s_{n+1}(x)[v_{n+1}(x) - v_{n+2}(x)] + s_{n+2}(x)[v_{n+2}(x) - v_{n+3}(x)] + \ldots
\]

\[
+ s_{n+p}(x)[v_{n+p-1}(x) - v_{n+p}(x)] + s_{n+p}(x)v_{n+p}(x) - s_n(x)v_{n+1}(x), \quad \text{...(i)}
\]

Now since the sequence \( \{v_n(x)\} \) is positive monotonic decreasing for every \( x \) in \([a, b]\) therefore terms \( [v_{n+1}(x) - v_{n+2}(x)], [v_{n+2}(x) - v_{n+3}(x)], \ldots [v_{n+p-1}(x) - v_{n+p}(x)] \) are positive fixed value of \( x \) in \([a, b]\).

Further since the sequence \( \{s_n(x)\} \) is uniformly bounded in \([a, b]\), we have a positive number \( k \) such that \( |s_n(x)| < k \) for all \( x \) in \([a, b]\) and all \( n \in \mathbb{N} \). \[\text{...(ii)}\]

Hence from (i), we have

\[
|R_{n, p}(x)| \leq |s_{n+1}(x)| [v_{n+1}(x) - v_{n+2}(x)] + |s_{n+2}(x)| [v_{n+2}(x) - v_{n+3}(x)] + \ldots
\]

\[
+ |s_{n+p-1}(x)| [v_{n+p-1}(x) - v_{n+p}(x)] + |s_{n+p}(x)| v_{n+p}(x) - |s_n(x)| v_{n+1}(x),
\]

\[\text{...(iii)}\]
Now by (i) \( v_n(x) \to 0 \) uniformly, we have for \( \varepsilon > 0 \) there exists positive integer \( m \) such that \( |v_n(x)| < \varepsilon /2k, \quad \forall \ n \geq m \) and all \( x \) in \([a, b]\). Hence \( |R_{n,p}(x)| < \varepsilon, \quad \forall \ n \geq m \) and all \( x \) in \([a, b]\). Hence the series \( \sum u_n(x)v_n(x) \) is converges uniformly on \([a, b]\).

**Example 2.4.3.** Test for uniform convergence of series \( \sum((-1)^{n-1}/n)x^n \) in \([0, 1]\).

**Solution.** Let \( v_n(x) = x^n \) and \( u_n(x) = (-1)^{n-1}/n \). Then clearly the sequence \( \{v_n(x)\} \) is uniformly bounded and monotonically increasing on \([0, 1]\). Also the series \( \sum((-1)^{n-1}/n)x^n \) is convergent (uniformly). Hence by Abel’s test the series \( \sum u_n(x)v_n(x) = \sum((-1)^{n-1}/n)x^n \) is uniformly convergent on \([0, 1]\).

**Example 2.4.4.** Test for uniform convergence of series \( \sum((-1)^{n-1}/(n + x^2)) \) for all real \( x \).

**Solution.** Let \( v_n(x) = 1/(n + x^2) \) and \( u_n(x) = (-1)^{n-1} \). Then since \( s_n(x) = \sum_{k=1}^{n} u_k(x) = 0 \) or 1 according as \( n \) is even or odd, therefore \( s_n(x) \) is bounded for all \( n \) (or uniformly bounded). Also the sequence \( \{v_n(x)\} \) is positive monotonically decreasing converging to zero for all real \( x \). Hence by Dirichlet’s test the series \( \sum u_n(x)v_n(x) = \sum((-1)^{n-1}/(n + x^2)) \) is uniformly convergent for all real \( x \).

**Example 2.4.5.** Show that the series \( \sin x + (1/2)\sin 2x+ (1/3)\sin 3x + \ldots \) is uniform convergent in \( 0 < a \leq x \leq b < 2\pi \).

**Solution.** Let \( v_n(x) = 1/n \) and \( u_n(x) = \sin nx \). Then
\[
s_n(x) = \sin x + (1/2)\sin 2x + (1/3)\sin 3x + \ldots + (1/n)\sin nx
= [\sin (n + 1/2)x] \sin (nx/2) / \sin (x/2)
= [\sin (n + 1/2)x] \sin (nx/2) / \sin (x/2)]
\]
Hence \( |s_n(x)| = [\sin ((n + 1)/2)x] \sin (nx/2) / |\sin (x/2)| \leq 1/|\sin (x/2)| = |\csc(x/2)| < k, \) where \( k \) is the least upper bound of \( \csc(x/2) \) in \((0, 2\pi)\). Therefore \( s_n(x) \) is uniformly bounded. Also the sequence \( \{v_n(x)\} = \{1/n\} \) is positive monotonically decreasing converging to zero. Hence by Dirichlet’s test the series \( \sum u_n(x)v_n(x) = \sum(1/n)\sin nx \) is uniformly convergent in \( 0 < a \leq x \leq b < 2\pi \).

2.5. Uniform Convergence and Continuity:

**Theorem 2.5.1.** Suppose that the sequence \( \{f_n\} \) of functions converges uniformly to \( f \) on a set \( E \) in a metric space \( X \). Let \( x \) be the limit point of \( E \) and such that \( \lim_{t \to x} f_n(t) = A_n \) (\( n = 1, 2, 3, \ldots \)). Then \( \{A_n\} \) converges and \( \lim_{t \to x} f(t) = \lim_{n \to \infty} A_n \).

In other words, the conclusion is that \( \lim_{t \to x} \lim_{n \to \infty} f_n(t) = \lim_{n \to \infty} \lim_{t \to x} f_n(t) \).

**Proof.** Let \( \varepsilon > 0 \) be given, since \( \{f_n\} \) is uniform convergent, then by Cauchy’s criterion for uniform convergence, there exists a positive integer \( N = N(\varepsilon) \) such that
\[
|f_n(t) - f_m(t)| < \varepsilon \quad \forall \ m, n \geq N \quad \text{and} \quad \forall \ t \in E \quad \ldots \quad \ldots \quad (i)
\]
Letting \( t \to x \) in (i), we have \( |A_n - A_m| < \varepsilon \quad \forall \ m, n \geq N \), therefore \( \{A_n\} \) is a Cauchy sequence and so by Cauchy’s criterion for convergence of real sequences \( \{A_n\} \) converges, say to \( A \).
Next since \( \{f_n\} \) is uniform converges to \( f \) and so for given \( \varepsilon > 0 \) there exists a positive integer \( m_1 = m_1(\varepsilon) \) such that \( |f_n(t) - f(t)| < \varepsilon/3 \quad \forall \ n \geq m_1 \) and \( \forall \ t \in E \) ... ... (ii) Also, by convergence of \( \{A_n\} \) to \( A \), for given \( \varepsilon > 0 \) there exists a positive integer \( m_2 \) such that \( |A_n - A| < \varepsilon/3 \quad \forall \ n \geq m_2 \) ... ... ... ... (iii)

Also, since \( \lim_{t \to x} f(t) = \lim_{n \to \infty} A_n \), therefore for given \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that

\[
|f_n(t) - A_n| < \varepsilon/3 \quad \forall \ n \geq m_2 \quad \text{and} \quad d(t, x) < \delta.
\]

\( \Rightarrow \lim_{t \to x} f(t) = \lim_{n \to \infty} A_n \) or \( \lim_{t \to x} \lim_{n \to \infty} f_n(t) = \lim_{n \to \infty} \lim_{t \to x} f_n(t) \).

**Theorem 2.5.2.** Suppose that the series \( \Sigma f_n \) of functions converges uniformly to \( f \) on a set \( E \) in a metric space \( X \). Let \( x \) be the limit point of \( E \) and such that \( \lim_{t \to x} f_n(t) = A_n \) (\( n = 1, 2, 3, \ldots \)). Then \( \Sigma A_n \) converges and \( \lim_{t \to x} f(t) = \Sigma_{n=1}^{\infty} A_n \).

In other words, the conclusion is that \( \lim_{t \to x} \Sigma_{n=1}^{\infty} f_n(t) = \Sigma_{n=1}^{\infty} [\lim_{t \to x} f_n(t)] \).

**Proof.** Let \( \varepsilon > 0 \) be given, since \( \Sigma f_n \) is uniform convergent, then by Cauchy’s criterion for uniform convergence, there exists a positive integer \( N = N(\varepsilon) \) such that

\[
| \sum_{k=n+1}^{k=n+p} f_k(t) | < \varepsilon \quad \forall \ n \geq N, \ p \geq 1 \quad \text{and} \quad \forall \ t \in E \quad \ldots \quad \ldots \quad (i)
\]

Letting \( t \to x \) in (i), we have

\[
| \sum_{k=k+n}^{k=n+p} A_k | < \varepsilon \quad \forall \ m, \ n \geq N.
\]

Hence by Cauchy’s criterion for convergence of real series \( \Sigma A_n \) converges, say to \( A \).

Next since \( \Sigma f_n \) is uniform converges to \( f \) and so for given \( \varepsilon > 0 \) there exists a positive integer \( m_1 = m_1(\varepsilon) \) such that \( |\Sigma f_n(t) - f(t)| < \varepsilon/3 \quad \forall \ n \geq m_1 \) and \( \forall \ t \in E \) (1 \( \leq k \leq n \))... ... (ii) Also, by convergence of \( \Sigma A_n \) to \( A \), for given \( \varepsilon > 0 \) there exists a positive integer \( m_2 \) such that \( |\Sigma A_k - A| < \varepsilon/3 \quad \forall \ n \geq m_2 \) (1 \( \leq k \leq n \)) ... ... ... ... (iii)

Also, since \( \lim_{t \to x} f(t) = \lim_{n \to \infty} A_n \), therefore for given \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that

\[
|f_n(t) - A_n| < \varepsilon/3n \quad \text{and} \quad d(t, x) < \delta, \quad \forall \ n \geq m_2 \quad \text{and} \quad d(t, x) < \delta.
\]

\( \Rightarrow \lim_{t \to x} f(t) = \lim_{n \to \infty} A_n \) or \( \lim_{t \to x} \lim_{n \to \infty} f_n(t) = \lim_{n \to \infty} \lim_{t \to x} f_n(t) \).

**Theorem 2.5.3.** Suppose that the sequence \( \{f_n\} \) of continuous functions on a set \( E \), converges uniformly to \( f \) on a set \( E \) then \( f \) is also continuous on \( E \).

**Proof.** Let \( c \) be arbitrary point in \( E \), then each \( f_n \) is continuous on \( c \), then for given \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( |f_n(x) - f_n(c)| < \varepsilon/3, \ |x - c| < \delta \) ... ... ... ... (i)

Again since \( f_n \to f \) uniformly then for \( \varepsilon > 0 \) there exists positive integer \( m(\varepsilon) \) such that

\[
|f_n(x) - f(x)| < \varepsilon/3 \quad \forall \ n \geq m \quad \text{and} \quad \forall \ x \in E \quad \ldots \quad \ldots \quad (ii)
\]

In particular, \( |f_m(x) - f(x)| < \varepsilon/3 \) and \( |f_m(c) - f(c)| < \varepsilon/3 \).

Hence for \( |x - c| < \delta \), we have
\[|f(x) - f(c)| = |f(x) - f_m(x) + f_m(x) - f_m(c) + f_m(c) - f(c)|
\leq |f(x) - f_m(x)| + |f_m(x) - f_m(c)| + |f_m(c) - f(c)| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.\]
\[\Rightarrow f \text{ is continuous on arbitrary } c \text{ in } E.\]

**Theorem 2.5.4.** Suppose that the series \( \Sigma f_n \) of continuous functions on a set \( E \) converges uniformly to \( f \) on a set \( E \) in a metric space \( X \), then the sum function \( f \) is also continuous on \( E \).

**Proof.** Let \( s_n(x) \) is the \( n^{th} \) partial sum of the series \( \Sigma f_n \) then \( s_n(x) \) is also continuous on \( E \). Hence by definition of uniform convergence of series, the sequence \( \{s_n(x)\} \) of continuous functions converges uniformly to \( f \) on \( E \). Hence by theorem 5.3 the limit function \( f \) of the sequence \( \{s_n(x)\} \) (or the sum function of the series \( \Sigma f_n \)) is also continuous on \( E \).

**Remark:** Here we note that the converse of theorems 5.3 & 5.4 are not true in general i.e., a sequence (or series) of continuous function may converge to a continuous function but it may not converge uniformly. These facts are illustrated by following examples:

**Example 2.5.1:** Consider the sequence \( f_n(x) = n^2x(1 - x)^n \) for all \( x \in [0, 1] \) and \( n \in \mathbb{N} \) of continuous functions.

**Solution.** Here \( f(x) = \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} [n^2x/(1 - x)^n] \) \([\infty/\infty \text{ form}]\)
\[= -\lim_{n \to \infty} [2nx/(1 - x)^n \log(1 - x)] \) \([\infty/\infty \text{ form, by L’ Hospital Law}]\)
\[= \lim_{n \to \infty} 2x/(1 - x)^n \log(1 - x)^2 \quad \text{[by L’ Hospital Law]}\]
\[= \lim_{n \to \infty} 2x(1 - x)^n/\log(1 - x)^2 \to 0, \text{ since } (1 - x)^n \to 0 \text{ as } n \to \infty.\]
Hence \( f(x) = 0, \forall x \in [0, 1] \) is continuous function. Hence the sequence \( \{f_n(x)\} \) of continuous functions converges to a continuous function.

Now \( M_n = \sup\{|f_n(x) - f(x)| : x \in [0, 1]\} = \sup\{|n^2x(1 - x)^n| : x \in [0, 1]\} \]
\[\geq n^2(1/n)[1 - (1/n)]^n \quad \text{[for } x = 1/n \in [0, 1]\} = n(1 - (1/n))^n \to \infty \neq 0 \text{ as } n \to \infty.\]
Hence by \( M_n \) test the given sequence is not converges uniformly.

**Example 2.5.2:** Consider the series \( \Sigma f_n(x) = \Sigma[n^2x/(1 + n^3x^2)] \) - \((n-1)^2x/(1 + (n-1)^3x^2))\) for all \( x \in [0, 1] \) of continuous functions.

**Solution.** Let \( s_n(x) \) is the \( n^{th} \) partial sum of the series \( \Sigma f_n(x) \)
\[\text{i.e., } s_n(x) = \{x/(1 + x^2) - 0\} + \{(2^2x/(1 + 2^3x^2))\} - \{x/(1 + x^2)\}
\[+ \ldots + \{n^2x/(1 + n^3x^2)\} - \{(n-1)^2x/(1 + (n-1)^3x^2))\}\}
\[= n^2x/(1 + n^3x^2)\]

Here we have \( f(x) = \lim_{n \to \infty} f_n(x) = \lim_{n \to \infty} [n^2x/(1 + n^3x^2)] \) \([\infty/\infty \text{ form}]\)
\[= \lim_{n \to \infty} x/((1/n^2) + nx^2) \to 0, \text{ as } n \to \infty.\]
Hence \( f(x) = 0, \forall x \in [0, 1] \) is continuous function. Hence the series \( \Sigma f_n(x) \) of continuous functions converges to a continuous function.

Now \( M_n = \sup\{|f_n(x) - f(x)| : x \in [0, 1]\} = \sup\{|n^2x/(1 + n^3x^2)| : x \in [0, 1]\} \]
\[\geq n^2(1/n)^2/[1 + n^3(1/n^2)] \text{[for } x = 1/n^2 \in [0,1]\} = n^{1/2}/2 \to \infty \neq 0 \text{ as } n \to \infty.\]
Hence by \( M_n \)-test the sequence \( \{s_n(x)\} \) and consequently the given series is not converges uniformly.

As we discussed above that the converse of theorems 5.3 & 5.4 are not true in general. The following theorem asserts the condition under which the converse of these theorems is true.
Theorem 2.5.5. Suppose that E is compact and \( \{f_n\} \) is a sequence of continuous functions on K converges pointwise to a continuous function f on E such that \( f_n(x) \geq f_{n+1}(x) \) \( \forall x \in E, n = 1, 2, 3, \ldots \). Then \( f_n \to f \) uniformly on K.

Proof. Let \( g_n(x) = f_n(x) - f(x) \), then \( g_n \) is continuous function as \( f_n \) and \( f \) are continuous functions and \( g_n \geq g_{n+1} \). Thus to prove the theorem it is enough to prove that \( g_n \to 0 \) uniformly on E. For suppose that \( \varepsilon > 0 \) be given, since \( g_n \to 0 \) on E therefore for each \( x \in E, \exists \) a positive integer \( N = N(\varepsilon, x) \) such that \( |g_n(x)| < \varepsilon/2, \forall n \geq N \).

In particular

\[
|g_n(x)| < \varepsilon/2 \quad \ldots \quad \ldots
\]

Again by continuity of \( g_n \) there exists an open nbd \( G_x \) of \( x \in E \) such that

\[
|g_n(y) - g_n(x)| < \varepsilon/2, \forall y \in G_x
\]

\[
\Rightarrow |g_n(y)| - |g_n(x)| < \varepsilon/2, \forall y \in G_x
\]

\[
\Rightarrow |g_n(y)| < \varepsilon/2 + |g_n(x)| < \varepsilon/2 + \varepsilon/2 = \varepsilon, \forall y \in G_x \quad \text{[using (i)]}
\]

Further since \( g_n \geq g_{n+1} \) then \( g_n(x) \geq g_n(x) \) for \( n \geq N \).

\[
... \quad \ldots \quad \ldots
\]

Therefore from (ii) and (iii), we have \( |g_n(y)| < \varepsilon, \forall y \in G_x \)

\[
... \quad \ldots \quad \ldots
\]

Clearly the family \( G_x : x \in E \) is an open covering of E and E is compact therefore there exists a finite sub-covering of E.

i.e., there exists finite set \( \{x_1, x_2, x_3, \ldots, x_m\} \) such that \( E \subseteq \cup G_{x_i}, i = 1, 2, 3, \ldots, m \).

Remark: Here we remarked that the compactness of E is essential for uniform convergence, for example consider the sequence \( f_n(x) = 1/(nx + 1), 0 < x < 1 \).

Since clearly, \( f_n(x) \geq f_{n+1}(x) \) \( \forall x \in (0, 1) \) for all \( n \in \mathbb{N} \).

Further \( f(x) = \lim_{n \to \infty} f_n(x) = 0, \forall x \in (0, 1) \). But if we take \( \varepsilon = \frac{1}{2} \) there exists positive integer \( N \) such that \( |f_n(x) - 0| < \frac{1}{2}, \forall x \in (0, 1) \) and \( n \geq N \) implies that \( \frac{1}{2} < \frac{1}{2}, \) for \( x = 1/n \in (0, 1) \), a contradiction. Hence the uniform convergence is not possible because of (0, 1) is not compact.

2.6. Uniform Convergence and Riemann–Stieltjes integration:

Theorem 2.6.1. Let \( \alpha \) be monotonically increasing function on \( [a, b] \) and \( f_n \in R(\alpha) \) on \( [a, b] \), for \( n = 1, 2, 3, \ldots \), such that \( f_n \to f \) uniformly on \( [a, b] \). Then \( f \in R(\alpha) \) on \( [a, b] \) and

\[
\int_a^b f d\alpha = \lim_{n \to \infty} \int_a^b f_n d\alpha
\]

Proof. Let \( \varepsilon_n = \sup_{x \in [a,b]}|f_n(x) - f(x)| \), then for all \( x \in [a, b] \)

\[
|f_n(x) - f(x)| \leq \varepsilon_n \quad \text{or} \quad |f(x) - f_n(x)| \leq \varepsilon_n
\]

\[
\Rightarrow f_n - \varepsilon_n \leq f \leq f_n + \varepsilon_n \quad \text{Therefore by upper and lower Reimann-Stieltjes integration}
\]

\[
\int_a^b (f_n - \varepsilon_n) d\alpha \leq \int_a^b f d\alpha \leq \int_a^b (f_n + \varepsilon_n) d\alpha \quad \ldots \quad \ldots
\]
\[0 \leq \int_a^b f \, d\alpha - \int_a^b f \, d\alpha \leq \int_a^b (f_n + \epsilon_n) \, d\alpha - \int_a^b (f_n - \epsilon_n) \, d\alpha = 2\epsilon_n \int_a^b d\alpha = 2\epsilon_n [\alpha(b) - \alpha(a)]\]

Since \(f_n \to f\) uniformly, therefore by M-test \(\epsilon_n \to 0\) as \(n \to \infty\). Hence

\[0 \leq \int_a^b f \, d\alpha - \int_a^b f \, d\alpha \leq 0 \Rightarrow \int_a^b f \, d\alpha = \int_a^b f \, d\alpha \Rightarrow f \in R(\alpha)\text{ on } [a, b]\]

Therefore from equation (i), we have

\[\int_a^b (f_n - \epsilon_n) \, d\alpha \leq \int_a^b f \, d\alpha \leq \int_a^b (f_n + \epsilon_n) \, d\alpha\]

\[\Rightarrow \int_a^b f_n \, d\alpha - \epsilon_n \int_a^b d\alpha \leq \int_a^b f \, d\alpha \leq \int_a^b f_n \, d\alpha + \epsilon_n \int_a^b d\alpha\]

\[\Rightarrow \left| \int_a^b f \, d\alpha - \int_a^b f_n \, d\alpha \right| \leq \epsilon_n \int_a^b d\alpha = \epsilon_n [\alpha(b) - \alpha(a)] \to 0 \text{ as } n \to \infty.\]

\[\Rightarrow \int_a^b f \, d\alpha = \lim_{n \to \infty} \int_a^b f_n \, d\alpha.\]

**Theorem 2.6.2.** Let \(\alpha\) be monotonically increasing function on \([a, b]\) and \(f_n \in R(\alpha)\) on \([a, b]\), for \(n = 1, 2, 3, \ldots\), such that series \(\sum f_n\) converges uniformly to \(f\) on \([a, b]\). Then \(f \in R(\alpha)\) on \([a, b]\) and

\[\int_a^b f \, d\alpha = \sum_{n=1}^{\infty} \int_a^b f_n \, d\alpha\]

In other words the series can be integrated term – by- term.

**Proof.** Let \(s_n(x) = f_1(x) + f_2(x) + f_3(x) + \ldots + f_n(x)\), then \(s_n(x) \in R(\alpha)\) for all \(i, f_i(x) \in R(\alpha)\). Also since the series \(\sum f_n\) converges uniformly to \(f\) on \([a, b]\), therefore the sequence \(\{s_n(x)\}\) also, converges uniformly to \(f\) on \([a, b]\). Hence by theorem 6.1, \(f \in R(\alpha)\) and

\[\int_a^b f \, d\alpha = \lim_{n \to \infty} \int_a^b s_n \, d\alpha.\]

\[\Rightarrow \int_a^b f \, d\alpha = \lim_{n \to \infty} \sum_{i=1}^{n} \int_a^b f_i \, d\alpha = \lim_{n \to \infty} \sum_{i=1}^{n} \int_a^b f_i \, d\alpha = \sum_{n=1}^{\infty} \int_a^b f_n \, d\alpha.\]

**2.7. Uniform Convergence and Differentiation:** Analogous to the theorems 5.3 & 6.1 we can expect that if \(f_n \to f\) uniformly on \([a, b]\) and if \(f_n'\) for each \(n\), then \(f'\) exists and \(f_n' \to f'\) uniformly on \([a, b]\). However from the example 2.3, we seen that it is not true i.e., the uniform convergence of \(\{f_n\}\) does not implies about the sequence \(\{f_n'\}\), then for the assertion that \(f_n' \to f'\) if \(f_n \to f\), stronger hypotheses is required.
Theorem 2.7.1. Let \( \{f_n\} \) be a sequence of differentiable functions on \([a, b]\) and for some point \( x_0 \) in \([a, b]\) the sequence \( \{f_n(x_0)\} \) is convergent. If sequence \( \{f'_n\} \) converges uniformly on \([a, b]\) then sequence \( \{f_n\} \) converges uniformly on \([a, b]\) to a function \( f \), and

\[
f'(x) = \lim_{n \to \infty} f'_n(x) \quad x \in [a, b].
\]

Proof. Since for some point \( x_0 \) in \([a, b]\) the sequence \( \{f_n(x_0)\} \) is convergent and the sequence \( \{f'_n\} \) converges uniformly on \([a, b]\), then by Cauchy’s criterion \( \varepsilon > 0 \) there exists positive integer \( N \) such that

\[
|f_n(x_0) - f_m(x_0)| < \varepsilon/2, \quad \forall m, n \geq N \quad \ldots \quad \ldots \quad (i)
\]

and

\[
|f'_n(t) - f'_m(t)| < \varepsilon/2(b - a), \quad \forall m, n \geq N \quad \forall t \in [a, b]. \quad \ldots \quad (ii)
\]

Now by Mean value theorem for \( m, n \geq N \), we have

\[
|f_n(x) - f_m(t)| \leq |x - t| |f'_n(\xi)| \quad \text{and} \quad |f_m(x) - f_n(t)| \leq |x - t| |f'_m(\xi)|, \quad \forall t, x \text{ and } x < \xi < t.
\]

Then for \( m, n \geq N \),

\[
|\{f_n(x) - f_n(t)\} - \{f_m(x) - f_m(t)\}| \leq |x - t| |f'_n(\xi)| - |f'_m(\xi)| \leq |x - t|\varepsilon/2(b - a) \leq \varepsilon/2. \quad \ldots \quad (iii)
\]

Further using (i) and (iii), we have

\[
|f_n(x) - f_m(t)| = |f_n(x) - f_n(x_0) + f_n(x_0) - f_m(x_0) + f_m(x_0) - f_m(t)|
\]

\[
\leq |f'_n(x)| + |f'_m(x_0) - f'_m(t)| < \varepsilon/2 + \varepsilon/2 = \varepsilon. \quad \forall m, n \geq N \quad \text{and} \quad x \in [a, b].
\]

Hence by Cauchy’s criterion for uniform convergence the sequence \( \{f_n\} \) converges uniformly, say \( f_n \to f \) uniformly.

Now for fixed \( x \) in \([a, b]\), define,

\[
\phi_n(t) = \{f_n(t) - f_n(x)\} / (t - x) \quad \text{and} \quad \phi(t) = \{f(t) - f(x)\} / (t - x), \quad \ldots \quad \ldots \quad (iv)
\]

for all \( t \in [a, b] \), \( t \neq x \). Then

\[
\lim_{t \to x} \phi_n(t) = \lim_{t \to x} \{f_n(t) - f_n(x)\} / (t - x) = f'_n(x); \quad n = 1, 2, 3, \ldots \quad \ldots \quad (v)
\]

Now by (iii), we have

\[
|\phi_n(t) - \phi_m(t)| \leq \varepsilon/2(b - a), \quad \forall m, n \geq N \quad \text{and} \quad t \in [a, b]. \quad \ldots \quad (vi)
\]

Hence for \( t \neq x \) by Cauchy’s criterion for uniform convergence the sequence, \( \{\phi_n\} \) converges uniformly, say \( \phi_n \to \phi \) uniformly.

Therefore by (v) and (vi), we have

\[
\lim_{t \to x} \phi(t) = \lim_{n \to \infty} f'_n(x)
\]

Or,

\[
\lim_{t \to x} \{f(t) - f(x)\} / (t - x) = \lim_{n \to \infty} f'_n(x) \quad \text{by (iv)}
\]

Or,

\[
f'(x) = \lim_{n \to \infty} f'_n(x), \quad x \in [a, b].
\]

Theorem 2.7.2. Let \( \Sigma u_n \) be a series of differentiable real valued functions on \([a, b]\) and for some point \( x_0 \) in \([a, b]\) the series \( \Sigma u_n(x_0) \) is convergent. If the series \( \Sigma u'_n \) converges uniformly on \([a, b]\) then the series \( \Sigma u_n \) converges uniformly on \([a, b]\) to a function \( f \), and

\[
f'(x) = \lim_{n \to \infty} \Sigma_{m=1}^n u'_m(x), \quad x \in [a, b].
\]

In other word, \((d/dx)\Sigma u_n(x) = \Sigma (d/dx)u_n(x), \quad x \in [a, b].\)

Proof. Suppose that \( f_n(x) = u_1(x) + u_2(x) + u_3(x) + \ldots + u_n(x) \), then

\[
f'_n(x) = u'_1(x) + u'_2(x) + u'_3(x) + \ldots + u'_n(x).
\]

Clearly, \( \{f_n\} \) is a sequence of differentiable functions on \([a, b]\) and for some point \( x_0 \) in \([a, b]\) the sequence \( \{f_n(x_0)\} \) is convergent, also the sequence \( \{f'_n\} \) converges uniformly on \([a, b]\) then by theorem 7.1 the sequence \( \{f_n\} \) converges uniformly on \([a, b]\) to a function \( f \), and

\[
f'(x) = \lim_{n \to \infty} f'_n(x) \quad x \in [a, b].
\]

\[
(d/dx)\Sigma u_n(x) = \Sigma (d/dx)u_n(x), \quad x \in [a, b].
\]
2.8. Weierstrass approximation theorem:

**Theorem 2.8.1.** Let \( f \) be any continuous function in \( C[a, b] \). Then for any \( \varepsilon > 0 \) there exists a polynomial \( P(x) \) such that \( |p(x) - f(x)| < \varepsilon, \quad \forall x \in [a, b] \).

**Proof.** To prove this theorem first we shall prove it for a special case \( a = 0 \) and \( b = 1 \). Since we observe that \( x = (b-a)x' + a \) provides that a continuous mapping of \([0,1]\) onto \([a, b]\), so that we can define a function \( g(x') = f[(b-a)x' + a] \) is a continuous real function on \([0, 1] \). So if our theorem is true for the case \( a = 0 \) and \( b = 1 \), then there exists a polynomial \( p' \) on \([0, 1] \) such that

\[
|g(x') - p'(x)| < \varepsilon, \quad \forall x' \in [0, 1].
\]

In term of \( x \) this inequality reduced to \( |f(x) - p'(x)/(b-a)| < \varepsilon, \quad \forall x \in [a, b] \).

Accordingly, we may assume that \( a = 0 \) and \( b = 1 \).

Let \( f \in C[0, 1] \), define a sequence of polynomials \( \{B_n\} \) as:

\[
B_n(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f(k/n) \quad (0 < x \leq 1), \quad n \in \mathbb{N}, \quad \text{where } \binom{n}{k} = n!/(k!(n-k)!) \quad \ldots \quad (A)
\]

This polynomial \( B_n(x) \) is called \( \text{th} \) Bernstein polynomial for \( f \).

Now by Binomial theorem for any \( p, q \in \mathbb{R} \), we have

\[
\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p + q)^n, \quad n \in \mathbb{N} \quad \ldots \quad \ldots \quad \ldots \quad (i)
\]

On differentiating \( (i) \) with respect to \( p \), we have

\[
\sum_{k=0}^{n} \binom{n}{k} k p^{k-1} q^{n-k} = n(p + q)^{n-1}, \quad n \in \mathbb{N} \quad \ldots \quad \ldots \quad \ldots \quad (ii)
\]

Again differentiating \( (ii) \) with respect to \( p \), we have

\[
\sum_{k=0}^{n} \binom{n}{k} (k^2/n) p^{k-1} q^{n-k} = (n-1)p(p + q)^{n-2} + (p + q)^{n-1}, \quad n \in \mathbb{N} \quad \ldots \quad \ldots \quad \ldots \quad (iii)
\]

Now for \( x \in [0, 1] \), \( p = x \) and \( q = 1 - x \) in \( (i) \), \( (ii) \) and \( (iii) \), we have

\[
\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = 1, \quad n \in \mathbb{N} \quad \ldots \quad \ldots \quad \ldots \quad (iv)
\]

\[
\sum (k/n) \binom{n}{k} x^k (1-x)^{n-k} = x, \quad n \in \mathbb{N} \quad \ldots \quad \ldots \quad \ldots \quad (v)
\]
\[
\sum_{k=0}^{n} (k^2/n^2) \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} \right\} = \{1-(1/n)\}x^2 + (x/n), \quad n \in \mathbb{N} \quad \ldots \quad (vi)
\]

Now by equations (iv), (v) and (vi), we have
\[
\sum_{k=0}^{n} [(k^2/n^2) - (2kx/n) + x^2] \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} \right\} = x^2 - (x^2/n) - 2x^2 + x^2, \quad x \in [0,1] \text{ and } n \in \mathbb{N}
\]
\[
\Rightarrow \sum_{k=0}^{n} [(k/n) - x]^2 \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} \right\} = x(1-x)/n, \quad x \in [0,1] \text{ and } n \in \mathbb{N} \quad \ldots \quad (vii)
\]

Since we know that “if \((M_1, \rho_1)\) be compact metric space and if \(f\) is continuous function from \((M_1, \rho_1)\) to another metric space \((M_2, \rho_2)\) then \(f\) is uniformly continuous on \(M_1\).”

Hence \(f\) is uniformly continuous on \([0,1]\), then for given \(\varepsilon > 0\) \(\exists \delta > 0\) such that
\[
|f(x) - f(y)| < \varepsilon/2 \quad \text{for} \quad |x - y| < \delta, \quad \forall x, y \in [0,1].
\]

Choose positive integer \(N\) such that
\[
1/N^{1/4} < \delta \quad \ldots \quad (viii)
\]
and such that
\[
1/\sqrt{N} < \varepsilon/4||f|| \quad \ldots \quad (ix)
\]
Fix \(x \in [0,1]\), multiply (iv) by \(f(x)\) and subtract from (A), we have
\[
f(x) - B_n(x) = \sum_{k=0}^{n} x^k (1 - x)^{n-k} [f(x) - f(k/n)] = \Sigma' + \Sigma'' (\text{say}), \quad \ldots \quad (x)
\]
where \(\Sigma'\) is the sum over those values of \(k\) such that \(|(k/n) - x| < 1/n^{1/4}\) \(\ldots \quad (xi)\)
while \(\Sigma''\) is the sum over all other those values of \(k\).

Now if \(k\) does not satisfies (xi)
i.e., \(|(k/n) - x| \geq 1/n^{1/4} \Rightarrow (k - nx)^2 \geq n^2/\sqrt{n} = n^{3/2} \quad \ldots \quad (xii)\)

Hence \(|\Sigma''| = |\Sigma''| \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} [f(x) - f(k/n)] \right\} \]
\[
\leq \Sigma'' \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} [f(x)] + |f(k/n)| \right\} \]
\[
\leq 2||f|| \Sigma'' \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} \right\} \]
\[
\leq [2||f||/n^{1/3}] \Sigma''(k-nx)^2 \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} \right\} \]
\[
\leq [2||f||/n^{1/3}] nx(1-x) \quad \text{by (vii)} \leq 2||f||/\sqrt{n}, \quad x \in [0,1].
\]

Now for \(n \geq N\) by (ix) \(1/\sqrt{n} \leq 1/\sqrt{N} < \varepsilon/4||f||\), therefore by (xiii)
\[
|\Sigma''| \leq 2||f||/\sqrt{n} < 2||f|| \varepsilon/4||f|| = \varepsilon/2, \quad x \in [0,1] \quad \ldots \quad (xiii)
\]

Further for \(n \geq N\) if \(k\) satisfies (xi), then by (viii), we have
\[
|(k/n) - x| < 1/n^{1/4} \leq 1/N^{1/4} < \delta \quad \text{implies} \quad |f(x/n) - f(x)| < \varepsilon/2.
\]

Hence \(|\Sigma'| = |\Sigma'| \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} [f(x) - f(k/n)] \right\} \]
\[
\leq \Sigma'' \left\{ \sum_{k=0}^{n} x^k (1 - x)^{n-k} [f(x)] \right\} \]
\[
\leq [2||f||/n^{1/3}] nx(1-x) \quad \text{by (vii)} \leq 2||f||/\sqrt{n}, \quad x \in [0,1].
\]
\[ k \left\{ \sum_{k}^{n} x^k(1-x)^{n-k} \right\} < \frac{\varepsilon}{2} \] [by (iv)].

Hence by (x), we have \[ |f(x) - B_n(x)| \leq |\sum'_{\varepsilon} + |\sum'_{\varepsilon}^{\prime} | < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \quad x \in [0,1], \quad n \geq N. \]

Since \( x \) being arbitrary, consequently \[ |f(x) - B_n(x)| \leq |\sum'_{\varepsilon} + |\sum'_{\varepsilon}^{\prime} | < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \quad x \in [0,1], \quad n \geq N. \]

2.9. Power Series:

**Power Series.** *(Definition 2.9.1):* A series of the form \( \sum_{n=0}^{\infty} c_n z^n \) or \( \sum_{n=0}^{\infty} c_n(z-a)^n \) is called power series. Here \( c_n(n = 0, 1, 2, 3, \ldots) \), \( z \) and \( a \) are complex numbers and the constants \( c_n \) are independent from \( z \). Second form can be reduced into the first form easily by changing origin.

With every power series there associated a disk, called the *disk of convergence*, such that the series converges absolutely for every \( z \) interior to this disk and diverges for every \( z \) outside this disk. The center of the disk is at \( a \) and its radius is called *radius of convergence* of the power series. Here we shall derive some properties of functions \( f(x) \) which are represented by these power series, these functions are called *analytic functions*. The following theorem established the existence of the disk of convergence and provides us with a way of calculating its radius.

**Theorem 2.9.1:** Let \( \sum_{n=0}^{\infty} c_n(z-a)^n \) be any power series and let \( \lambda = \lim_{n \to \infty} \sup |c_n|^{1/n}, \quad R = 1/\lambda. \)

Then the series converges absolutely if \( |z-a| < R \) and diverges if \( |z-a| > R \). Further the series converges uniformly on every compact subset interior to the disk of convergence.

**Proof.** Let \( u_n = c_n(z-a)^n \) then by Cauchy’s root test, we have

\[
\lim_{n \to \infty} \sup |u_n|^{1/n} = \lim_{n \to \infty} \sup |c_n(z-a)^n|^{1/n} = |z-a|/R.
\]

Hence the series converges absolutely if \( |z-a|/R < 1 \) and diverges if \( |z-a|/R > 1 \), Or, converges absolutely if \( |z-a|/R < 1 \) and diverges if \( |z-a| > R \).

\( (|z-a| < R \quad \text{and} \quad R \quad \text{are called disk of convergence and radius of convergence respectively}) \)

For the second part we observe that if \( E \) is any compact subset in the disk of convergence, then there exists a point \( p \) in \( T \) such that \( z \in T \) implies that \( |z-a| \leq |p-a| < R \).

Therefore \( |c_n(z-a)| \leq |c_n(|p-a)| \) for each \( z \) in \( T \), therefore by Weierstrass M-Test the series converges uniformly on \( E \) in the disk of convergence.

*(Here we note that if \( \lim_{n \to \infty} |c_n/z_n| \) exists then it is equal to the radius of convergence of the series)*

**Example 2.9.1:**

(i) \( \sum_{n=0}^{\infty} n^a z^n \) (ii) \( \sum_{n=0}^{\infty} (z^n/n!) \) (iii) \( \sum_{n=0}^{\infty} (z^n/n) \) (iv) \( \sum_{n=0}^{\infty} (z^n/n^2) \)

**Solution.** Here for (i) \( R = 0 \), for (ii) \( R = \infty \), for (iii) \( R = 1 \) and the series is divergent if \( |z| = 1 \), because of \( \lim_{n \to \infty} |z| \neq 0 \), when \( |z| = 1 \), for (iv) \( R = 1 \) and the series is divergent if \( z = 1 \) and convergent if \( |z| = 1 \), for (v) \( R = 1 \) and the series is convergent for all \( z \) for which \( |z| = 1 \).

**Example 2.9.2:** Find the radius of convergence of the following power series.
Weirstrass M-test

Proof. Let $\varepsilon > 0$ be given, now for $|x| < R - \varepsilon$, we have $|c_n x^n| < |c_0 (R - \varepsilon)^n|$. Therefore by root test $\sum c_n (R - \varepsilon)^n$ converges absolutely in the interior of its interval of convergence. Now by Weirstrass M-test for series (i) converges uniformly on $[-R + \varepsilon, -R - \varepsilon]$ ($\varepsilon > 0$). The function $f(x)$ is continuous and differentiable in $(-R, R)$ and
\[
f'(x) = \sum_{n=1}^{\infty} c_n x^{n-1} \quad (|x| < R) \quad \ldots \quad (iii)
\]

\(42\)
\[
\lim_{n \to \infty} \sup (n|c_n|)^{1/n} = \lim_{n \to \infty} \sup |c_n|^{1/n}.\text{Hence the series (ii) and (iii) have same interval of convergence. Also (ii) is a power series, it converges uniformly in } [-R+\varepsilon, R-\varepsilon] \text{ for every } \varepsilon > 0.\text{ Hence (ii) holds if } |x| \leq R - \varepsilon. \text{ Now since for any } x \text{ such that } |x| < \varepsilon \text{ we can find an } \varepsilon > 0 \text{ such that } |x| < R - \varepsilon. \text{ This implies that (ii) holds for } |x| < R. \text{ The continuity of } f \text{ follows from the existence of } f'.
\]

2. 10. Abel’s and Tauber’s theorems:

**Theorem 2.10.1. (Abel’s limit theorem)** Suppose the series \( \sum c_n \) converges. Let

\[
f(x) = \sum_{n=0}^{\infty} c_n x^n \ ( -1 < x < 1). \text{ Then } \lim_{x \to 1} f(x) = \sum_{n=0}^{\infty} c_n.
\]

**Proof.** Let \( s_n = c_0 + c_1 + \ldots + c_n, s_1 = 0. \text{ Then } s_n - s_{n-1} = c_n \text{ implies that}
\]

\[
\sum_{m=0}^{n} c_n x^n = \sum_{m=0}^{n} (s_m - s_{m-1}) x^n = (s_0 - s_1) + (s_1 - s_0) x + (s_2 - s_1) x^2 + \ldots + (s_m - s_{m-1}) x^m.
\]

Since \(|x| < 1\) then on letting \( m \to \infty, \) we have

\[
f(x) = (1-x) \sum_{n=0}^{\infty} s_n x^n. \quad \ldots \quad \ldots \quad (i)
\]

Now since \( \sum c_n \) converges therefore \( \{s_n\} \) converges also say \( s = \lim_{n \to \infty} s_n, \) then for given \( \varepsilon > 0 \) there exists positive integer \( N \) such that

\[
\sup |s_n - s| < \varepsilon/2 \quad \forall n \geq N.
\]

Again since

\[
(1-x) \sum_{n=0}^{\infty} x^n = 1. \quad \sup |x| < 1 \quad \ldots \quad \ldots \quad (ii)
\]

Hence from (i) and (ii), we have

\[
|f(x) - s| = |(1-x) \sum_{n=0}^{\infty} s_n x^n - s(1-x) \sum_{n=0}^{\infty} x^n| = |(1-x) \sum_{n=0}^{\infty} (s_n - s) x^n| \leq (1-x) \sum_{n=0}^{N} |s_n - s| x^n + \varepsilon/2
\]

Now since if \( x \to 1 \) then \( |f(x) - s| < \varepsilon. \) Hence \( \lim_{x \to 1} f(x) = \sum_{n=0}^{\infty} c_n. \)

**Multiplication of power series:** If \( \sum a_n x^n, \sum b_n x^n, \) are power series then their multiplication is defined by \( (\sum a_n x^n)(\sum b_n x^n) = \sum c_n x^n \) where \( c_n = a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \ldots + a_n b_0. \) If \( \sum a_n, \sum b_n, \) and \( \sum c_n \) are converges to \( A, B \) and \( C \) respectively then \( AB = C. \)

Let \( f(x) = \sum a_n x^n, g(x) = \sum b_n x^n \) and \( h(x) = \sum c_n x^n \) for \( 0 \leq x < 1. \)

Then \( f(x) g(x) = h(x) \) for \( 0 \leq x < 1. \) Letting \( x \to 1 \) and using Abel’s theorem , we have \( f(x) \to A, g(x) \to B \) and \( h(x) \to C. \) Hence \( AB = C. \)
The converse of Abel’s theorem is not true in general, i.e. if \( f(x) = \sum c_n x^n \) \((-1 < x < 1)\) and if \( \lim_{x \to 1} f(x) \) is may exists but yet the series \( \sum c_n \) may not convergence. For example if \( c_n = (-1)^n \), then \( f(x) = 1/(1 + x) \) \((-1 < x < 1)\), \( f(x) \to \frac{1}{2} \) as \( x \to 1 \), but the series \( \sum (-1)^n \) is not convergent. By placing further restrictions on the coefficient \( c_n \), Tauber established the converse of the Abel’s theorem.

**Theorem 2.10.2: (Tauber’s theorem):** Let \( f(x) = \sum c_n x^n \) \((-1 < x < 1)\) and suppose that \( \lim_{n \to \infty} nc_n = 0 \). If \( f(x) \to s \) as \( x \to 1 \), then \( \sum c_n \) converges and has sum \( s \).

**Proof.** Let \( s_n = c_1 + c_2 + c_3 + \ldots + c_n \) and \( \{\sigma_n\} \) is the sequence of arithmetic means defined by

\[
\sigma_n = (s_1 + s_2 + s_3 + \ldots + s_n)/n
\]

or \( n\sigma_n = (s_1 + s_2 + s_3 + \ldots + s_n) = c_1 + c_1 + c_2 + c_1 + c_2 + c_3 + \ldots + c_1 + c_2 + c_3 + \ldots + c_n \)

\[
= c_1 + 2c_2 + 3c_3 + \ldots + n c_n = \sum_{k=1}^{n} kc_k
\]

and since \( nc_n \to 0 \) as \( n \to \infty \). Then \( \sigma_n \to 0 \) as \( n \to \infty \). Consider \( x_n = 1 - 1/n \) then as \( n \to \infty \), \( x_n \to 1 \), therefore \( \lim_{n \to \infty} f(x_n) = s \). Hence for given \( \varepsilon > 0 \) there exists positive integer \( N \) such that

\[
|f(x_n) - s| < \varepsilon/3, |\sigma_n - 0| < \varepsilon/3 \quad \text{and} \quad |ncn - 0| < \varepsilon/3.
\]

Now for \(-1 < x < 1\), we can write

\[
s_n - s = \sum_{k=0}^{n} c_k x^k - s = \sum_{k=0}^{n} c_k x^k - \sum_{k=0}^{\infty} c_k x^k = f(x) - s + \sum_{k=n+1}^{\infty} c_k (1 - x^k) - \sum_{k=0}^{n} c_k x^k.
\]

Now for \( x \in (0, 1) \), we have

\[
(1 - x^k) = (1 - x)(1 + x + x^2 + \ldots + x^{k-1}) \leq k(1 - x) \quad \text{for each} \ k.
\]

Hence for \( n \geq N \) and \( x \in (0, 1) \), we have

\[
|s_n - s| \leq |f(x) - s| + (1 - x) \sum_{k=0}^{n} |c_k| + \{\varepsilon/n(1 - x)\}.
\]

Taking \( x = x_n = 1 - 1/n \), we have \( |s_n - s| < (\varepsilon/3) + (\varepsilon/3) + (\varepsilon/3) = \varepsilon \). Hence the series \( \sum c_n \) converges and has sum \( s \).

### 2.11. Unit Summary.

**2.11.1.** Let \( \{f_n\} \) be a sequence of functions defined on a set \( E \) and suppose that for each \( x \in E \) the sequence of numbers \( \{f_n(x)\} \) converges to a function \( f(x) \), i.e., \( f(x) = \lim_{n \to \infty} f_n(x) \), for each \( x \in E \). Then we say that \( \{f_n\} \) converges on \( E \) and \( f \) is the limit or limit function of \( \{f_n\} \) and the convergence of \( \{f_n\} \) is called **pointwise convergence** of \( \{f_n\} \) on \( E \).

**2.11.2.** A sequence \( \{f_n\} \) of functions defined on a set \( E \) is said to be **converges uniformly** on \( E \) to a function \( f \) if for \( \varepsilon > 0 \) \( \exists \ n_0 = n_0(\varepsilon) \in \mathbb{N} \) such that \( |f_n(x) - f(x)| < \varepsilon \), \( \forall \ n \geq n_0 \), for all \( x \in E \). Clearly uniform convergence implies pointwise convergence but not conversely.

**2.11.3. Cauchy Criterion for uniform convergence for sequence:** Any sequence \( \{f_n\} \) of functions defined on a set \( E \) is converges uniformly on \( E \) to a function \( f \) iff the following condition (called the Cauchy condition) hold:

For every \( \varepsilon > 0 \), \( \exists \ p \in \mathbb{N} \) such that \( |f_n(x) - f_m(x)| < \varepsilon \quad \forall n, m \geq p \) and \( \forall x \in E \).
**Cauchy Criterion for uniform convergence for series:** A series $\Sigma f_n$ of functions defined on a set $E$ converges uniformly on $E$ to a function $f$ iff the following condition (called the Cauchy condition) hold:

For every $\varepsilon > 0$, there exists $M \in \mathbb{N}$ such that $|\Sigma f(x)| < \varepsilon$ for all $n \geq m$, $p = 1, 2, 3, \ldots$ and $\forall x \in E$.

**2.11.4. Test for uniform convergence:**

**$M_n$-test for sequence:** A sequence $\{f_n\}$ of functions defined on a set $E$ and if $\lim_{k\to\infty} f_n(x) = f(x)$ $(x \in E)$ and $M_n = \sup_{x \in E} |f_n(x) - f(x)|$. Then $f_n \to f$ uniformly on $E$ iff $M_n \to 0$ as $n \to \infty$.

**Weierstrass M-test for series:** A sequence $\{f_n\}$ of functions defined on a set $E$ and suppose that $|f_n(x)| \leq M_n$ $(\forall x \in E, n = 1, 2, 3, \ldots)$. Then series $\Sigma f_n$ converges uniformly (and absolutely) on $E$ if $\Sigma M_n$ converges.

**Abel’s test for series:** A series $\Sigma u_n(x)v_n(x)$ converges uniformly on $[a,b]$ if

(i) $\Sigma u_n(x)$ is uniformly convergent in $[a,b]$
(ii) $\{v_n(x)\}$ is monotonic for every $x$ in $[a,b]$
(iii) $\{v_n(x)\}$ is uniformly bounded in $[a,b]$.
(iv) $s_n(x) = \Sigma_{m=1}^n u_r(x)$ is uniformly bounded in $[a,b]$.

**Dirichlet’s test for series:** A series $\Sigma u_n(x)v_n(x)$ is converges uniformly on $[a,b]$ if

(iii) $\{v_n(x)\}$ is positive monotonic decreasing sequence converging uniformly to zero for all $x$ in $[a,b]$
(iv) $s_n(x) = \Sigma_{n=1}^n u_r(x)$ is uniformly bounded in $[a,b]$.

**2.11.5. Some Results:**

(1) Suppose that the sequence $\{f_n\}$ of continuous functions on a set $E$, converges uniformly to $f$ on a set $F$ then $f$ is also continuous on $E$.

(2) Suppose that $E$ is compact and $\{f_n\}$ is a sequence of continuous functions on $K$ converges pointwise to a continuous function $f$ on $E$ such that $f_n(x) \geq f_{n+1}(x)$ $\forall x \in E$, $n = 1, 2, 3, \ldots$. Then $f_n \to f$ uniformly on $K$.

(3) Let $\{f_n\}$ be a sequence of differentiable functions on $[a,b]$ and for some point $x_0$ in $[a,b]$ the sequence $\{f_n(x_0)\}$ is convergent. If sequence $\{f'_n\}$ converges uniformly on $[a,b]$ then sequence $\{f_n\}$ converges uniformly on $[a,b]$ to a function $f$, and $f'(x) = \lim_{n\to\infty} f'_n(x)$ $x \in [a,b]$.

(4) Let $f$ be any continuous function in $C[a,b]$. Then for any $\varepsilon > 0$ there exists a polynomial $P(x)$ such that $|P(x) - f(x)| < \varepsilon$, $\forall x \in [a,b]$.

**2.11.6. Power Series.** A series of the form $\Sigma c_n z^n$ or $\Sigma c_n (z-a)^n$ is called power series, where $c_n$ $(n = 0, 1, 2, 3, \ldots)$, $z$ and $a$ are complex numbers and the constants $c_n$ are independent from $z$.

(1) Let $\Sigma c_n (z-a)^n$ be any power series and let $\lambda = \lim_{n\to\infty} \sup |c_n|^{1/n}$, $R = 1/\lambda$. Then the series converges absolutely if $|z-a| < R$ and diverges if $|z-a| > R$.

(2) Suppose the series $\Sigma c_n z^n$ converges for $|x| < R$ and let $f(x) = \Sigma c_n x^n$ $(|x| < R)$. Then $\Sigma c_n z^n$ converges uniformly on $[-R + \varepsilon, -R- \varepsilon]$ ($\varepsilon > 0$). The function $f(x)$ is continuous and differentiable in $(-R, R)$ and $f'(x) = \Sigma c_n x^{n-1}$ $(|x| < R)$.

**2.11.7. Abel’s and Tauber’s theorems:**

**Abel’s limit theorem** Suppose the series $\Sigma c_n$ converges. Let $f(x) = \Sigma c_n x^n$ ($-1 < x < 1$). Then $\lim_{x\to1} f(x) = \Sigma c_n$.

**Tauber’s theorem:** Let $f(x) = \Sigma c_n x^n$ ($-1 < x < 1$) and suppose that $\lim_{n\to\infty} n c_n = 0$. If $f(x) \to s$ as $x \to 1$, then $\Sigma c_n$ converges and has sum $s$. 
2.12. Assignments/Check your Progress.

1. Prove that every uniformly convergent sequence of bounded functions is uniformly bounded.
2. If \{f_n\} and \{g_n\} converge uniformly on a set E, prove that \{f_n + g_n\} converges uniformly on E. If, in addition, \{f_n\} and \{g_n\} are sequences of bounded functions prove that \{f_n g_n\} converges uniformly on E.
3. Construct sequences \{f_n\} and \{g_n\} which converge uniformly on some set E, but such that \{f_n g_n\} does not converge uniformly on E.
4. Prove that the series \[ \sum_{n=1}^{\infty} (-1)^n \frac{x^2 + n}{n^2} \] converges uniformly in every bounded interval, but does not converge absolutely for any value of x.
5. For \( n = 1, 2, 3, \ldots \), let \( f_n(x) = \frac{x}{1 + nx^2} \). Show that \{f_n\} converges uniformly to a function f, and that the equation \( f'(x) = \lim_{n \to \infty} f'_n(x) \) is correct if \( x \neq 0 \) but false if \( x = 0 \).
6. Let \{f_n\} be a sequence of continuous functions which converges uniformly to a function \( f \) on a set E. Prove that \( f(x) = \lim_{n \to \infty} f_n(x_n) \) for every sequence of points \( x_n \in E \) such that \( x_n \to x \) and \( x \in E \). Is the converse of this true?
7. Suppose \{f_n\}, \{g_n\} are defined on E, and
   (a) \( \sum f_n \) has uniformly bounded partial sums (b) \( g_n \to 0 \) uniformly on E.
   (c) \( g_1(x) \geq g_2(x) \geq g_3(x) \geq \ldots \) for all \( x \in E \). Prove that \( \sum f_n g_n \) converges uniformly on E.
8. Let \{f_n\} be a uniformly bounded sequence of functions which are Reimann Integrable on \([a, b]\) and \( F_n(x) = \int_a^x f_n(t) dt \) (\( a \leq x \leq b \)). Prove that there exists a subsequence \{F_{n_k}\} which converges uniformly on \([a, b]\).
9. If \( f \) is continuous on \([0, 1]\) and if \( \int_0^1 f(x)x^n dx = 0 \), \( n = 1, 2, 3, \ldots \). Prove that \( f(x) = 0 \) on \([0, 1]\).
10. If \( f_n \to f \) uniformly on \( S \) and suppose there is a constant \( M > 0 \) such that \( |f_n(x)| \leq M \) for all \( x \) in \( S \) and all \( n \). Let \( g \) be continuous on the closure of the disk \( B(0; M) \) and define \( h_n(x) = g[f_n(x)] \), \( h(x) = g[f(x)] \), if \( x \in S \). Prove that \( h_n \to h \) uniformly on \( S \).
11. (a) Let \( f_n(x) = 1/(nx + 1) \) if \( 0 < x < 1, n = 1, 2, \ldots \). Prove that \{f_n\} converges pointwise but not uniformly on \((0, 1)\).
    (b) Let \( g_n(x) = x/(nx + 1) \) if \( 0 < x < 1, n = 1, 2, \ldots \). Prove that \( g_n \to 0 \) uniformly on \((0, 1)\).
12. Let \( f_n(x) = x^n \), the sequence \{f_n\} converges pointwise but not uniformly on \([0, 1]\). Let \( g \) be continuous on \((0, 1)\) with \( g(1) = 0 \). Prove that the sequence \{g(x)x^n\} converges uniformly on \([0, 1]\).
13. Assume that \( f_n \to f \) uniformly on \( S \), and that each \( f_n \) is continuous on \( S \). If \( x \in S \), let \( \{x_n\} \) be a sequence of points in \( S \) such that \( x_n \to x \). Prove that \( f_n(x_n) \to f(x) \).

14. Let \( f_n(x) = n^2x(1 - x^2)^n \) for \( x \) real and \( n \geq 1 \). Prove that \( \{f_n\} \) converges pointwise on \([0, 1]\) for every real \( c \). Determine those \( c \) for which the convergence is uniform on \([0, 1]\) and those for which term-by-term integration on \([0, 1]\) leads to a correct result.

15. Prove that \( \sum x^n (1 - x) \) converges pointwise but not uniformly on \([0, 1]\), whereas \( \sum (-1)^n x^n (1 - x) \) converges uniformly on \([0, 1]\).

16. Let \( g_{n+1}(x) \leq g_n(x) \) for each \( x \) in \( T \) and each \( n = 1, 2, \ldots \), and suppose \( g_n \to 0 \) uniformly on \( T \). Prove that \( \sum (-1)^{n+1} g_n(x) \) converges uniformly on \( T \).

17. Let \( f_n(x) = 1/(1 + n^2x^2) \) if \( 0 \leq x \leq 1 \), \( n = 1, 2, \ldots \). Prove that \( \{f_n\} \) converges pointwise but not uniformly on \([0, 1]\). Is term-by-term integration permissible?

18. Prove that \( \sum_{n=1}^{\infty} \frac{x}{n^a(1 + nx^2)} \) converges uniformly on every finite interval in \( \mathbb{R} \) if \( a > \frac{1}{2} \). Is the convergence uniform on \( \mathbb{R} \)?

19. Prove that the series \( \sum_{n=1}^{\infty} \frac{(-1)^n}{\sqrt{n}} \sin(1 + (x/n)) \) converges uniformly on every compact subset of \( \mathbb{R} \).

20. Prove that the series \( \sum_{n=0}^{\infty} \left( \frac{x^{2n+1}}{2n+1} - \frac{x^{n+1}}{2n+2} \right) \) converges pointwise but not uniformly on \([0, 1]\).

21. Prove that \( \sum_{n=1}^{\infty} a_n \sin nx \) and \( \sum_{n=1}^{\infty} a_n \cos nx \) are uniformly convergent on \( \mathbb{R} \) if \( \sum_{n=1}^{\infty} |a_n| \) converges.

22. Let \( \{a_n\} \) be a decreasing sequence of positive terms. Prove that the series \( \sum a_n \sin x \) converges uniformly on \( \mathbb{R} \) if, and only if, \( na_n \to 0 \) as \( n \to \infty \).

2.13. Points for discussion/Clarification.
2.14. Suggested Study material

5. H.K. Pathak - Real Analysis
6. Lecture Notes on Real analysis Richard F. Bass

=*=*=*=*=*=*=*=*=*=*=*=****
Functions of Several Variables,

Introduction: As we studied earlier the partial derivatives of the functions from $\mathbb{R}^n$ to $\mathbb{R}$ In this unit our main aim to introduced the theory of derivatives of functions from $\mathbb{R}^n$ into $\mathbb{R}^m$. The partial derivative is some what unsatisfactory generalization of the usual derivative because the existence of all the partial derivatives at a particular point does not necessarily imply the continuity of the function at that point, because of in the partial derivatives we treat a function of several variables as a function of one variable at a time. The partial derivatives describes the rate of change of a function in the direction of each co-ordinate axis. There is a slight generalization, called the directional derivative, which studies the rate of change of a function in an arbitrary direction. When we have a system of several equations involving several variables and we want to solve these equations for some of these variables in terms of remaining variables, the implicit function theorem provided a description of some conditions and conclusions about the solution of these equations.

Objectives: In this unit we shall discus various properties of sets of vectors in Euclidean n-space $\mathbb{R}^n$. The algebraic facts presented here extend without change to finite-dimensional vector spaces over any field of scalars. However, for our purposes it is quite sufficient to stay within the familiar framework provided by the Euclidean space. Our study is focused on the concepts of functions of several variables, linear transformations, derivatives and partial derivatives in an open subset of $\mathbb{R}^n$, Taylor’s theorem, Inverse function theorem, Implicit function theorem, Differentiation of integrals etc.

3.1. Functions of Several Variables.

3.1.1. Definitions. As you are familiar to the concept of vector spaces and their properties in earlier classes for the convenience here we recall some definitions and preliminaries of the same in the frame work in Euclidean n-space $\mathbb{R}^n$.

(a) A nonempty set $X \subseteq \mathbb{R}^n$ is said to be a vector space if $x + y \in X$ and $cx \in X$ for all $x \in X$, $y \in X$, and for all scalars $c$, where $x = (a_1, a_2, ..., a_n)$, $y = (b_1, b_2, ..., b_n) \in \mathbb{R}^n$ and $a_1, a_2, ... , a_n$; $b_1, b_2, ... , b_n$ are scalars etc.

(b) If $x_1, ... , x_k \in \mathbb{R}^n$ and $c_1, ..., c_k$ are scalars, then the vector $c_1x_1 + ... + c_kx_k$ is called a linear combination of $x_1, ..., x_k$.

(c) If $S = \{ x_1, ... , x_k \} \subseteq \mathbb{R}^n$ and if $E$ is the set of all linear combinations of elements of $S$, i.e. $E = \{ c_1x_1 + ... + c_kx_k : x_i \in S, c_i \text{ are scalars} \}$ then we that $S$ spans $E$, or that $E$ is the span of $S$ and we write $E = L(S)$. Note that every $L(S)$ is a vector space.

(d) A set $S = \{ x_1, ... , x_k \} \subseteq \mathbb{R}^n$ is said to be linearly independent if the relation $c_1x_1 + ... + c_kx_k = 0$ implies that $c_1 = ... = c_k = 0$. Otherwise $S$ is said to be linearly dependent. Note that no independent set contains the null vector.

(e) If $S = \{ x_1, ... , x_k \}$ is linearly independent subset if a vector space $X$ but any subset $S_1 = \{ x_1, ... , x_k, x_{k+1} \}$ which is not linearly independent then $k$ is called the dimension of $X$, and we write, $\dim X = k$ and $X$ is called finite dimensional. Note that the set consisting of 0 alone is a vector space its dimension is 0.

(f) Let $B$ is a linearly independent subset of a vector space $X$ such that $L(B) = X$, then $B$ is called a basis of $X$. Here we note that if $B = \{ x_1, ... , x_r \}$ is a basis of $X$, then every $x \in X$ has a unique representation of the form $x = \Sigma c_jx_j$. Such a representation exists since $L(B) = X$. 
and it is unique since \( B \) is linearly independent. The numbers \( c_1, \ldots, c_r \) are called the coordinates of \( x \) with respect to the basis \( B \) and we write \( [x]_B = (c_1, \ldots, c_r) \). Note that each basis of \( X \) possess equal number of elements.

The most familiar example of a basis is the set \( \{e_1, \ldots, e_n\} \), where \( e_j \) is the vector in \( \mathbb{R}^n \) whose \( j \)th coordinate is 1 and whose other coordinates are all 0, i.e., \( e_j = (0, 0, \ldots, 1, \ldots, 0) \). If \( x \in \mathbb{R}^n, \ x = (a_1, a_2, \ldots, a_n) \), then \( x = \sum a_j e_j \) and the set \( \{e_1, \ldots, e_n\} \) is called the standard basis of \( \mathbb{R}^n \).

### 3.1.2. Theorem

If a vector space \( X \) is spanned by a set of \( r \) vectors, then \( \text{dim} X \leq r \), where \( r \) is a positive integer.

**Proof.** Suppose on the contrary that \( X \) is a vector space which contains a linearly independent set \( Q = \{y_1, \ldots, y_{i+1}\} \) such that \( Q = L(S_0) \), where \( S_0 = \{x_1, x_2, \ldots, x_r\} \).

Suppose for \( 0 \leq i < r \), \( S_i \) be a set such that

(i) \( L(S_i) = X \)  and  (ii) \( S_i = \{y_1, \ldots, y_{i+1}, x_1, x_2, \ldots, x_{i+1}\} \).

Since \( S_i \) spans \( X \) so \( y_{i+1} \in L(S_i) \) , hence there are scalars \( a_1, \ldots, a_{i+1}, b_1, \ldots, b_{r-i} \), with \( a_{i+1} = 1 \) such that \( \sum a_j y_j + \sum b_k x_k = 0 \), where \( j = 1 \) to \( i+1 \) and \( k = 1 \) to \( r-i \).

If all \( b_k = 0 \) then by the linear independence of \( Q \) we have all \( a_j = 0 \), a contradiction. It follows that some \( x_k \in S_i \) is a linear combination of the other members of \( T_i = S_i \cup \{y_{i+1}\} \). On removing this \( x_k \) from \( T_i \) and call the remaining set \( S_{i+1} \). Then clearly \( L(S_{i+1}) = L(T_i) = X \), so that \( S_{i+1} \) has the properties postulated for \( S_i \) with \( i+1 \) in place of \( i \).

Starting with \( S_0 \), we thus construct sets \( S_1, \ldots, S_r \). Clearly the last of these consists of \( y_1, \ldots, y_r \) such that \( L(S_r) = X \). But \( Q \) is linearly independent, hence \( y_{r+1} \) is not in the span of \( S_r \) this contradicts that \( L(S_r) = X \). Hence the theorem is established.

### 3.1.3. Corollary

\( \text{dim} \mathbb{R}^n = n \).

**Proof.** Since \( \{e_1, \ldots, e_n\} \) spans \( \mathbb{R}^n \), the theorem shows that \( \text{dim} \mathbb{R}^n \leq n \). Since \( \{e_1, \ldots, e_n\} \) is independent implies that \( \text{dim} \mathbb{R}^n \geq n \). Hence \( \text{dim} \mathbb{R}^n = n \).

### 3.1.4. Theorem

Suppose \( X \) is a vector space and \( \text{dim} X = n \).

(a) A set \( E = \{x_1, \ldots, x_n\} \) in \( X \) spans \( X \) if and only if \( E \) is linearly independent.

(b) \( X \) has a basis and every basis consists of \( n \) vectors.

(c) If \( 1 \leq r \leq n \) and \( \{y_1, \ldots, y_r\} \) is an independent set in \( X \), then \( X \) has basis containing \( \{y_1, \ldots, y_r\} \).

**Proof.** (a) Suppose \( E = \{x_1, \ldots, x_n\} \) be a subset of \( X \). Since \( \text{dim} X = n \) then the set containing \( n + 1 \) or more vectors is linearly dependent. Therefore the set \( \{x_1, \ldots, x_n, y\} \) is linearly dependent for every \( y \in X \). Now if \( E \) is linearly independent, we have \( y \in L(E) \). Hence \( L(E) = X \) i.e., \( E \) span \( X \).

Conversely, if \( E \) is linearly dependent, one of its members can be removed without changing the span of \( E \). Hence \( E \) cannot span \( X \) (by Theorem 3.1.1). This proves (a).

(b) Since \( \text{dim} X = n \), then by the definition of basis and part (a) and if \( X \) contains an linearly independent set of \( n \) vectors it follows that every such set is a basis of \( X \).
(c) Let \( \{ x_1, \ldots, x_n \} \) be a basis of \( X \). Consider a set \( S = \{ y_1, \ldots, y_r, x_1, \ldots, x_n \} \) such that (i) \( L(S) = X \) and (ii) is linearly dependent, because of it contains more than \( n \) vectors. Now for \( 1 \leq i \leq n, x_i \in X = L(S) \) is a linear combination of the other members of \( S \). On removing this \( x_i \) from \( S \), the remaining set still spans \( X \). This process can be repeated \( r \) times we have a linearly independent set of \( X \) which span \( X \) and so this set will be the basis of \( X \) which contains \( \{ y_1, \ldots, y_r \} \).

3.2. Linear Transformations

3.2.1. Definitions. A mapping \( A \) of a vector space \( X \) into a vector space \( Y \) is said to be a linear transformation if
\[
A(x_1 + x_2) = Ax_1 + Ax_2 \quad \text{and} \quad A(cx) = cA(x)
\]
for all \( x, x_1, x_2 \in X \) and all scalars \( c \).

Note
(i) We often write \( Ax \) instead of \( A(x) \) if \( A \) is linear.
(ii) \( A0 = 0 \) if \( A \) is linear.
(iii) A linear transformation \( A \) of \( X \) into \( Y \) is completely determined by its action on any basis. For if \( \{ x_1, \ldots, x_n \} \) is a basis of \( X \) then every \( x \in X \) has a unique representation of the form
\[
x = \Sigma c_i x_i, \quad \text{where} \quad i = 1 \to n.
\]
Now by the linearity of \( A \), we can compute \( Ax \) from the vectors \( Ax_1, \ldots, Ax_n \) and the coordinates \( c_1, \ldots, c_n \) by the formula
\[
Ax = \Sigma c_i Ax_i, \quad \text{where} \quad i = 1 \to n.
\]
(iv) Linear transformations of \( X \) into \( X \) are often called linear operators on \( X \).
(v) If \( A \) is a linear operator on \( X \) which is one-to-one and onto, we say that \( A \) is invertible. In this case we define an operator \( A^{-1} \) on \( X \) by setting that \( A^{-1}(Ax) = x \) for all \( x \in X \). It is trivial to verify that we then also have \( A(A^{-1}x) = x \), for all \( x \in X \), and that \( A^{-1} \) is linear.
(vi) If \( A \) is a linear operator on a finite dimensional vector space \( X \) then \( a \) is one-one if and only if \( A \) is onto. In other words, a linear operator on a finite dimensional vector space is one-one or onto then \( a \) is invertible. We prove this fact as follows:

3.2.2. Theorem. A linear operator \( A \) on a finite-dimensional vector space \( X \) is one-to-one if and only if the range of \( A \) is all of \( X \) \([i.e., A(X) = X]\) or \( A \) is onto.

Proof. Let \( B = \{ x_1, \ldots, x_n \} \) be a basis of \( X \) then \( Q = \{ Ax_1, \ldots, Ax_n \} \). Let \( R(A) \) is the range of \( A \) and suppose that \( y \in R(A) \) then \( y = Ax \) for some \( x \in X \). Since \( L(B) = X \) therefore there exist scalars \( c_1, c_2, c_3, \ldots, c_n \) such that \( x = \Sigma c_i x_i \), where \( i = 1 \to n \). Therefore by the linearity of \( A \), \( y = Ax = A(\Sigma c_i x_i) = \Sigma c_i A(x_i) \), where \( i = 1 \to n \). This shows that \( y \in L(Q) \) implies that \( L(Q) = R(A) \). Now we know that \( L(Q) = R(A) = X \) if and only if \( Q \) is linearly independent. We have to prove that this happens if and only if \( A \) is one-to-one.

Suppose \( A \) is one-to-one and \( \Sigma c_i Ax_i = 0 \) then by linearity of \( A \), \( A(\Sigma c_i x_i) = 0 = A0 \) implies that \( \Sigma c_i x_i = 0 \) (since \( A \) is one-one) implies that \( c_1 = \ldots = c_n = 0 \) (since \( B \) is linearly independent). Hence \( Q \) is linearly independent.

Conversely, suppose \( Q \) is independent. Let \( Ax = 0 \) \( \iff \) \( A(\Sigma c_i x_i) = 0 \) \( \iff \) \( \Sigma c_i Ax_i = 0 \) (by linearity of \( A \)) \( \iff \) \( c_1 = \ldots = c_n = 0 \) (since \( Q \) is linearly independent) \( \iff \) \( x = 0 \).
Then for \( x, y \in X \), if \( Ax = Ay \implies A(x - y) = Ax - Ay = 0 \), so that \( x = y \), and this says that \( A \) is one-to-one.

Hence we conclude that \( Q \) is linearly independent if and only if \( A \) is one-to-one. Consequently \( R(A) = X \), i.e. \( A \) is onto.

### 3.2.3 Definitions and Notations.

(a) Let \( X \) and \( Y \) be vector spaces then \( L(X, Y) \) denote the set of all linear transformations of \( X \) into \( Y \). The set of all linear transformations of \( X \) into itself is denoted by \( L(X, X) \), instead of \( L(X, X) \) we simply write \( L(X) \). If \( A_1, A_2 \in L(X, Y) \) and if \( c_1, c_2 \) are scalars, define \( (c_1A_1 + c_2A_2)x = c_1A_1x + c_2A_2x \), \( x \in X \).

Clearly, \( c_1A_1 + c_2A_2 \in L(Y, Y) \).

(b) If \( X, Y, Z \) are vector spaces, and if \( A \in L(X, Y) \) and \( B \in L(Y, Z) \) then their product \( BA \) to be the composition of \( A \) and \( B \) is defined by \( (BA)x = B(Ax), \ x \in X \).

Clearly \( BA \in L(X, Z) \). Here we note that \( BA \) need not be the same as \( AB \) even if \( X = Y = Z \).

(c) For \( A \in L(R^n, R^n) \), the norm \( ||A|| \) of \( A \) is defined to be the sup of all numbers \( |Ax| \), where \( x \) ranges over all vectors in \( R^n \) with \( |x| \leq 1 \), i.e., \( ||A|| = \sup \{|Ax| : x \in R^n, |x| \leq 1 \} \).

(d) Here we observe that the inequality \( |Ax| \leq ||A|||x| \), holds for all \( x \in R^n \). Since if \( x \in R^n \) with \( |x| \leq 1 \), then \( ||A|| = \sup \{|Ax| : x \in R^n, |x| \leq 1 \} \).

Now we consider \( x = x/|x| \) for any \( x \in R^n \). Then \( |x| = 1 \) and \( A \) is linear therefore

\[
||A|| = \sup \{|A(x/|x|) : x \in R^n\} = \sup \{1/|x||Ax| : x \in R^n\},
\]

which implies that \( \{1/|x||Ax| \leq ||A|| \} or \{Ax \leq ||A|||x| \}, x \in R^n \).

(e) If \( \lambda \) is such that \( |Ax| \leq \lambda |x| \) for all \( x \in R^n \), then \( ||A|| \leq \lambda \). Since from the given inequality \( |Ax|/|x| \leq \lambda \) or \( A(x/|x|) \leq \lambda \) or \( |Ax| \leq \lambda \) for \( x = x/|x| \in R^n \) with \( |x| = 1 \).

It follows that \( ||A|| = \sup \{|Ax| : x \in R^n, |x| = 1\} \leq \lambda \).

### 3.2.4. Theorem.

(a) If \( A \in L(R^n, R^m) \), then \( ||A|| < \infty \) and \( A \) is a uniformly continuous mapping of \( R^n \) into \( R^m \).

(b) If \( A, B \in L(R^n, R^m) \), and \( c \) is a scalar, then \( ||A+B|| \leq ||A|| + ||B|| \), \( ||cA|| = |c||||A|| \). With the distance between \( A \) and \( B \) defined as \( ||A - B|| \), \( L(R^n, R^m) \) is metric space.

(c) If \( A \in L(R^n, R^m) \) and \( B \in L(R^m, R^k) \) then \( ||BA|| \leq ||B||||A|| \).

**Proof.** (a) Let \( E = \{e_1, ... , e_n\} \) be the standard basis of \( R^n \) and suppose that \( x \in R^n \) with \( |x| \leq 1 \).

Then by the definition of the basis there exists scalars \( c_1, c_2, c_3, ... , c_n \) such that \( x = \Sigma c_i e_i \) (\( i = 1, ... , n \)), so that \( ||\Sigma c_i e_i|| = |x| \leq 1 \) implies that \( |c_i| \leq 1 \) for \( i = 1, ... , n \). Then \( |Ax| = |A(\Sigma c_i e_i)| = |\Sigma c_i A e_i| \) (by linearity of \( A \)) \( \leq \Sigma |c_i||A e_i| \leq \Sigma |A e_i| \).

Therefore \( ||A|| \leq \Sigma |A e_i| < \infty \). Since \( |Ax - Ay| \leq ||A|| |x - y| \) \( \forall x, y \in R^n \), therefore for given \( \varepsilon > 0 \) there exists \( \delta = \varepsilon / ||A|| > 0 \) such that \( |Ax - Ay| < \varepsilon \), \( |x - y| < \delta \), proves the uniform continuity of \( A \).

(b) We have \( ||A + B|| x = ||A + Bx|| \leq ||A|| + ||B|| \) \( \forall x \in R^n \),

\[
(A + B)(x/|x|) \leq ||A|| + ||B||, \ \forall x \in R^n.
\]

\[
\sup \{|(A + B)(y)| : y \in R^n, |y| = 1\} \leq ||A|| + ||B||,
\]

Similarly if \( c \) is any scalar \( ||cA|| = \sup \{|cAx| : x \in R^n, |x| \leq 1\} = |c||\sup \{|Ax| : x \in R^n, |x| \leq 1\}| = |c||||A|| \).

Further \( \forall A, B, C \in L(R^n, R^m) \) let \( d(A, B) = ||A - B|| \), then we observe that
(i) \( d(A, B) = ||A - B|| \geq 0 \)
(ii) If \( d(A, B) = ||A - B|| = 0 \leftrightarrow A = B \).
(iii) \( d(A, B) = ||A - B|| = ||-1(B - A)|| = ||B - A|| = d(B, A) \).
(iv) \( d(A, B) = ||A - B|| = ||A - C + C - B|| \leq ||A - C|| + ||C - B|| = d(A, C) + d(C, B) \)

Hence we conclude that \((L(R^n, R^m), d)\) is a metric space.

(c) Since \( \forall A, B \in L(R^n, R^m) \), we have

\[ ||(BA)x|| = ||B(Ax)|| \leq ||B|| \cdot ||Ax|| \leq ||B|| \cdot ||A|| \cdot ||x|| \forall x \in R^n \]

\[ \Rightarrow \sup \{||(BA)x|| : x \in R^n, ||x|| \leq 1\} \leq ||B|| ||A|| \]

\[ \Rightarrow ||BA|| \leq ||B|| ||A||. \]

Since \( L(R^n, R^m) \) is a metric space therefore the concepts of open set, continuity, etc., make sense for these spaces. Our next theorem utilizes these concepts.

3.2.5. Theorem. Let \( \Omega \) be the set of all invertible linear operators on \( R^n \).

(a) If \( A \in \Omega, B \in L(R^n) \) and \( ||B - A|| \cdot ||A^{-1}|| < 1 \), then \( B \in \Omega \).

(b) \( \Omega \) is an open subset of \( L(R^n) \) and the mapping \( A \rightarrow A^{-1} \) is continuous on \( \Omega \). (This mapping is also obviously a 1-1 mapping of \( \Omega \) onto \( \Omega \), which is its own inverse)

Proof. (a) Let \( ||A^{-1}|| = 1/\alpha \) and \( ||B - A|| = \beta \), then \( ||B - A|| \cdot ||A^{-1}|| < 1 \) implies that \( \beta/\alpha < 1 \) or \( \beta < \alpha \). Now for every \( x \in R^n \), we have

\[ \alpha |x| = \alpha |A^{-1}Ax| \leq \alpha ||A^{-1}|| |Ax| = |Ax| \leq |(A - B)x| + |Bx| \leq \beta |x| + |Bx|, \]

so that \( (\alpha - \beta) |x| \leq |Bx| (x \in R^n) \). Since \( \alpha - \beta > 0 \) shows that \( Bx \neq 0 \) if \( x \neq 0 \).

Therefore for all \( x, y \in R^n \), suppose that \( Bx = By \Rightarrow Bx - By = 0 \Rightarrow B(x - y) = 0 \)

\[ \Rightarrow x - y = 0 \text{ or } x = y. \] Hence \( B \) is 1-1 and \( 0 \) it is onto also, consequently \( B \in \Omega \).

(b) As in part (a) if \( A \in \Omega \) then \( \alpha = 1/||A^{-1}|| > 0 \) is such that \( ||B - A|| < \alpha \Rightarrow B \in \Omega \). This shows that \( \Omega \) is open. If we replace \( x \) by \( B^{-1}y \) in \( (\alpha - \beta) |x| \leq |Bx| (x \in R^n) \), we have \( (\alpha - \beta) ||B^{-1}y|| \leq ||B|| ||y|| = ||y|| \).

\[ \Rightarrow \sup \{|B^{-1}(y/|y|)| : y/|y| \in R^n, |y/|y|| = 1\} \leq (\alpha - \beta)^{-1}. \]

Furthemore,

\[ ||f(B) - f(A)|| = ||B^{-1} - A^{-1}|| = ||B^{-1}(A - B)A^{-1}|| \leq ||B^{-1}|| ||A - B|| ||A^{-1}|| \leq \beta/\alpha(\alpha - \beta). \]

Since \( \beta \rightarrow 0 \) as \( B \rightarrow A \), proves the continuity of \( f: \Omega \rightarrow \Omega \) by \( f(A) = A^{-1} \).

3.2.6. Matrices. Suppose \( B = \{x_1, ..., x_n\} \) and \( B' = \{y_1, ..., y_m\} \) are bases of vector space \( X \) and \( Y \), respectively. Then every \( A \in L(X, Y) \) determines a set of numbers \( a_{ij} \) such that

\[ Ax_j = \Sigma a_{ij}y_j \quad (1 \leq i \leq m) \quad (1 \leq j \leq n) \quad \ldots \quad \ldots \quad (i) \]

We observe that these numbers in a rectangular array of \( m \times n \) rows and \( n \) columns, called \( m \times n \) matrix:

\[ [A] = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} = [a_{ij}]_{m \times n} \]
Here we observe that (i) expresses the difference where the "remainder" is small, in the sense that it tends to zero as \( h \to 0 \). Thus we can regard the derivative of \( f \) at \( x \) as the sum of a linear function that takes \( h \) to \( f(x)h \) plus a small remainder. Therefore we can regard the derivative of \( f \) at \( x \), not as a real number but as the linear operator on \( \mathbb{R}^1 \) that takes \( h \) to \( f(x)h \)
We observe that every real number \( \alpha \) gives rise to a linear operator on \( \mathbb{R}^1 \); the operator in question is simply multiplication by \( \alpha \). For example define a function \( T_\alpha : \mathbb{R}^1 \to \mathbb{R}^1 \) by \( T_\alpha(x) = \alpha x \ \forall \ x \in \mathbb{R}^1 \), clearly \( T_\alpha \) is a linear operator. Conversely we see that every linear function that carries \( \mathbb{R}^1 \) to \( \mathbb{R}^1 \) is multiplication by some real number. Thus there is a 1-1 correspondence between \( \mathbb{R}^1 \) and \( L(\mathbb{R}^1) \) which motivates the preceding statement.

Next we consider a function \( f \) from \( (a, b) \subset \mathbb{R}^1 \) into \( \mathbb{R}^m \). Here if \( f'(x) \) is exists then we define \( f'(x) \) to be a vector \( y \in \mathbb{R}^m \) for which

\[
\lim_{h \to 0} \frac{|f(x + h) - f(x) - f'(x)h|}{|h|} = 0.
\]

It can be rewrite as

\[
f(x + h) - f(x) = hy + r(h)/h, \quad r(h)/h \to 0 \quad \text{as} \quad h \to 0.
\]

Here we observe that the main term on the right side is again a linear function of \( h \).

Thus we see that every \( y \in \mathbb{R}^m \) induces a linear transformation of \( \mathbb{R}^1 \) into \( \mathbb{R}^m \) by associating to each \( h \in \mathbb{R}^1 \) the vector \( hy \in \mathbb{R}^m \).

This identification of \( \mathbb{R}^m \) with \( L(\mathbb{R}^1, \mathbb{R}^m) \) allows us to regard \( f'(x) \) as a member of \( L(\mathbb{R}^1, \mathbb{R}^m) \).

Hence if \( f \) is a differentiable mapping of \( (a, b) \subset \mathbb{R}^1 \) into \( \mathbb{R}^m \) and if \( x \in (a, b) \) then \( f'(x) \) is the linear transformation of \( \mathbb{R}^1 \) into \( \mathbb{R}^m \) satisfies

\[
\lim_{h \to 0} \frac{|f(x + h) - f(x) - f'(x)h|}{|h|} = 0
\]

or equivalently

\[
\lim_{h \to 0} \frac{|f(x + h) - f(x) - f'(x)h|}{|h|} = 0
\]

From the above discussion we are now ready to define the differentiability of a function \( f \) for the case \( n > 1 \).

**3.3.2. Definition.** Suppose \( E \) is an open set in \( \mathbb{R}^n \), \( f \) maps \( E \) into \( \mathbb{R}^m \) and \( x \in E \). If there exists a linear transformation \( A \) of \( \mathbb{R}^n \) into \( \mathbb{R}^m \) such that

\[
\lim_{h \to 0} \frac{|f(x + h) - f(x) - Ah|}{|h|} = 0
\]

then \( f \) is said to be differentiable at \( x \), and we write \( f'(x) = A \). If \( f \) is differentiate at every \( x \in E \), we say that \( f \) is differentiable in \( E \). Here we note that, of course, \( h \in \mathbb{R}^n \). Also since \( E \) is open then if \( |h| \) is small enough, then \( x + h \in E \). Thus \( f(x + h) \) is defined, \( f(x + h) \in \mathbb{R}^m \), and since \( A \in L(\mathbb{R}^n, \mathbb{R}^m) \), \( Ah \in \mathbb{R}^m \). Therefore

\[
f(x + h) - f(x) - Ah \in \mathbb{R}^m.
\]

The norm in the numerator of (i) is that of \( \mathbb{R}^m \) and in the denominator we have the \( \mathbb{R}^n \)-norm of \( h \).

Now there arise a question that \( f'(x) = A \) defined in term of linear transformation is unique or not. The following result deals the answer of this question.

**3.3.3. Theorem.** Suppose \( E \) is an open set in \( \mathbb{R}^n \), \( f \) maps \( E \) into \( \mathbb{R}^m \) and \( x \in E \). If there exists a linear transformation \( A \) of \( \mathbb{R}^n \) into \( \mathbb{R}^m \) such that

\[
\lim_{h \to 0} |f(x + h) - f(x) - Ah| = 0
\]
Theorem (chain rule). For all \( x \) we have defined derivatives of functions carrying derivative \( (d) \) which \( (b) \) transformation applied to left hand side of (ii) is approximately equal to By saying that for fixed \( m \), yields that the above equality is independent of \( t \). Hence, \( Bh = 0 \) for every \( h \in \mathbb{R}^n \). Consequently \( B = 0 \) or \( A_1 = A_2 \).

3.3.4 Some Remarks
(a) The relation \( \lim_{h \to 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = 0 \) can be rewritten in the form \( f(x+h) - f(x) = f'(x)h + r(h) \) where the remainder \( r(h) \) satisfies \( \lim_{h \to 0} |r(h)|/|h| = 0 \).

(b) Suppose \( f \) and \( E \) are as in Definition 3.3.3, and \( f \) is differentiable in \( E \). Then we can interpret \( f'(x) \) for every \( x \in E \), a function namely, a linear transformation of \( \mathbb{R}^n \) into \( \mathbb{R}^m \). But \( f' \) is also a function: \( f' \) maps \( E \) into \( L(\mathbb{R}^n, \mathbb{R}^m) \).

(c) Since as differentiability implies continuity it follow that that \( f \) is continuous at any point at which \( f \) is differentiable.

(d) The derivative defined by (ii) or (iii) is often called the \textit{differential} of \( f \) at \( x \), or the \textit{total derivative} of \( f \) at \( x \).

(e) We have defined derivatives of functions carrying \( \mathbb{R}^n \) to \( \mathbb{R}^m \) to be linear transformations of \( \mathbb{R}^n \) into \( \mathbb{R}^m \). What is the derivative of such a linear transformation? For the answer of this question consider \( A \in L(\mathbb{R}^n, \mathbb{R}^m) \) and if \( x \in \mathbb{R}^n \), then \( A'(x) = A \). Since \( x \) appears on the left side of \( A'(x) = A \), but not on the right. However both \( A'(x) \) and \( A \) are members of \( L(\mathbb{R}^n, \mathbb{R}^m) \), whereas \( Ax \in \mathbb{R}^m \). Then by linearity of \( A \), we have \( A(x+h) - Ax = Ah \).

Therefore for \( f(x) = Ax \) from (ii), we have \( \lim_{h \to 0} \frac{|f(x+h) - f(x) - Ah|}{|h|} = \lim_{h \to 0} \frac{|Ah - Ah|}{|h|} = 0 \) for all \( h \in \mathbb{R}^n \), yields that \( f'(x) = A'x = A \).

Now we extend the chain rule in present situation.

3.3.5 Theorem (chain rule). Suppose \( E \) is an open set in \( \mathbb{R}^n \), \( f \) maps \( E \) into \( \mathbb{R}^m \), \( f \) is differentiable at \( x_0 \in E \), \( g \) maps an open set containing \( f(E) \) into \( \mathbb{R}^k \), and \( g \) is differentiable at \( f(x_0) \). Then the mapping \( F \) of \( E \) into \( \mathbb{R}^k \), defined by \( F(x) = g(f(x)) \) is differentiable at \( x_0 \) and \( F'(x_0) = g'(f(x_0))f'(x_0) \) \( (i) \).
Here we note that the right side of (i) is the product of two linear transformations.

**Proof.** Put \( y_0 = f(x_0) \), \( A = f'(x_0) \), \( B = g'(y_0) \), and write
\[
\begin{align*}
\mathbf{u}(h) &= f(x_0 + h) - f(x_0) - Ah; \\
\mathbf{v}(k) &= g(y_0 + k) - f(y_0) - Bk
\end{align*}
\]
for all \( h \in \mathbb{R}^n \) and \( k \in \mathbb{R}^m \) for which \( f(x_0 + h) \) and \( g(y_0 + k) \) are defined. Then by differentiability of \( f \) and \( g \) at \( x_0 \) and \( y_0 \) respectively, we have
\[
\begin{align*}
\lim_{h \to 0} \frac{|f(x_0 + h) - f(x_0) - Ah|}{|h|} &= \lim_{h \to 0} \frac{u(h)}{|h|} = 0 \\
\text{and} \quad \lim_{k \to 0} \frac{|g(y_0 + k) - g(y_0) - Bk|}{|k|} &= \lim_{h \to 0} \frac{|v(k)|}{|k|} = 0
\end{align*}
\]
which implies that
\[
|u(h)| = \varepsilon(h) |h| \quad \text{and} \quad |v(k)| = \eta(k) |k|,
\]
where \( \varepsilon(h) \to 0 \) as \( h \to 0 \) and \( \eta(k) \to 0 \) as \( k \to 0 \).

Now for given \( h \) put \( k = f(x_0 + h) - f(x_0) \). Then \( |k| = |Ah + u(h)| \leq ||A|| + \varepsilon(h)|h| \) and
\[
\begin{align*}
F(x_0 + h) - F(x_0) - BAh &= g(f(x_0 + h)) - g(f(x_0)) - BAh = g(y_0 + k) - g(y_0) - BAh \\
&= v(k) + Bk - BAh = B(k - Ah) + v(k) = B(u(h) + v(k)).
\end{align*}
\]
Hence for \( h \neq 0 \), we have
\[
\begin{align*}
\frac{|F(x_0 + h) - F(x_0) - BAh|}{|h|} &= \frac{|Bu(h) + v(k)|}{|h|} \leq \frac{|Bu(h)| + |v(k)|}{|h|} \\
&\leq \frac{||B|| \varepsilon(h)|h| + \eta(k) |k|}{|h|} \\
&\leq ||B|| \varepsilon(h) + ||A|| + \varepsilon(h) \eta(k).
\end{align*}
\]
Letting \( h \to 0 \), then \( \varepsilon(h) \to 0 \), \( k \to 0 \) and \( \eta(k) \to 0 \).

It follows that
\[
\lim_{h \to 0} \frac{|F(x_0 + h) - F(x_0) - BAh|}{|h|} = 0
\]
i.e., \( F'(x_0) = BA = g'(f(x_0))f'(x_0) \).

### 3.4. Partial derivatives.

Let \( f \) be a function from an open set \( E \subset \mathbb{R}^n \) into \( \mathbb{R}^m \). Let \( \{e_1, \ldots, e_n\} \) and \( \{u_1, \ldots, u_m\} \) be the standard bases of \( \mathbb{R}^n \) and \( \mathbb{R}^m \) respectively. Then the components of \( f \) are the real functions \( f_1, \ldots, f_m \) defined by
\[
f(x) = (f_1(x), f_2(x), \ldots, f_m(x)) = f_1(x)(1, 0, \ldots, 0) + \ldots + f_m(x)(0, \ldots, 1)
\]
\[
= f_1(x)u_1 + \ldots + f_m(x)u_m = \sum f_i(x)u_i,
\]
where \( i = 1 \) to \( m \) and \( x \in E \) .

**3.4.1. Definition:** For \( x \in E \), \( 1 \leq i \leq m \), \( 1 \leq j \leq n \), we define
\[
D_jf_i(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t}
\]
provided the limit exists. Write \( f_i(x) = f_i(x_1, \ldots, x_n) \), we see that \( D_jf_i \) is the derivative of \( f_i \) with respect to \( x_j \), keeping the other variables \( x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n \) fixed. The notation \( \partial f_i / \partial x_j \) is often used in place of \( D_jf_i \), it is called a **partial derivative** of \( f_i \) with respect to \( x_j \).

In many cases when we deal with functions of one variable, the existence of a derivative is sufficient. On the other hand we need continuity or at least boundedness of the partial derivatives while dealing with the function of several variables. However, if \( f \) is known to be differentiate
at a point \( \mathbf{x} \), then its partial derivatives exist at \( \mathbf{x} \), and they determine the linear transformation \( f'(x) \) completely.

### 3.4.2. Theorem.\(^1\) Suppose \( f \) maps an open set \( E \subset \mathbb{R}^n \) into \( \mathbb{R}^m \) and \( f \) is differentiable at a point \( x \in E \). Then the partial derivatives \( (D_f)_i(x) \) exist and

\[
(f'_i(x))_{ij} = \sum (D_f)_j(x) u_i
\]

where \( \{e_1, \ldots, e_n\} \) and \( \{u_1, \ldots, u_m\} \) are the standard bases of \( \mathbb{R}^n \) and \( \mathbb{R}^m \).

**Proof.** Since \( f \) is differentiable at \( x \), then take \( h = te_j \) for fix \( j \), we have

\[
f(x + te_j) - f(x) = f'(x)(te_j) + r(te_j)
\]

where \( |r(te_j)/t| \to 0 \) as \( t \to 0 \).

Now by the linearity of \( f'(x) \), we have \( f'(te_j) = t f'(e_j) \), therefore by (ii), we have

\[
\lim_{t \to 0} \frac{f(x + te_j) - f(x)}{t} = f'(x)e_j
\]

On representing \( f \) in terms of its components (iii) becomes

\[
\lim_{t \to 0} \frac{1}{t} \sum [f_i(x + te_j) - f_i(x)] u_i = f'_i(x)e_j \quad (1 \leq i \leq m) \quad \ldots \quad \ldots \quad (iv)
\]

Since right side of (iv) exits therefore each quotient in this sum has a limit as \( t \to 0 \), so that each

\[
\lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t}
\]

i.e., \( (D_f)_i(x) \) exists. Hence from (iv), we have \( f'(x)e_j = \sum (D_f)_j(x) u_i \) \( (1 \leq i \leq m, 1 \leq j \leq n) \).

### 3.4.3. Directional Derivatives: Some consequences of Theorem 3.4.2:

1. Let \( [f'(x)] \) be the matrix that represents \( f'(x) \) with respect to bases \( \{e_1, e_2, \ldots, e_m\} \) and \( \{u_1, u_2, \ldots, u_n\} \). Then \( f'(x)e_j \) is the \( j \)-th column vector of \( [f'(x)] \) and (i) shows that the number \( (D_f)_j(x) \) occupies the spot in the \( i \)-th row and \( j \)-th column of \( [f'(x)] \). Thus

\[
[f'(x)] = \begin{pmatrix}
(D_f)_1(x) & \ldots & (D_f)_m(x) \\
\ldots & \ldots & \ldots \\
(D_f)_1(x) & \ldots & (D_f)_m(x)
\end{pmatrix}
\]

If \( h = \sum h_j e_j \) \((1 \leq j \leq n)\) be any vector in \( \mathbb{R}^n \). Then by (i) we have

\[
f'(x)h = f'(x) \sum h_j e_j \quad (1 \leq j \leq n)
\]

\[
= \sum h_j f'(x) e_j \quad (1 \leq j \leq n)
\]

\[
= \sum h_j \sum (D_f)_j(x) u_i \quad (1 \leq i \leq m, 1 \leq j \leq n)
\]

\[
= \sum \{\sum (D_f)_j(x) h_j\} u_i \quad (1 \leq i \leq m, 1 \leq j \leq n).
\]

\[
\quad \ldots \quad \ldots \quad (v)
\]

2. Let \( \gamma \) be a differentiable mapping of the segment \( (a, b) \subset \mathbb{R} \) into an open set \( E \subset \mathbb{R}^n \), that is \( \gamma \) be a differentiable curve in \( E \). Further let \( f \) be a real-valued differentiable function with domain in \( E \). Thus we may consider \( f \) to be a differentiable mapping of \( E \) into \( \mathbb{R} \). Let

\[
g(t) = f(\gamma(t)) \quad (a < t < b) \quad \ldots \quad \ldots \quad (v)
\]

Then by the chain rule \( g'(t) = f'(\gamma(t)) \gamma'(t) \) \((a < t < b)\) \( \ldots \quad \ldots \quad (vi)\)

Here we observe that \( \gamma'(t) \in L(\mathbb{R}, \mathbb{R}^n) \) and \( f'(\gamma(t)) \in L(\mathbb{R}^n, \mathbb{R}) \) therefore \( g'(t) \) is a linear operator on \( \mathbb{R} \), since if \( g \) maps \( (a, b) \) into \( \mathbb{R} \), furthermore \( g'(t) \) can be regarded as a real number which is
computed in terms of the partial derivatives of \( f \) and the derivatives of the components of \( \gamma \) as follows:

Let \( \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n) \), then \( \gamma(t) = \gamma_1(t)e_1 + \gamma_2(t)e_2 + \cdots + \gamma_n(t)e_n \), and \( \gamma'(t) = \gamma_1'(t)e_1 + \gamma_2'(t)e_2 + \cdots + \gamma_n'(t)e_n \) and so \( [\gamma'(t)] \) is a column matrix of order \( n \) by \( 1 \) in which \( \gamma_i'(t) \) is in the \( i^{th} \) row with respect to the standard basis \( \{e_i, e_1, \ldots, e_n\} \) of \( \mathbb{R}^n \). Again for all \( x \in E \), \( [f \ '](x) \) is a row matrix of order \( 1 \) by \( n \) which has \( (D_j f)(x) \) in the \( j^{th} \) column. Hence \( [g'(t)] \) is the \( 1 \) by \( 1 \) matrix of real number

\[
g'(t) = \sum (D_j f) \gamma'_i(t) \quad (1 \leq i \leq n) \quad \ldots \quad \ldots \quad \ldots \quad (vii).
\]

Now we define the **gradient** of \( f \) at each \( x \in E \) by \( (\Delta f)(x) = \sum (D_j f)(x)e_j \) \( (1 \leq i \leq n) \) and since \( \gamma'_i(t) = \sum \gamma'_i(x) e_j \) \( (1 \leq i \leq n) \). Hence we have

\[
g'(t) = (\Delta f)(\gamma(t)) \quad \ldots \quad \ldots \quad \ldots \quad (ix),
\]

the scalar product of the vectors \( (\Delta f)(\gamma(t)) \) and \( \gamma'(t) \). Now we define \( \gamma(t) = x + tu \quad \ldots \quad (x) \) for \( t \in (-\infty, \infty) \) for fix \( x \in E \) and unit vector \( u \in \mathbb{R}^n \). Then \( \gamma(t) = u \quad \text{for all } t. \) Hence for \( t = 0 \) \( (ix) \) becomes

\[
g'(0) = (\Delta f)(x) \cdot u \quad \ldots \quad \ldots \quad \ldots \quad (xi).
\]

Hence by \( (x) \), we have

\[
g(t) - g(0) = f(\gamma(t)) - f(\gamma(0)) = f(x + tu) - f(x) + tu.
\]

Further

\[
g'(0) = \lim_{t \to 0} \frac{g(t) - g(0)}{t} = \lim_{t \to 0} \frac{f(x + tu) - f(x)}{t}.
\]

Hence by \( (xi) \) \( \lim_{t \to 0} \frac{f(x + tu) - f(x)}{t} = (\Delta f)(x) \cdot u \quad \ldots \quad \ldots \quad \ldots \quad (xii) \). Here in \( (xii) \) is usually called the **directional derivative** of \( f \) at \( x \) in the direction of unit vector \( u \) and is denoted by \( (D_u f)(x) \).

On keeping \( f \) and \( x \) fixed and \( u \) varies, then \( (xii) \) shows that the direction derivatives of \( f \) at \( x \) attains its maximum when \( u \) is a positive scalar multiple of \( (\Delta f)(x) \). Here we exclude the case \( (\Delta f)(x) = 0 \). If we put \( u = \sum u, e_i \) \( (1 \leq i \leq n) \), then \( (D_u f)(x) = \sum (D_i f)(x)u_i \) \( (1 \leq i \leq n) \).

3.4.4.**Theorem.** Suppose \( f \) maps a convex open set \( E \subset \mathbb{R}^n \) into \( \mathbb{R}^m \), \( f \) is differential in \( E \), and there is real number \( M \) such that

\[
\| f'(x) \| \leq M
\]

for every \( x \in E \). Then

\[
|f(b) - f(a)| \leq M|b - a| \quad \text{for all } a, b \in E.
\]

**Proof.** Fix \( a, b \in E \) and define \( \gamma(t) = (1-t)a + tb \quad \ldots \quad \ldots \quad \ldots \quad (i) \) for all \( t \in \mathbb{R}^1 \) such that \( \gamma(t) \in E \). Since if \( 0 \leq t \leq 1 \) then by convexity of \( E \), \( \gamma(t) \in E \). Now we put \( g(t) = f(\gamma(t)) \). Then by chain rule and by \( (i) \), we have

\[
g'(t) = f'(\gamma(t)) \gamma'(t) = f'(\gamma(t))(b - a).
\]

So that

\[
|g'(t)| \leq \| f'(\gamma(t)) \| \| b - a \| \leq M|b - a| \quad \text{for all } t \in [0, 1] \ldots \ldots \ldots \quad (ii)
\]

Now for all \( t \in [0, 1] \) by mean value theorem for vector valued function, we have

\[
|g(1) - g(0)| \leq (1-0)|g'(t)| \leq M|b - a|
\]

Now since \( g(0) = f(\gamma(0)) = f(a) \) and \( g(1) = f(\gamma(1)) = f(b) \). Hence \( f(b) - f(a) \leq M|b - a| \).

This complete the proof.

3.4.5.**Corollary.** If \( f'(x) = 0 \) for all \( x \in E \) in theorem 3.4.5, then \( f \) is constant.

**Proof.** Since if \( f'(x) = 0 \) for all \( x \in E \) then by \( (ii) \), we have \( g'(t) = 0 \).
Therefore \(|g(1) - g(0)| \leq 0\),
i.e., \(g(1) = g(0)\) or \(f(b) = f(a)\), proves that \(f\) is constant.

3.4.6. Definition: \((Continuously \ differentiable \ function)\): A differentiable mapping \(f\) of an open set \(E \subset \mathbb{R}^n\) into \(\mathbb{R}^m\) is said to be \(\mathbb{R}^n\) continuously differentiable in \(E\) if \(f'\) is a continuous mapping of \(E\) into \(L(\mathbb{R}^n, \mathbb{R}^m)\).

In other word, a differentiable mapping \(f\) of an open set \(E \subset \mathbb{R}^n\) into \(\mathbb{R}^m\) is said to be continuously differentiable if for every \(x \in E\) and to every \(\varepsilon > 0\) there exists a \(\delta > 0\) such that \(|f'(y) - f'(x)| < \varepsilon\) if \(y \in E\) and \(|x - y| < \delta\).

A continuously differentiable mapping \(f\) is also known as a \(\zeta\)-mapping, or that \(f \in \zeta(E)\).

3.4.7. Theorem: Suppose \(f\) maps an open set \(E \subset \mathbb{R}^n\) into \(\mathbb{R}^m\). Then \(f \in \zeta(E)\) if and only if the partial derivatives \(D_{j}\) exist and are continuous on \(E\) for \(1 \leq i \leq m, 1 \leq j \leq n\).

**Proof.** Assume first that \(f \in \zeta(E)\) and if \(\{e_1, e_2, \ldots, e_n\}\) and \(\{u_1, u_2, \ldots, u_m\}\) are the standard basis of \(\mathbb{R}^n\) and \(\mathbb{R}^m\) respectively. Since the partial derivative of \(f_i\) w.r.t. \(x_j\) is given by \((D_{j}f_i)(x) = (f'(x)e_j)u_i\) for all \(i, j\) and for all \(x \in E\). Hence

\[
(D_{j}f_i)(y) - (D_{j}f_i)(x) = \left\{ (f'(y) - f'(x))e_j \right\} u_i.
\]

and since \(|u_i| = |e_j| = 1\), it follows that

\[
|D_{j}f_i(y) - (D_{j}f_i)(x)| = \left| \left\{ (f'(y) - f'(x))e_j \right\} u_i \right| \leq \left| (f'(y) - f'(x))e_j \right| \|u_i\| \leq \left| (f'(y) - f'(x))e_j \right|,
\]

(since \(|x - y| \leq |x| \leq |y|\))

\[
= \left| (f'(y) - f'(x))e_j \right| \leq \left| (f'(y) - f'(x)) \right| \leq ||f'(y) - f'(x)|| \leq |f'(y) - f'(x)| \leq \varepsilon.
\]

(i)

Since \(f'\) is continuous therefore for \(\varepsilon > 0\) there exists \(\delta > 0\) such that

\[
|x - y| < \delta \Rightarrow |f'(x) - f'(y)| < \varepsilon \quad \forall \ x, y \in E.
\]

Thus from (i) and (ii) for \(\varepsilon > 0\) there exists \(\delta > 0\) such that

\[
|x - y| < \delta \Rightarrow |D_{j}f_i(y) - (D_{j}f_i)(x)| < \varepsilon/n \quad (y \in B_{r}(x), 1 \leq j \leq n) \quad \ldots \quad \ldots \quad (iii)
\]

Suppose \(h = \sum_{l} h_{l}e_{j}, \ |h| < r \quad (1 \leq j \leq n), \) put \(v_{0} = 0\) and \(v_{k} = h_{1}e_{1} + \ldots + h_{n}e_{n}\), for \(1 \leq k \leq n\). Then

\[
f(x + h) - f(x) = f(x + h_{1}e_{1} + \ldots + h_{n}e_{n}) - f(x) = \left[ f(x + h_{1}e_{1} + \ldots + h_{n}e_{n}) - f(x + h_{1}e_{1} + \ldots + h_{n}e_{n}) \right] + \left[ f(x + h_{1}e_{1} + \ldots + h_{n}e_{n} - f(x + h_{1}e_{1} + \ldots + h_{n}e_{n}) \right] + \ldots + \left[ f(x + e_{j}) - f(x + 0) \right].
\]

\[
= \left[ f(x + v_{0}) - f(x + v_{0}) \right] + \left[ f(x + e_{j}) - f(x + e_{j}) \right] + \left[ f(x + v_{1}) - f(x + v_{1}) \right] + \ldots + \left[ f(x + v_{n}) - f(x + v_{n}) \right] = \sum [f(x + v_{j}) - f(x + v_{j})] (1 \leq j \leq n) \quad \ldots \quad (iv)
\]

Since \(|v_{k}| < r\) for \(1 \leq k \leq n\) and since \(B_{r}(x)\) is convex, the segments with end points \(x + v_{j}\) and \(x + v_{j+1}\) lie in \(B_{r}(x)\). Since \(v_{j} = v_{j+1} + h_{j}e_{j}\), the mean value theorem shows that the \(j^{th}\) summand in (iv) is given by \(f(x + v_{j+1}) - f(x + v_{j}) = h_{j}(D_{j}f)(x - v_{j-1}; 0, h_{j}e_{j}) \) for some \(0 < \theta < 1\), and then \(|h_{j}(D_{j}f)(x - v_{j-1}; 0, h_{j}e_{j}) - (D_{j}f)(x)| = \left| h_{j} \right| \left| (D_{j}f)(x - v_{j-1}0, h_{j}e_{j}) - (D_{j}f)(x) \right|\)
By (iv), it follows that for \(1 \leq j \leq n\),
\[
|\Sigma h_i(D_jf)(x-v_j, \theta[h_i e_j] - \Sigma h_i(D_jf)(x)| < \Sigma |h_i|e/n < |h|e < re,
\]
Or, \(|f(x + h) - f(x) - \Sigma h_i(D_jf)(x)| < re. This inequality shows that \(f\) is differentiable at \(x\) and
that \(f'(x)\) is the linear function which assigns the number \(\Sigma h_i(D_jf)(x)\) to the vector \(h = \Sigma h_i e_j\).
The matrix \([f'(x)]\) consists of the row \((D_1f)(x), \ldots, (D_nf)(x)\), and since \((D_1f), \ldots, (D_nf)\) are
continuous functions on \(E\).

3.5. Derivatives of Higher Order.

Definition 3.5.1. Let \(f\) is a real valued function defined in an open set \(E \subset \mathbb{R}^n\), is called the
function of \(n\)-variables, with partial derivatives \(D_1f, D_2f, \ldots, D_nf\). If the functions \(D_jf\) are
themselves differentiable, then the second-order partial derivatives of \(f\) are defined by \(D_{ij}f = D_i(D_jf)\)
\((i, j = 1, \ldots, n)\). For \(i = j\), the second order partial derivative written as \(D_i^2f\). If all these
functions \(D_{ij}f\) are continuous in \(E\), we say that \(f\) is of class \(\zeta''\) in \(E\), or that \(f \in \zeta''(E)\). A
mapping \(f\) of \(E\) into \(\mathbb{R}^m\) is said to be of class \(\zeta''\) if each component \(f_k\) \((k = 1, 2, \ldots, m)\)
of \(f = (f_1, f_2, \ldots, f_m)\) is of class \(\zeta''\) i.e., \(f \in \zeta''(E)\). The third order partial derivatives of \(f\) are
defined by \(D_{ijk}f = D_i(D_jD_kf)\) \((i, j, k = 1, \ldots, n)\). For \(j = k\), \(i = j\) and \(i = j = k\) the third order
partial derivative written as \(D_i^3f, D_i^2D_kf\) and \(D_i^3f\) respectively. Similarly fourth order, fifth
order, \ldots partial order can be defined. \(D_1^{m_1}D_2^{m_2} \cdots D_n^{m_n}\) is called the \(m^{th}\) order partial derivatives,
wher \(m = m_1 + m_2 + \ldots + m_n\). If all \(m^{th}\) order partial derivatives are continuous in \(E\), we call \(f \in \zeta''(E)\).

Repeated partial derivatives. If the derivatives \((D_{ij}f)(x) = (D_iD_j)(x)\) exists, it is called a
repeated partial derivatives.

Interchange of order of differentiation. In general, the order of repeated partial derivatives can not be interchanged, i.e., it can happen that \(D_{ij}f \neq D_{ji}f\) at some points, although both derivatives
exist. For example consider the function \(f(x, y) = \begin{cases} xy^2 - y^2, & (x, y) \neq (0, 0), \\
0, & (x, y) = (0, 0) \end{cases}\), we have

\[
(D_xf)(x, y) = y \frac{(3x^2 - y^2)(x^2 + y^2) - 2x^2(x^2 - y^2)}{(x^2 + y^2)^2} = y \frac{x^4 + 4x^2y^2 - y^4}{(x^2 + y^2)^2} , (x, y) \neq (0, 0), \text{ and}
\]

\[
D_xf(0, 0) = \lim_{h \to 0} \frac{f(h, 0) - f(0, 0)}{h} = 0 . \text{ Hence } (D_xf)(0, y) = -y \text{ for all } y, \text{ therefore}
\]

\[
(D_{xy}f)(0, y) = -1 \text{ and } (D_{xy}f)(0, 0) = -1 . \text{ But}
\]

\[
(D_yf)(x, y) = x \frac{(3x^2 - 3y^2)(x^2 + y^2) - 2y^2(x^2 - y^2)}{(x^2 + y^2)^2} = x \frac{x^4 - 4x^2y^2 - y^4}{(x^2 + y^2)^2}, (x, y) \neq (0, 0), \text{ and}
\]

\[
D_yf(0, 0) = \lim_{k \to 0} \frac{f(0, k) - f(0, 0)}{k} = 0 . \text{ Hence } (D_yf)(0, x) = x \text{ for all } x, \text{ therefore}
\]

\[
(D_{yx}f)(x, 0) = 1 \text{ and } (D_{yx}f)(0, 0) = 1 . \text{ Thus } (D_{yx}f)(0, 0) \neq (D_{xy}f)(0, 0).
\]

However, our next results give the criteria that \(D_{ij}f = D_{ji}f\), whenever these derivatives are
continuous.
3.5.2. Theorem (Mean Value): Suppose $f$ is defined in an open set $E \subset \mathbb{R}^2$ and $D_1f$, $D_2f$ exist at every point of $E$. Suppose $Q \subset E$ is a closed rectangle with sides parallel to the coordinate axes, having $(a, b)$ and $(a + h, b + k)$ as opposite vertices ($h \neq 0, k \neq 0$). Put

$$\Delta(f, Q) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b).$$

Then there is a point $(x, y)$ in the interior of $Q$ such that

$$\Delta(f, Q) = hk(D_{21}f)(x, y).$$

**Proof.** Put $u(t) = f(t, b + k) - f(t, b)$. Now by the mean value theorem for real valued function stated as “If $f$ is a continuous real valued function on $[a, b]$, which is differentiable in $(a, b)$, then there exists a point $x$ in $(a, b)$ such that

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

Two applications of this theorem are, there is a $x$ between $a$ and $a + h$, and a $y$ between $b$ and $b + k$, such that

$$\Delta(f, Q) = u(a + h) - u(a) = hu'(x)$$

$$= h[(D_1f)(x, b + k) - (D_1f)(x, b)] = hk(D_{21}f)(x, y).$$

3.5.3. Theorem. Suppose $f$ is a real value function defined in an open set $E \subset \mathbb{R}^2$, suppose that $D_1f$, $D_2f$ and $D_{21}f$ exist at every point of $E$, and $D_{22}f$ is continuous at some point $(a, b) \in E$. Then $D_{12}f$ exists at $(a, b)$ and $(D_{12}f)(a, b) = (D_{21}f)(a, b)$.

**Or.** Suppose $f$ is a real value function defined in an open set $E \subset \mathbb{R}^2$, suppose that the partial derivatives $D_1f$, $D_2f$, $D_{12}f$ and $D_{22}f$ and continuous in $E$. Then

$$(D_{12}f)(a, b) = (D_{21}f)(a, b), \ (a, b) \in E, \text{ or } D_{21}f = D_{12}f \text{ if } f \in \zeta''(E).$$

**Proof.** Let $(a, b) \in E$ be fix arbitrary element, since $E$ is open and so there exists an open ball $B(0, r)$ such that $(a, b) + (h, k) \in E$ for all $(h, k) \in B(0, r)$. Now define a function $\Delta : B(0, r) \rightarrow r$ by $\Delta(h, k) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b)$ for $(h, k) \in B(0, r)$. Consider the function $u(t) = f(t, b + k) - f(t, b)$. Then by mean value theorem $u(a + h) - u(a) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b) = hu'(a + \theta_1h), \ \theta_1 \in (0, 1)$

$$= h[(D_1f)(a + \theta_1h, b + k) - (D_1f)(a + \theta_1h, b)].$$

Further $\Delta(h, k) = u(a + h) - u(a) = hu'(x)$

$$= h[(D_1f)(a + \theta_1h, b + k) - (D_1f)(a + \theta_1h, b)]$$

$$= hk(D_{21}f)(a + \theta_1h, b + \theta_2k), \ \theta_2 \in (0, 1).$$

Now by the continuity of $D_{21}f$, we have

$$\lim_{(h,k) \to (0,0)} \frac{\Delta(h,k)}{hk} = (D_{21}f)(a,b).$$

Similarly

$$\lim_{(h,k) \to (0,0)} \frac{\Delta(h,k)}{hk} = (D_{21}f)(a,b).$$

Hence $(D_{12}f)(a, b) = (D_{21}f)(a, b) \forall (a, b) \in E$.

This theorem can be extended to the function of $n$-variables as:

“Suppose $f$ is a real value function defined in an open set $E \subset \mathbb{R}^n$, has continuous partial derivatives of order two $(D_{ij}f)(x) = (D_{ji}f)(x)$.”

3.5.4. Taylor’s Theorem. In this section we extend the one dimensional Taylor’s theorem to real-valued function $f$ defined on subset of $\mathbb{R}^n$. Here we use some notations:

$L(x, y) = \{(1 - t)x + ty : t \in [0, 1]\}$, the line segment joining the point $x, y$ in $\mathbb{R}^n$.

An open ball $B(x_0, r)$ contains line segment $L(x, y)$, whenever $x, y \in B(x_0, r)$.

A real-valued function $f : E \subset \mathbb{R}^n \rightarrow \mathbb{R}$ of $n$-variables defined on an open set $E$ of $\mathbb{R}^n$ is said to be of class $\zeta$ if all repeated partial derivatives $D_{ijk...}f$ are continuous in $E$ and we write $f \in \zeta(E)$. 
If $x$ is a point in $\mathbb{R}^n$ where second order, third order, ... partial derivatives of $f$ exist at $x$ and if $u = (u_1, u_2, \ldots, u_n)$ is an arbitrary point in $\mathbb{R}^n$, we write
\[
(D_a^2 f)(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} (D^2_{ij}f)(x) u_i u_j,
\]
and these sum are analogous to the formula for the directional derivatives of the function $f : E \subset \mathbb{R}^n \to \mathbb{R}$, which is differentiable at $x$; i.e.,
\[
(D_a f)(x) = \sum_{j=1}^{n} (D_j f)(x) u_j.
\]

**Theorem (Taylor’s).** Let $f : E \subset \mathbb{R}^n \to \mathbb{R}$ be a function, where $E$ is an open subset $\mathbb{R}^n$. Suppose $f \in C^{(m)}$, and $a \in E$ and $x \in \mathbb{R}^n$, further suppose that $L(a, a + x) \subset E$. Then there exist $\theta \in (0, 1)$ such that
\[
f(a + x) = f(a) + \frac{(D_x f)(a)}{1!} \frac{1}{x} + \frac{(D_x^2 f)(a)}{2!} \frac{1}{2} + \cdots + \frac{(D_x^{m-1} f)(a)}{(m-1)!} \frac{1}{m!} + \frac{(D_x^m f)(a + \theta x)}{m!},
\]
where $D_x = x \cdot \nabla$.

**Proof.** Fix an arbitrary $a \in E$ and choose $a \in \mathbb{R}^n$ such that $L(a, a + x) \subset E$. Define a function $g : I \to \mathbb{R}^n$ by $g(t) = a + tx$, where $I$ is an open interval in $\mathbb{R}$ containing $[0, 1]$ as a subset. Then the function $h : I \to \mathbb{R}$ defined by $h(t) = f(g(t))$ satisfies the conditions of Taylor’s theorem for the function of one variable. Hence
\[
h(1) = h(0) + h'(0) + \frac{h''(0)}{2} + \cdots + \frac{h^{(m-1)}(0)}{(m-1)!} + \frac{h^{(m)}(\theta)}{m!},
\]
for some $\theta \in (0, 1)$ ... (i)

Since here $h$ is a composite function $h(t) = f(g(t))$, where $g(t) = a + tx$. The $k$-th component of $p$ has derivative $p_k(t) = x_k$. Then by the chain rule, we have $h'(t)$ exists in $I$ and is given by
\[
h'(t) = \sum_{j=1}^{n} (D_j f)(p(t)) x_j = (D_x f)(p(t)).
\]
Again by chain rule
\[
h''(t) = \sum_{i=1}^{n} \sum_{j=1}^{n} (D^2_{ij} f)(p(t)) x_i x_j = (D_x^2 f)(p(t)), \ldots,
\]
\[
h^{(m)}(t) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_m=1}^{n} (D^m_{i_1 \cdots i_m} f)(p(t)) x_{i_1} x_{i_2} \cdots x_{i_m} = (D_x^m f)(p(t)).
\]
Now at $t = 1$, $t = \theta$ and $t = 0$, we have $h(1) = f(g(1)) = f(a + x)$, $h(\theta) = f(g(\theta)) = f(a + \theta x)$ and $h(0) = f(g(0)) = f(a)$
\[
h'(0) = \sum_{j=1}^{n} (D_j f)(a) x_j = (D_x f)(a), h''(0) = \sum_{i=1}^{n} \sum_{j=1}^{n} (D^2_{ij} f)(a) x_i x_j = (D_x^2 f)(a), \ldots,
\]
\[
h^{(m-1)}(0) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_{m-1}=1}^{n} (D^{m-1}_{i_1 \cdots i_{m-1}} f)(a) x_{i_1} x_{i_2} \cdots x_{i_{m-1}} = (D_x^{m-1} f)(a)
\]
\[
h^{(m)}(\theta) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_m=1}^{n} (D^m_{i_1 \cdots i_m} f)(a + \theta x) x_{i_1} x_{i_2} \cdots x_{i_m} = (D_x^m f)(a + \theta x).
\]
Then by (i), we have
\[
f(a + x) = f(a) + \frac{(D_x f)(a)}{1!} + \frac{(D_x^2 f)(a)}{2!} + \cdots + \frac{(D_x^{m-1} f)(a)}{(m-1)!} + \frac{(D_x^m f)(a + \theta x)}{m!}
\]
\[
\lim_{x \to 0} \frac{r_m(x)}{x^m} = 0.
\]

Some forms. 1. Replace \( x \) by \( x - a \) in (i), we have

\[
f(x) = f(a) + \frac{(D f)(a)}{1!} + \frac{(D^2 f)(a)}{2!} + \ldots + \frac{(D^{m-1} f)(a)}{(m-1)!} + \frac{(D^m f)(z)}{m!},
\]

where \( D = (x-a) \cdot \nabla \) and \( z \in \mathbb{R}(a, x) \). This identity is called Taylor's formula with remainder.

If \( \lim_{m \to \infty} \frac{D^m f(z)}{m!} = 0 \), then, we have

\[
f(x) = f(a) + \frac{(D f)(a)}{1!} + \frac{(D^2 f)(a)}{2!} + \ldots + \frac{(D^m f)(z)}{m!} + \ldots.
\]

This is called the Taylor's expansion or Taylor's series of \( f \) at \( a \).

3.6. The Inverse Function Theorem. The inverse function theorem states, roughly speaking, that a continuously differentiable mapping \( f \) is invertible in a neighborhood of any point \( x \) at which the linear transformation \( f'(x) \) is invertible:

3.6.1. Theorem (The inverse function). Suppose \( f \) is a \( \zeta' \)-mapping of an open set \( E \subset \mathbb{R}^n \) into \( \mathbb{R}^n \), \( f'(a) \) is invertible for some \( a \in E \) and \( b = f(a) \). Then (a) there exist open sets \( U \) and \( V \) in \( \mathbb{R}^n \) such that \( a \in U, b \in V, \ f \) is one-to-one on \( U \), and \( f(U) = V \) (b) If \( g \) is the inverse of \( f \), defined in \( V \) by \( g(f(x)) = x \) \((x \in U)\), then \( g \in \zeta'(V) \). In component form the conclusion of this theorem can be stated as:

Suppose \( y = f(x) \) is represented in component form as \( y_i = f_i(x_1, x_2, \ldots, x_n)(1 \leq i \leq n) \). Then this system of equations can be solved for \( x_1, x_2, \ldots, x_n \) in terms of \( y_1, y_2, \ldots, y_n \) if we restrict \( x \) and \( y \) in sufficiently small neighbourhood of \( a \) and \( b \) respectively. Moreover, the solutions are unique and continuously differentiable.

Proof. (a) Put \( f'(a) = A \), and choose \( \lambda \) so that \( 2\lambda \cdot \| A^{-1} \| = 1 \) \( \ldots \ldots \ldots \ldots \) (i)

Since \( f \) is a \( \zeta' \)-mapping i.e., \( f' \) is continuous on \( E \). In particular \( f' \) is continuous at \( a \) in \( E \). Then there is an open ball \( U \subset E \), with center at \( a \) such that

\[
\|f'(x) - f'(a)\| < \lambda \quad (x \in U), \quad \text{i.e.,} \quad \|f'(x) - A\| < \lambda \quad (x \in U) \ldots \ldots \ldots (ii)
\]

We associate to each \( y \in \mathbb{R}^n \) a function \( \phi \), defined by

\[
\phi(x) = x + A^{-1}(y - f(x)) \quad \text{for} \quad (x \in U) \ldots \ldots \ldots (iii)
\]

Note that \( f(x) = y \) then \( \phi(x) = x \) and if \( \phi(x) = x \), then \( x = x + A^{-1}(y - f(x)) \),

\( \text{or} \quad 0 = A^{-1}(y - f(x)) \) which implies that \( y = f(x) = 0 \) \text{or} \( y = f(x) \).

Hence \( y = f(x) \) iff \( x \) is a fixed point of \( \phi \). Now since \( \phi'(x) = I - A^{-1}(f'(x)) = A^{-1}A - A^{-1}(f'(x)) = A^{-1}(A - f'(x)) \).

Then (i) and (ii) imply that
\[ \|\phi'(x)\| = \|A^{-1}(A - f'(x))\| \leq \|A^{-1}\| \|A - f'(x)\| < (1/2\lambda)\|A\| = \frac{1}{2} \quad (x \in U) \quad \ldots \quad \ldots \quad (iv) \]

Hence by theorem 3.4.4, we have \[ |\phi(x_1) - \phi(x_2)| \leq \frac{1}{2} |x_1 - x_2| \] for \( x_1, x_2 \in U \) \( \ldots \quad \ldots \quad (v) \)

It follows that \( \phi \) is a contraction mapping, hence by Banach contraction principle \( \phi \) has unique fixed point in \( U \), so that \( f(x) = y \) for at most one \( x \in U \). Thus \( f \) is 1-1 in \( U \).

Next, let \( V = f(U) \), and choose an arbitrary \( y_0 \in V \), then \( y_0 = f(x_0) \) for some \( x_0 \in U \). Let \( B_r(x_0) \) be an open ball with center at \( x_0 \) and radius \( r > 0 \), so small that its closure \( \overline{B_r(x_0)} \) lies in \( U \), it is possible since \( U \) is an open set. Now our aim is to prove that \( V \) is open, to prove it we shall prove that \( y \in V \) whenever \( |y - y_0| < \lambda r \). Fix \( y \) such that \( |y - y_0| < \lambda r \), now by definition of \( \phi \), we see that

\[ |\phi(x_0) - x_0| = |A^{-1}(y - f(x_0))| = |A^{-1}(y - y_0)| \leq \|A^{-1}\| |y - y_0| < (1/2\lambda)\lambda r = r/2. \]

If \( x \in \overline{B_r(x_0)} \), then from (5), we have

\[ |\phi(x) - x| = |\phi(x) - \phi(x_0) + \phi(x_0) - x_0| \leq |\phi(x) - \phi(x_0)| + |\phi(x_0) - x_0| < (1/2)|x - x_0| + r/2 < (r/2) + (r/2) = r. \]

Hence \( \phi(x) \in \overline{B_r(x_0)} \). Since as per our choice \( B_r(x_0) \) lies in \( U \) and so (v) holds if \( x_1, x_2 \in \overline{B_r(x_0)} \). Thus \( \phi \) is a contraction of \( \overline{B_r(x_0)} \) into \( \overline{B_r(x_0)} \). Since \( \overline{B_r(x_0)} \) is closed subset of \( \mathbb{R}^n \) so that \( \overline{B_r(x_0)} \) is complete. Therefore by Banach contraction principle \( \phi \) has a fixed poin \( x \in \overline{B_r(x_0)} \) and for this \( x, f(x) = y \). Thus \( y \in f(\overline{B_r(x_0)}) \subset f(U) = V \). This proves part (a) of the theorem.

(b) Choose \( y, y + k \in V \) and since \( f(U) = V \), there exist \( x, x + h \in U \) such that \( y = f(x), y + k = f(x + h) \). Now by definition of \( \phi \) as in (iii), we have

\[ \phi(x + h) - \phi(x) = [x + h + A^{-1}(y - f(x + h))] - [x + A^{-1}(y - f(x))] \]

\[ = h + A^{-1}[f(x) - f(x + h)] = h + A^{-1}[y - y - k] = h - A^{-1}k. \]

Then by (5), we have \[ |h| - A^{-1}|k| \leq |\phi(x + h) - \phi(x)| \leq (1/2)|h| + (1/2)|h - x| = (1/2)|h|, \] which implies that \[ |A^{-1}k| = |h - A^{-1}k| \leq |h| - |A^{-1}k| \leq |h| - (1/2)|h| = 1/2|h|, \] therefore \[ |h| \leq 2\lambda^{-1}|k| \leq 2 \|A^{-1}\| |k| = \lambda^{-1}|k| \quad \ldots \quad \ldots \quad (vi) \]

Now by (i), (ii) and Theorem 3.2.5, \( f'(x) \) has an inverse, say \( T \). Since \( g(y + k) - g(y) - Tk = g(f(x + h) - g(f(x))) - Tk = x + h - x - Tk = h - Tk \)

\[ = -T[f(x + h) - f(x) - f'(x)h]. \]

Therefore by (vi), we have

\[ \frac{1}{|k|} \frac{|g(y + k) - g(y) - Tk|}{|k|} \leq \|T\| \frac{|f(x + h) - f(x) - f'(x)h|}{|h|} \rightarrow 0 \quad \text{as} \quad k \rightarrow 0 \] (since by (vi) \( h \rightarrow 0 \) as \( k \rightarrow 0 \)).

Hence \( g'(y) = T = \{f'(y)\}^{-1} = \{f'(g(y))\}^{-1} \quad (y \in V) \quad \ldots \quad \ldots \quad (vii) \)

Finally, note that \( g \) is a continuous mapping of \( V \) onto \( U \) (since \( g \) is differentiable), that \( f' \) is a continuous mapping of \( U \) into the set \( \Omega \) of all invertible elements of \( L(\mathbb{R}^n) \), and that inversion is a continuous mapping of \( \Omega \) onto \( \Omega \), by Theorem 3.2.5. If we combine these facts with (i), we see that \( g \in \mathcal{C}'(V) \). This completes the proof.

The following is an immediate consequence of part(a) of the inverse function theorem.
3.6.2 Theorem. If \( f \) is a \( \zeta' \) -mapping of an open set \( E \subset \mathbb{R}^n \) into \( \mathbb{R}^n \) and if \( f'(x) \) is invertible for every \( x \in E \), then \( f(W) \) is an open subset of \( \mathbb{R}^n \) for every open set \( W \subset E \). In other words, \( f \) is an open mapping of \( E \) into \( \mathbb{R}^n \).

3.7. The Implicit Function Theorem. If \( f \) is a continuously differentiable real function in the plane, then the equation \( f(x, y) = 0 \) can be solved for \( y \) in terms of \( x \) in a neighborhood of any point \((a, b)\) at which \( f(a, b) = 0 \) and \( \partial f/\partial y \neq 0 \). Likewise, one can solve for \( x \) in terms of \( y \) near \((a, b)\) if \( \partial f/\partial x \neq 0 \) at \((a, b)\). For example consider \( f(x, y) = x^2 + y^2 - 1 \) and a point \((0, 1)\) so that \((i)\) \( f(0, 1) = 0 \) and \((ii)\) \( \partial f/\partial y \neq 0 \) at \((0, 1)\). Here the possible solutions are \( y = +\sqrt{1-x^2} \) is the implicit function in a neighbourhood of \((0, 1)\), where \(|x| < 1\), \( y > 0 \) and \( y = -\sqrt{1-x^2} \) is the implicit function in a neighbourhood of \((0, 1)\) where \(|x| < 1\), \( y < 0 \). Our next result is the very informal statement is the simplest case (the case \( m = n = 1 \) of "implicit function theorem." Its proof makes strong use of the fact that continuously differentiable transformation behave locally very much like their derivatives. Accordingly, we first prove the linear version of implicit function theorem as:

Notation. In our further discussion we shall make the use of the following notations: If \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \) and \( y = (y_1, \ldots, y_m) \in \mathbb{R}^m \), then we write \((x, y)\) for the point (or vector) as \((x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_m) \in \mathbb{R}^{n+m}\). Then an arbitrary \( A \in L(\mathbb{R}^{n+m}, \mathbb{R}^n) \) can be split into two linear transformations \( A_x \) and \( A_y \) defined by \( A_xh = A(h, 0) \) and \( A_yk = A(0, k) \) for all \( h \in \mathbb{R}^n \), \( k \in \mathbb{R}^m \). Then clearly \( A_x \in L(\mathbb{R}^n) \) and \( A_y \in L(\mathbb{R}^m, \mathbb{R}^n) \) and \( A(h, k) = A_xh + A_yk \).

3.7.1. Theorem(Implicit function theorem, linear version). If \( A \in L(\mathbb{R}^{n+m}, \mathbb{R}^n) \) and if \( A_x \) is invertible, then there corresponds to every \( k \in \mathbb{R}^m \) a unique \( h \in \mathbb{R}^n \) such that \( A(h, k) = 0 \), and \( h \) can be computed from \( k \) by the formula \( h = -(A_x)^{-1}A_yk \).

Proof. Since by \( A(h, k) = A_xh + A_yk \), then \( A(h, k) = 0 \Leftrightarrow A_xh + A_yk = 0 \). Now since \( A_x \) is invertible then we get \( h = -(A_x)^{-1}A_yk \). The conclusion of Theorem 3.7.1 can be restated as: "The equation \( A(h, k) = 0 \) can be solved (uniquely) for \( h \) if \( k \) is given and that the solution \( h \) is a linear function of \( k \)."

3.7.2. Theorem(Implicit function theorem, non-linear version). Let \( f \) be a \( \zeta' \) -mapping of an open set \( E \subset \mathbb{R}^{n+m} \) into \( \mathbb{R}^n \), such that \( f(a, b) = 0 \) for some point \((a, b) \in E \). Put \( A = f'(a, b) \) and assume that \( A_x \) is invertible, then there exist open sets \( U \subset \mathbb{R}^{n+m} \) and \( W \subset \mathbb{R}^m \) with \((a, b) \in U \) and \((a, b) \in W \), having the following property:

To every \( y \in W \) corresponds a unique \( x \) such that \((x, y) \in U \) and \( f(x, y) = 0 \) ... ... (i)

If this \( x \) is defined to be \( g(y) \), then \( g \) is a \( \zeta' \) -mapping of \( W \) into \( \mathbb{R}^n \), \( g(b) = a \),

\[ f(g(y), y) = 0 \quad (y \in W) \] ... ... (ii)

and

\[ g(b) = -(A_x)^{-1}A_y. \] ... ... (iii)

The function \( g \) is "implicitly" defined by (ii), hence the name of the theorem.

Note: The equation \( f(x, y) = 0 \) can be written as a system of \( n \) equations in \( n + m \) variables as:

\[ f_1(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0 \]
\[ f_2(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0 \]
\[ \ldots \] ... ... ... ... ... ... (iv)

\[ f_n(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0. \]

The assumption that \( A_x \) is invertible means that the \( n \) by \( n \) matrix
evaluated at \((a, b)\) defines an invertible linear operator in \(\mathbb{R}^n\); in other words, its column vectors should be independent, or equivalently, its determinant should be \(\neq 0\). Also if (iv) holds when \(x = a\) and \(y = b\), then the conclusion of the theorem is that (iv) can be solved for \(x_1, \ldots, x_n\) in terms of \(y_1, \ldots, y_m\), for every \(y\) near \(b\), and that these solutions are continuously differentiable functions of \(y\).

**Proof.** Define \(F\) by

\[
F(x, y) = (f(x, y), y) \quad ((x, y) \in E) \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (v)
\]

Since \(f\) is a \(\zeta'\)-mapping of \(E \subset \mathbb{R}^{n+m}\) into \(\mathbb{R}^n\) and \(F\) is a mapping of \(E \subset \mathbb{R}^{n+m}\) into \(\mathbb{R}^{n+m}\), hence by (v) \(F\) is a \(\zeta'\)-mapping of \(E\) into \(\mathbb{R}^{n+m}\).

Since \(f(a, b) = 0\) for some \((a, b) \in E\) and \(A = f'(a, b)\), we then have

\[
F(a + h, b + k) = F(a + h, b + k) - F(a, b) = A(h, k) + r(h, k),
\]

where \(r\) is the remainder that occurs in the definition of \(f'(a, b)\). Now since by definition of \(F\),

\[
F(a + h, b + k) - F(a, b) = (f(a + h, b + k), b + k) - (f(a, b), b),
\]

\[
= (f(a + h, b + k), b + k) - (0, b) = (f(a + h, b + k), k)
\]

\[
= (A(h, k) + r(h, k), k + 0)
\]

\[
= (A(h, k), k) + (r(h, k), 0)
\]

It follows that \(F'(a, b)\) is the linear operator on \(\mathbb{R}^{n+m}\) that maps \((h, k)\) to \((A(h, k), k)\), i.e., \(F'(a, b) \in L(\mathbb{R}^{n+m})\) in the sense that \(F'(a, b):(h, k) \mapsto (A(h, k), k)\) for each \((h, k) \in \mathbb{R}^{n+m}\). If the image vector \((A(h, k), k) = 0 \in \mathbb{R}^n\), then \(A(h, k) = 0 \in \mathbb{R}^n\), and \(k = 0 \in \mathbb{R}^m\). Hence \(A(h, 0) = 0\), i.e., \(A, h = 0\) and the preceding theorem implies that \(h = 0\). Thus by properties of linear transformation \(F'(a, b)\) one-one and hence invertible. Therefore by inverse function theorem there exists open sets \(U\) and \(V\) in \(\mathbb{R}^{n+m}\) with \((a, b) \in U, (0, b) \in V\), such that \(F\) is one-one mapping of \(U\) onto \(V\). If we suppose \(W\) be the set of all \(y \in \mathbb{R}^m\), such that \((0, y) \in V\), i.e., \(W = \{y \in \mathbb{R}^m : (0, y) \in V\}\). We see that \(b \in W\), as \((0, b) \in V\), which is an open set, it follows that \(W\) is also open. If \(y \in W\), then \((0, y) = F(x, y)\) for some \((x, y) \in U\). Then by (v), we have \(f(x, y) = 0\) for this \(x\). We now prove that for the \(y \in W\) there exists a unique \(x\) such that \((x, y) \in U\) and \(f(x, y) = 0\). Suppose on the contrary that it is possible that for this \(y \in W\) there exists another \((x', y) \in U\) and \(f(x', y) = 0\), then \(F(x', y) = (f(x', y), y) = (0, y) = (f(x, y), y) = F(x, y)\), and since \(F\) is injective so \(x = x'\).

Further define \(g(y)\) for \(y \in W\) such that \((g(y), y) \in U\) and suppose that (ii) hold i.e., \((g(y), y) = 0\). Then \(F(g(y), y) = (f(g(y), y), y) = (0, y) \in W\). \(\ldots \quad \ldots \quad \ldots \quad \ldots \quad (vi)\)

Let \(G\) be a mapping from \(V\) onto \(U\) i.e., \(F^{-1} = G\). Since \(F\) is a \(\zeta'\)-mapping of an open set \(E\) into \(\mathbb{R}^{n+m}\), \(F'(a, b)\) is invertible for some \((a, b) \in E\), \((0, b) = (f(a, b), b) = F(a, b) (b \in W)\), keeping in view the equation (vi), and using the inverse function theorem, we have \(G \in \zeta'(E)\). Further by (vi), we have \((g(y), y) = F^{-1}(0, y) = G(0, y) \quad (y \in W)\). Since \(G \in \zeta'(E)\) gives \(g \in \zeta'(E)\). Now put \((g(y), y) = \phi(y)\), then

\[
\phi'(y) = \lim_{t \to 0} \frac{\phi(y + tk) - \phi(y)}{t} = \lim_{t \to 0} \frac{(g(y + tk), y + tk) - (g(y), y)}{t}
\]
\[
\lim_{t \to 0} \frac{g(y + tk) - g(y)}{tk} = \lim_{t \to 0} \frac{(g(y + tk) - g(y))}{tk} = (g'(y)k, k).
\]

Thus we have \( \phi'(y)k = (g'(y), k) \) (\( y \in W, k \in \mathbb{R}^m \)) and by (ii) \( f(\phi(y)) = f(g(y), y) \). Thus by chain rule, we have \( f'(\phi(y)) = \phi'(y) = 0 \). Also when \( y = b \), we have \( \phi(y) = (g(b), b) = (a, b) \) and \( f'(y) = f'(a, b) = A \). Then \( A\phi'(b) = 0 \) also \( A(h, k) = A_b + A_k, A_kg'(b)k + A_kk = A(g'(b)k, k) = A\phi'(b)k = 0 \), for all \( k \in \mathbb{R}^m \). Hence \( A_kg'(b) + A_y = 0 \) or \( g'(b) = -(A_x)^{-1}A_y \).

3.8. Differentiation of Integrals.

Suppose \( \phi \) is a function of two variables \( x \) and \( t \), which can be integrated with respect to \( x \) and which can be differentiated with respect to \( t \). There arise a natural question that, under what conditions will the result be the same if these two limit processes are carried out in the opposite order? More precisely: Under what conditions on \( \phi \) can one prove that the equation

\[
\frac{d}{dt} \int_a^b \phi(x, t)dx = \int_a^b \frac{\partial \phi(x, t)}{\partial t} dx \quad \ldots \quad \ldots \quad (i)
\]

is true? For the sake of convenience we use the notation \( \phi'(x) = \phi(x, t) \quad \ldots \quad \ldots \quad (ii) \)

Thus \( \phi' \) is, for each \( t \), a function of one variable.

3.8.1. Theorem. Suppose

(a) \( \phi(x, t) \) is defined for \( a \leq x \leq b, c \leq t \leq d \);
(b) \( \alpha \) is an increasing function on \( [a, b] \);
(c) \( \phi' \in \mathcal{R}(\alpha) \) for every \( t \in [c, d] \);
(d) \( c < s < d \), and to every \( \varepsilon > 0 \) corresponds a \( \delta > 0 \) such that

\[
|\phi'_{x}(t, s) - (D_2\phi)(x, s)| < \varepsilon \quad \text{for all } x \in [a, b] \text{ and for all } t \in (s - \delta, s + \delta).
\]

Define

\[
f(t) = \int_a^b \phi(x, t)d\alpha(x) \quad (c \leq t \leq d) \quad \ldots \quad \ldots \quad (iii)
\]

Then \( (D_2\phi)^s \in \mathcal{R}(\alpha) \), \( f^s(t) \) exists, and

\[
f^s(t) = \int_a^t (D_2\phi)(x, s) d\alpha(x). \quad \ldots \quad \ldots \quad \ldots \quad (iv)
\]

Note that (c) simply asserts the existence of the integrals (iii) for all \( t \in [c, d] \) and (d) certainly holds whenever \( D_2\phi \) is continuous on the rectangle on which \( \phi \) is defined.

Proof. For \( 0 < |t - s| < \delta \), consider the difference quotients

\[
\psi(x, t) = \frac{\phi(x, t) - \phi(x, s)}{t - s} \quad \ldots \quad \ldots \quad \ldots \quad (v)
\]

Now by mean value theorem there corresponds to each \( (x, t) \) there exists a number \( u \) between \( s \) and \( t \) such that

\[
\psi(x, t) = (D_2\phi)(x, u).
\]

Hence (d) implies that

\[
|\psi(x, t) - (D_2\phi)(x, s)| < \varepsilon \quad (a \leq x \leq b, 0 < |t - s| < \delta) \quad \ldots \quad \ldots \quad (vi)
\]

Note that by (v), we have

\[
\frac{f(t) - f(s)}{t - s} = \frac{1}{t - s} \left[ \int_a^b \phi(x, t)d\alpha(x) - \int_a^b \phi(x, s)d\alpha(x) \right] = \int_a^b \psi(x, t)d\alpha(x)
\]

\[
\ldots \quad \ldots \quad \ldots \quad (vii)
\]

By (vi), \( \psi^s \to (D_2\phi)^s \), uniformly on \([a, b]\), as \( t \to s \). Since each \( \psi^s \in \mathcal{R}(\alpha) \), it follows that \( (D_2\phi)^s \in \mathcal{R}(\alpha) \). Now from (vii), we have
\[f'(s) = \lim_{t \to s} \frac{f(t) - f(s)}{t - s} = \lim_{t \to s} \int_a^b \psi(x, t)d\alpha(x)
\]
\[= \int_a^b \lim_{t \to s} \psi(x, t)d\alpha(x) = \int_{\alpha} \int_a^b \psi(x, s)d\alpha(x) = \int_a^b \psi'd\alpha(x) \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \text{(viii)} \]

Since \(\psi^s \in \mathfrak{H}(\alpha)\) for each \(s \in [c, d]\), which implies that the integral on the RHS of (viii) exists and so \(f'(s)\) exists for all \(s \in [c, d]\). Further from (v), we have

\[\psi' = \psi(x, t) = \lim_{t \to s} \frac{\phi(x, t) - \phi(x, s)}{t - s} = (D_2\phi)(x, s). \]

Hence by (viii), we have

\[f'(s) = \int_a^b (D_2\phi)(x, s). \text{ This completes the proof.} \]

### 3.9 Unit summary

1. (a) A nonempty set \(X \subset \mathbb{R}^n\) is said to be a vector space if \(x + y \in X\) and \(cx \in X\) for all \(x \in X, y \in X\), and for all scalars \(c\), where \(x = (a_1, a_2, \ldots, a_n)\), \(y = (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n\) and \(a_1, a_2, \ldots, a_n; b_1, b_2, \ldots, b_n\) are scalars etc.

(b) If \(x_1, \ldots, x_k \in \mathbb{R}^n\) and \(c_1, \ldots, c_k\) are scalars, then the vector \(c_1x_1 + \ldots + c_kx_k\) is called a linear combination of \(x_1, \ldots, x_k\).

(c) The set of all linear combinations of elements of a set \(S\) called a linear span of \(S\) and denoted by \(L(S)\). Note that every \(L(S)\) is a vector space.

(d) A set \(S = \{x_1, \ldots, x_k\} \subset \mathbb{R}^n\) is said to be linearly independent if the relation \(c_1x_1 + \ldots + c_kx_k = 0\) implies that \(c_1 = \ldots = c_k = 0\). Otherwise \(S\) is said to be linearly dependent.

(e) Let \(B\) be a linearly independent subset of a vector space \(X\) such that \(L(B) = X\), then \(B\) is called a basis of \(X\).

2. If a vector space \(X\) is spanned by a set of \(r\) vectors, then \(\dim X \leq r\), where \(r\) is a positive integer and \(\dim \mathbb{R}^n = n\).

3. Suppose \(X\) is a vector space and \(\dim X = n\).
   (a) A set \(E\) of \(n\) vectors in \(X\) spans \(X\) if and only if \(E\) is linearly independent.
   (b) \(X\) has a basis and every basis consists of \(n\) vectors.
   (c) If \(1 \leq r \leq n\) and \(\{y_1, \ldots, y_r\}\) is an independent set in \(X\), then \(X\) has basis containing \(\{y_1, \ldots, y_r\}\).

4. A mapping \(A\) of a vector space \(X\) into a vector space \(Y\) is said to be a linear transformation if \(A(x_1 + x_2) = Ax_1 + Ax_2\) and \(A(cx) = cA(x)\) for all \(x, x_1, x_2 \in X\) and all scalars \(c\). Linear transformations of \(X\) into \(X\) are often called linear operators on \(X\). A linear operator on a finite dimensional vector space is one-one or onto then it is invertible. A linear operator \(A\) on a finite-dimensional vector space \(X\) is one-to-one if and only if the range of \(A\) is all of \(X\) [i.e., \(A(X) = X\) or \(A\) is onto].

5. Let \(X\) and \(Y\) be vector spaces then \(L(X, Y)\) denote the set of all linear transformations of \(X\) into \(Y\). The set of all linear transformations of \(X\) into itself is denoted by \(L(X, X)\) or \(L(X)\).

6. For \(A \in L(\mathbb{R}^n, \mathbb{R}^m)\), the norm \(\|A\|\) of \(A\) is defined by \(\|A\| = \sup\{|Ax| : x \in \mathbb{R}^n, |x| \leq 1\}\). If \(A \in L(\mathbb{R}^n, \mathbb{R}^m)\), then \(\|A\| < \infty\) and \(A\) is a uniformly continuous mapping of \(\mathbb{R}^n\) into \(\mathbb{R}^m\). If \(A, B \in L(\mathbb{R}^n, \mathbb{R}^m)\), and \(c\) is a scalar, then \(\|A + B\| \leq \|A\| + \|B\|, \quad \|cA\| = |c|\|A\|\). If \(A \in L(\mathbb{R}^n, \mathbb{R}^m)\) and \(B \in L(\mathbb{R}^m, \mathbb{R}^k)\) then \(\|BA\| \leq \|B\|\|A\|\).

7. Suppose \(E\) is a nonempty open set in \(\mathbb{R}^n\), \(f\) maps \(E\) into \(\mathbb{R}^m\) and \(x \in E\). If there exists a linear transformation \(A\) of \(\mathbb{R}^n\) into \(\mathbb{R}^m\) such that
\[ \lim_{h \to 0} \frac{|f(x + h) - f(x) - Ah|}{|h|} = 0 \quad \ldots \quad \ldots \quad \ldots \] (ii)

then \( f \) is said to be differentiable at \( x \), and we write \( f'(x) = A \). If \( f \) is differentiable at every \( x \in E \), we say that \( f \) is differentiable in \( E \).

8. **Chain rule.** Suppose \( E \) is an open set in \( \mathbb{R}^n \), \( f \) maps \( E \) into \( \mathbb{R}^m \), \( f \) is differentiable at \( x_0 \in E \), \( g \) maps an open set containing \( f(E) \) into \( \mathbb{R}^k \), and \( g \) is differentiable at \( f(x_0) \). Then the mapping \( F \) of \( E \) into \( \mathbb{R}^k \), defined by \( F(x) = g(f(x)) \), is differentiable at \( x_0 \) and \( F'(x_0) = g'(f(x_0))f'(x_0) \).

9. Let \( f \) be a function from an open set \( E \subset \mathbb{R}^n \) and for \( x \in E \), \( 1 \leq i \leq m, 1 \leq j \leq n \), we define

\[ D_j f_i(x) = \lim_{t \to 0} \frac{f_i(x + te_j) - f_i(x)}{t} \]

provided the limit exists. Write \( f_i(x) = f_i(x_1, \ldots, x_n) \), we see that \( D_j f_i \) is the derivative of \( f_i \) with respect to \( x_j \), keeping the other variables \( x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n \) fixed. The notation \( \partial f_i / \partial x_j \) is often used in place of \( D_j f_i \); it is called a partial derivative of \( f_i \) with respect to \( x_j \).

10. **Continuously differentiable function.** A differentiable mapping \( f \) of an open set \( E \subset \mathbb{R}^n \) into \( \mathbb{R}^m \) is said to be continuously differentiable in \( E \) if \( f' \) is a continuous mapping of \( E \) into \( L(\mathbb{R}^n, \mathbb{R}^m) \).

11. **Taylor’s Theorem.** Let \( f : E \subset \mathbb{R}^n \to \mathbb{R} \) be a function, where \( E \) is an open subset \( \mathbb{R}^n \). Suppose \( f \in \zeta^{(m)} \), and \( a \in E \) and \( x \in \mathbb{R}^n \), further suppose that \( L(a, a + x) \subset E \). Then there exist \( \theta \in (0, 1) \) such that

\[ f(a + x) = f(a) + \frac{(D_1 f)(a)}{1!} + \frac{(D_2^2 f)(a)}{2!} + \ldots + \frac{(D_n^m f)(a)}{m!}, \]

where \( D_x = x \cdot \nabla \).

12. **The Inverse Function Theorem.** Suppose \( f \) is a \( \zeta \)-mapping of an open set \( E \subset \mathbb{R}^n \) into \( \mathbb{R}^n \), \( f'(a) \) is invertible for some \( a \in E \) and \( b = f(a) \). Then (a) there exist open sets \( U \) and \( V \) in \( \mathbb{R}^n \) such that \( a \in U, b \in V \), \( f \) is one-to-one on \( U \), and \( f(U) = V \). (b) If \( g \) is the inverse of \( f \), defined in \( V \) by \( g(f(x)) = x \) (\( x \in U \)), then \( g \in \zeta(V) \). 13. **The Implicit Function Theorem.** (linear version). If \( A \in L(\mathbb{R}^{m+n}, \mathbb{R}^n) \) and if \( A_x \) is invertible, then there corresponds to every \( k \in \mathbb{R}^m \) a unique \( h \in \mathbb{R}^n \) such that \( A(h, k) = 0 \), and \( h \) can be computed from \( k \) by the formula \( h = -(A_x)^{-1}A_k \).

(non-linear version). Let \( f \) be a \( \zeta \)-mapping of an open set \( E \subset \mathbb{R}^{m+n} \) into \( \mathbb{R}^n \), such that \( f(a, b) = 0 \) for some point \( (a, b) \in E \). Put \( A = f'(a, b) \) and assume that \( A_x \) is invertible, then there exist open sets \( U \subset \mathbb{R}^{m+n} \) and \( W \subset \mathbb{R}^m \), with \( (a, b) \in U \) and \( b \in W \), having the following property: To every \( y \in W \) corresponds a unique \( x \) such that \( (x, y) \in U \) and \( f(x, y) = 0 \). \ldots \ldots \ldots (i)

If this \( x \) is defined to be \( g(y) \), then \( g \) is a \( \zeta \)-mapping of \( W \) into \( \mathbb{R}^n \), \( g(b) = a \), \( g(y) \in U \). \ldots \ldots \ldots (ii)

and \( g(b) = -(A_x)^{-1}A_y \). \ldots \ldots \ldots (iii)

3.10 Check your progress/ assignment

1. If \( S \) is a nonempty subset of a vector space \( X \), prove that the span of \( S \) is a vector space.

2. Prove that \( BA \) is linear if \( A \) and \( B \) are linear transformations. Prove also that \( A^{-1} \) is linear and invertible.
3. Assume \( A \in L(X, Y) \) and \( Ax = 0 \) only when \( x = 0 \). Prove that \( A \) is then 1-1.
4. Prove that null spaces and ranges of linear transformations are vector spaces.
5. Prove that to every \( A \in L(R^n, R^1) \) corresponds a unique \( y \in R^n \) such that \( Ax = xy \).
   Prove also that \( ||A|| = |y| \).
6. If \( f(0, 0) = 0 \) and
   \[ f(x, y) = \frac{xy}{x^2 + y^2}, \] if \((x, y) \neq (0, 0)\).
   Prove that \((D_1 f)(x, y)\) and \((D_2 f)(x, y)\) exist at every point of \( R^2 \), although \( f \) is not continuous at \((0, 0)\).
7. Suppose that \( f \) is a real-valued function defined in an open set \( E \subset R^n \), and that the partial derivatives \( D_1 f, \ldots, D_n f \) are bounded in \( E \). Prove that \( f \) is continuous in \( E \).
8. Suppose that \( f \) is a differentiable real function in an open set \( E \subset R^n \), and that \( f \) has a local maximum at a point \( x \in E \). Prove that \( f'(x) = 0 \).
9. If \( f \) is a differentiable mapping of a connected open set \( E \subset R^n \) into \( R^m \), and if \( f'(x) = 0 \) for every \( x \in E \), prove that \( f \) is constant in \( E \).
10. If \( f \) is a real function defined in a convex open set \( E \subset R^n \), such that \( D_1 f(x) = 0 \) for every \( x \in E \), prove that \( f(x) \) depends only on \( x_2, \ldots, x_n \).
11. Suppose \( f \) is a differentiable mapping of \( R^1 \) into \( R^3 \) such that \( |f(t)| = 1 \) for every \( t \). Prove that \( f'(t), f(t) = 0 \).
12. Define \( f(0, 0) = 0 \) and \( f(x, y) = \frac{x^2}{x^2 + y^2} \) if \((x, y) = (0, 0)\).
   (a) Prove that \( D_1 f \) and \( D_2 f \) are bounded functions in \( R^1 \). (Hence \( f \) is continuous.)
   (b) Let \( u \) be any unit vector in \( R^2 \). Show that the directional derivative \( (D_uf)(0, 0) \) exists, and that its absolute value is at most 1.
   (c) Let \( \gamma \) be a differentiable mapping of \( R^1 \) into \( R^2 \) with \( y(0) = (0, 0) \) and \( |y'(0)| > 0 \).
      Put \( g(t) = f(\gamma(t)) \) and prove that \( g \) is differentiable for every \( t \in R^1 \). If \( \gamma \in \zeta \) prove that \( \gamma \in \zeta \).
   (d) In spite of this, prove that \( f \) is not differentiable at \( (0, 0) \).
13. Show that the continuity of \( f' \) at the point \( a \) is needed in the inverse function theorem, even in the case \( n = 1 \): If \( f(t) = t + 2t^2 \sin \left( \frac{1}{t} \right) \) for \( t \neq 0 \), and \( f(0) = 0 \), then \( f'(0) = 1 \), \( f' \) is bounded in \((-1, 1)\), but \( f \) is not one-to-one in any neighborhood of \( 0 \).
14. Show that the system of equations \( 3x + y - z + u^2 = 0, x - y + 2z + u = 0, 2x + 2y - 3z + 2u = 0, \) can be solved for \( x, y, u \) in terms of \( z \); for \( x, z, u \) in terms of \( y \); for \( y, z, u \) in terms of \( x \); but not for \( x, y, z \) in terms of \( u \).
15. Take \( n = m = 1 \) and prove the implicit function theorem.
16. Define \( f \) in \( R^2 \) by \( f(x, y_1, y_2) = x^2 y_1 + e^x + y_2 \), show that \( f(0, 1, -1) = 0, (D_1 f)(0, 1, -1) \neq 0, \) and that there exists therefore a differentiable function \( g \) in some neighborhood of \((1, -1) \) in \( R^2 \), such that \( g(1, -1) = 0 \) and \( f(g(y_1, y_2), y_1, y_2) = 0 \). Find \((D_1 g)(1, -1)\) and \((D_2 g)(1, -1)\).
16. Put \( f(0, 0) = 0 \) and \( f(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2} \) if \((x, y) \neq (0, 0)\).
   (a) \( f, D_1 f, D_2 f \) are continuous in \( R^2 \):
(b) $D_1f$ and $D_2f$ exists at every point of $\mathbb{R}^2$ and are continuous except at $(0, 0)$; 
(c) $(D_1f)(0, 0) = 1$ and $(D_2f)(0, 0) = -1$.

17. For $t \geq 0$, put 
$$
\phi(x,t) = \begin{cases} 
    x, & 0 \leq x \leq \sqrt{t} \\
    -x + 2\sqrt{t}, & \sqrt{t} \leq x \leq 2\sqrt{t} \\
    0, & \text{otherwise}
\end{cases}
$$
and put $\phi(x, t) = -\phi(x, |t|)$ if $t < 0$.

Show that $\phi$ is continuous on $\mathbb{R}^2$, and $(D_2\phi)(x, 0) = 0$ for all $x$. Define 
$f(t) = \int_{-1}^{1} \phi(x,t) \, dx$. Show that $f(t) = t$ if $|t| < \frac{1}{4}$. Hence $f'(0) \neq \int_{-1}^{1} (D_2\phi)(x,0) \, dx$.

### 3.11 Points for discussion and clarification

A ____________________________________________________________

B ____________________________________________________________

C ____________________________________________________________

D ____________________________________________________________

### 3.12 Suggested study material

5. H.K. Pathak - Real Analysis
6. Lecture Notes on Real analysis Richard F. Bass
Lebesgue outer measure.

Introduction: The Riemann integral, dealt with in calculus courses, is well suited for computations but less suited for dealing with limit processes. In this course we will introduce the so called Lebesgue integral, which keeps the advantages of the Riemann integral and eliminates its drawbacks. At the same time we will develop a general measure theory which serves as the basis of contemporary analysis and probability.

Objectives: Main objective of this unit is to understand about Lebesgue outer measure. Measurable sets, Regularity, Measurable functions, Borel and Lebesgue measurability, Non-measurable sets, Integration of Non-negative functions, The General integral, Integration of Series. Reimann and Lebesgue Integrals, The Four derivatives, Functions of Bounded variation, Lebesgue Differentiation Theorem, Differentiation and Integration, Before this we have to understand some basic ideas about algebra.

4.1 Lebesgue outer measure

Measurable Sets

4.1.1 Definition: A Set E is said to be measurable if for each set A we have

\[ m^*(A) = m^*(A \cap E) + m^*(A \cap E^c) \]

Regularity

4.1.2 Definition: A function \( n \) on the collection of all subsets satisfying

(a) \( n(\emptyset) = 0 \);
(b) if \( A \subset B \), then \( n(A) \leq n(B) \);
(c) \( n(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} n(A_i) \)

is called an outer measure.

4.1.3 Proposition: \( m^* \) defined by \( m^*(A) = \inf \{ \sum_{i=1}^{\infty} (b_i - a_i) : A \subset \bigcup_{i=1}^{\infty} (a_i, b_i) \} \) is an outer measure.

Proof: (a) and (b) are obvious.
To prove (c), let \( \varepsilon > 0 \). For each \( i \) there exist intervals \( I_{i1}, I_{i2}, \ldots \), each of the form \( (a_{ij}, b_{ij}] \), such that \( A_i \subset \bigcup_{j=1}^{\infty} I_{ij} \) and

\[ \sum_{i,j} m(I_{ij}) \leq \sum_i m^*(A_i) + \sum_i \varepsilon/2^i = \sum_i m^*(A_i) + \varepsilon \]

Since \( \varepsilon \) is arbitrary,

\[ m^* (\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} m^*(A_i) \]

4.1.4 Definition: Let \( m^* \) be an outer measure. A set \( A \subset X \) is \( m^* \)-measurable if

\[ m^*(E) = m^*(E \cap A) + m^*(E \cap A^c) \]  

(4.1)

for all \( E \subset X \).

4.1.5 Theorem: If \( m^* \) is an outer measure on \( X \), then the collection \( A \) of \( m^* \) measurable sets is a \( \sigma \)-algebra and the restriction of \( m^* \) to \( A \) is a measure. Moreover, \( A \) contains all the null sets.
**Proof:** By Proposition 4.1.3,

\[ m^* (E) \leq m^* (E \cap A) + m^* (E \cap A^c) \] for all \( E \subseteq X \).

So to check (1) it is enough to show

\[ m^* (E) \geq m^* (E \cap A) + m^* (E \cap A^c). \]

This will be trivial in the case \( m^* (E) = \infty \).

If \( A \in A \), then \( A^c \in A \) by symmetry and the definition of \( A \).

Suppose \( A, B \in A \) and \( E \subseteq X \). Then

\[ m^* (E) = m^* (E \cap A) + m^* (E \cap A^c) \]

\[ = (m^* (E \cap A \cap B) + m^* (E \cap A \cap B^c)) + (m^* (E \cap A^c \cap B) + m^* (E \cap A^c \cap B^c)). \]

The first three terms on the right have a sum greater than or equal to \( m^* (E \cap (A \cup B)) \) because \( A \cup B \subseteq (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B) \). Therefore

\[ m^* (E) \geq m^* (E \cap (A \cup B)) + m^* (E \cap (A \cup B)^c), \]

which shows \( A \cup B \in A \). Therefore \( A \) is algebra.

Let \( A_i \) be disjoint sets in \( A \), let \( B_n = \bigcup_{i=1}^{n} A_i \), and \( B = \bigcup_{i=1}^{\infty} A_i \). If \( E \subseteq X \),

\[ m^* (E \cap B_n) = m^* (E \cap B_n \cap A_n) + m^* (E \cap B_n \cap B_n^c) \]

\[ = m^* (E \cap A_n) + m^* (E \cap B_{n-1}). \]

Repeating for \( m^* (E \cap B_{n-1}) \), we obtain \( m^* (E \cap B_n) \geq \sum_{i=1}^{n} m^* (E \cap A_i) \).

Since \( B_n \in A \), then

\[ m^* (E) = m^* (E \cap B_n) + m^* (E \cap B_n^c) \]

\[ \geq \sum_{i=1}^{n} m^* (E \cap A_i) + m^* (E \cap B_n^c). \]

Let \( n \to \infty \). Recalling that \( m^* \) is an outer measure,

\[ m^* (E) \geq \sum_{i=1}^{n} m^* (E \cap A_i) + m^* (E \cap B_n^c). \]

\[ \geq m^* \bigcup_{i=1}^{n} (E \cap A_i) + m^* (E \cap B_n^c). \]

\[ = m^* (E \cap B) + m^* (E \cap B_n^c). \]

\[ \geq m^* (E). \]

This shows \( B \in A \).

If we set \( E = B \) in this last equation, we obtain

\[ m^* (B) = \sum_{i=1}^{\infty} m^* (A_i), \]

or \( m^* \) is countably additive on \( A \).

If \( m^* (A) = 0 \) and \( E \subseteq X \), then

\[ m^* (E \cap A) + m^* (E \cap A^c) = m^* (E \cap A^c) \leq m^* (E), \]

which shows \( A \) contains all null sets.

**Measurable functions:**

Suppose we have a set \( X \) together with a \( \sigma \)-algebra \( A \).
4.1.6 **Definition:** \( f : X \to \mathbb{R} \) is measurable or \( A \)-measurable if \( \{ x : f(x) > a \} \in A \) for all \( a \in \mathbb{R} \).

**Example 1** Suppose \( f \) is identically constant. Then \( \{ x : f(x) > a \} \) is either empty or the whole space, so \( f \) is measurable.

**Example 2** Suppose \( f(x) = 1 \) if \( x \in A \) and 0 otherwise. Then \( \{ x : f(x) > a \} \) is either \( \emptyset \), \( A \), or \( X \). So \( f \) is measurable if and only if \( A \) is in the \( \sigma \)-algebra.

**Example 3** Suppose \( X \) is the real line with the Borel \( \sigma \)-algebra and \( f(x) = x \). Then \( \{ x : f(x) > a \} = (a, \infty) \), and so \( f \) is measurable.

4.1.7 **Proposition:** The following are equivalent.

(a) \( \{ x : f(x) > a \} \in A \) for all \( a \);
(b) \( \{ x : f(x) \leq a \} \in A \) for all \( a \);
(c) \( \{ x : f(x) < a \} \in A \) for all \( a \);
(d) \( \{ x : f(x) \geq a \} \in A \) for all \( a \).

**Proof.** The equivalence of (a) and (b) and of (c) and (d) follow from taking complements. The remaining equivalences follow from the equations

\[
\{ x : f(x) \geq a \} = \bigcap_{n=1}^{\infty} \{ x : f(x) > a - 1/n \},
\]

\[
\{ x : f(x) > a \} = \bigcup_{n=1}^{\infty} \{ x : f(x) \geq a + 1/n \}.
\]

4.1.8 **Proposition:** If \( f \) and \( g \) are measurable, so are \( f + g \), \( cf \), \( fg \), \( \max(f, g) \), and \( \min(f, g) \).

**Proof:** If \( f(x) + g(x) < \alpha \), then \( f(x) < \alpha - g(x) \), and there exists a rational \( r \) such that \( f(x) < r < \alpha - g(x) \).

So

\[
\{ x : f(x) + g(x) < \alpha \} = \bigcup_{r=\text{rational}} \{ x : f(x) < r \} \cap \{ x : g(x) < \alpha - r \}.
\]

\( f^2 \) is measurable since

\[
\{ x : f(x)^2 > a \} = \{ x : f(x) > \sqrt{a} \} \cup \{ x : f(x) < -\sqrt{a} \}.
\]

The measurability of \( fg \) follows since

\[
f, g = \frac{1}{2} [(f + g)^2 - f^2 - g^2]. \{ x : \max(f(x), g(x)) > a \}
\]

\[
= \{ x : f(x) > a \} \cup \{ x : g(x) > a \}, \text{ and the argument for } \min(f, g) \text{ is similar.}
\]

4.1.9 **Proposition:** If \( f_i \) is measurable for each \( i \), then so are \( \sup_{i \to \infty} f_i \), \( \inf_{i \to \infty} f_i \), \( \limsup_{i \to \infty} f_i \), and \( \liminf_{i \to \infty} f_i \).

**Proof:** By using the definitions. We have

\[
\{ x : \sup_i f_i > a \} = \bigcap_{n=1}^{\infty} \{ x : f_i(x) > a \}
\]

We have the result for the sup and inf.

The result for \( \limsup_{i \to \infty} f_i \) and \( \liminf_{i \to \infty} f_i \) follows once.

The proof for \( \inf f_i \) is similar.

4.1.10 **Definition:** We say \( f = g \) almost everywhere, written \( f = g \ a.e. \), if \( \{ x : f(x) = g(x) \} \) has measure zero. Similarly, we say \( f_i \to f \ a.e. \), if the set of \( x \) where this fails has measure zero.

We saw, if \( X \) is a metric space, \( A \) contains all the open sets, and \( f \) is continuous, then \( f \) is measurable, hence all continuous functions are Borel measurable. The same is true for monotone functions on the real line.

4.1.11 **Proposition:** If \( f : \mathbb{R} \to \mathbb{R} \) is nondecreasing or nonincreasing, then \( f \) is Borel measurable.
Proof: Let us suppose $f$ is nondecreasing. The set $A_a = \{x : f(x) > a\}$ is then either a semi-infinite open interval or semi-infinite closed interval. This can be seen by a picture. To be more careful, given $a \in \mathbb{R}$, let $x_0 = \sup\{y : f(y) \leq a\}$. If $f(x_0) = a$, then $A_a = (x_0, \infty)$, while if $f(x_0) \neq a$, then $A_a = [x_0, \infty)$. In each case $A_a$ is a Borel set.

4.1.12 Proposition: Let $X$ be a space, $\mathcal{A}$ a $\sigma$-algebra on $X$, and $f : X \rightarrow \mathbb{R}$ a $\mathcal{A}$-measurable function. If $\mathcal{A}$ is in the Borel $\sigma$-algebra on $\mathbb{R}$, then $f^\leftarrow(A) \in \mathcal{A}$.

Proof: Let $B$ be the Borel $\sigma$-algebra on $\mathbb{R}$ and $C = \{A \in B : f^\leftarrow(A) \in \mathcal{A}\}$.

If $A_1, A_2, \ldots \in C$, then since $f^\leftarrow\left(\bigcup_i A_i\right) = \bigcup_i f^\leftarrow(A_i) \in \mathcal{A}$, we have that $C$ is closed under countable unions.

Similarly $C$ is closed under countable intersections and complements, so $C$ is a $\sigma$-algebra. Since $f$ is measurable, $C$ contains $(a, \infty)$ for every real $a$, hence $C$ contains the $\sigma$-algebra generated by these intervals, that is, $C$ contains $B$.

Example 4 We want to construct a set that is Lebesgue measurable, but not Borel measurable. Let $F$ be the Cantor-Lebesgue function of Example 4.1 and define $f(x) = \inf\{y : F(y) \geq x\}$.

Although $f$ is not continuous, observe that $f$ is strictly increasing (hence one-to-one) and maps $[0, 1]$ into $C$, the Cantor set.

Since $f$ is nondecreasing, $f^\leftarrow$ maps Borel measurable sets to Borel measurable sets.

Let $A$ be the non-measurable set we constructed in Proposition 4.1.18. Let $B = f(A)$. Since $f(A) \subseteq C$ and $m(C) = 0$, then $f(A)$ is a null set, hence is Lebesgue measurable.

On the other hand, $f(A)$ is not Borel measurable, because if it were, then $A = f^\leftarrow(f(A))$ would be Borel measurable, a contradiction.

Borel and Lebesgue Measurability

Let $X$ is a set. We will use the notation: $A^c = \{x \in X : x \not\in A\}$ and $A - B = A \cap B^c$. (The notation $A / B$ is also commonly used.)

4.1.13 Definition: ($\sigma$-algebra) An algebra is a collection $\mathcal{A}$ of subsets of $X$ such that

(a) $\emptyset, X \in \mathcal{A}$;
(b) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$;
(c) if $A_1, \ldots, A_n \in \mathcal{A}$, then $\bigcup_{i=1}^n A_i$ and $\bigcap_{i=1}^n A_i$ are in $\mathcal{A}$.

$\mathcal{A}$ is a $\sigma$-algebra (or $\sigma$-field) if in addition
(d) if $A_1, A_2, \ldots$ are in $\mathcal{A}$, then $\bigcup_{i=1}^n A_i$ and $\bigcap_{i=1}^n A_i$ are in $\mathcal{A}$.

In (d) we allow countable unions and intersections only; we do not allow uncountable unions and intersections.

Example 5 Let $X = \mathbb{R}$ and $\mathcal{A}$ be the collection of all subsets of $\mathbb{R}$.

Example 6 Let $X = \mathbb{R}$ and let $\mathcal{A} = \{A \subseteq \mathbb{R} : A$ is countable or $A^c$ is countable$\}$. 
Parts (a) and (b) of the definition are easy as $\emptyset$, $X = \mathbb{R}$ are elements of power set of $X$ and $A^c$ is the element of $A$. to prove (c) - Suppose $A_1, A_2, \ldots$ are all in $A$. If each of the $A_i$ are countable, then $\bigcup A_i$ is countable, and so in $A$. If $A_{i_0}$ is countable for some $i_0$, then
\[
\bigcap_{i} A_i = \bigcap_{i=1}^{i_0} A_i^c \subset \bigcap_{i} A_i^c
\]
is countable, and again $\bigcup A_i$ is in $A$. Since $\bigcap A_i = \left( \bigcap_{i} A_i^c \right)^c$, then the countable intersection of sets in $A$ is again in $A$.

**Example 7** Let $X = [0, 1]$ and, $A = \{\emptyset, X, [0, \frac{1}{2}], [\frac{1}{2}, 1]\}$.

**Example 8** $X = \{1, 2, 3\}$ and $A = \{X, \emptyset, \{1\}, \{2, 3\}\}$.

**Example 9** Let $X = [0, 1]$, and $B_1, \ldots, B_8$ subsets of $X$ which are pair wise disjoint and whose union is all of $X$. Let $A$ be the collection of all finite unions of the $B_i$'s as well as the empty set. (So $A$ consists of $2^8$ elements.)

**Remark:** If we take an intersection of $\sigma$-algebras, we get a $\sigma$-algebra; this is just a matter of checking the definition.

If we have a collection $C$ of subsets of $X$, there is at least one $\sigma$-algebra containing $C$, namely, the one consisting of all subsets of $X$. We can take the intersection of all $\sigma$-algebras that contain $C$; we denote this intersection by $\sigma(C)$.

If $A$ is any $\sigma$-algebra containing $C$, then $A \supset \sigma(C)$.

If $X$ has some additional structure, say, it is a metric space, then we can talk about open sets.

If $G$ is the collection of open subsets of $X$, then we call $\sigma(G)$ the Borel $\sigma$-algebra on $X$, and this is often denoted $B$. We will see later that when $X$ is the real line, that $B$ is not equal to the collection of all subsets of $X$.

We end this section with the following proposition.

**4.1.14 Proposition:** If $X = \mathbb{R}$, then the Borel $\sigma$-algebra is generated by each of the following collection of sets:

1. $C_1 = \{(a, b) : a, b \in \mathbb{R}\}$.
2. $C_2 = \{[a, b] : a, b \in \mathbb{R}\}$.
3. $C_3 = \{(a, b) : a, b \in \mathbb{R}\}$.
4. $C_4 = \{(a, \infty) : a \in \mathbb{R}\}$.

**Proof.** (1) Let $G$ be the collection of open sets. Then $C_1 \subset C \subset \sigma(G)$. Now $\sigma(G)$ is the Borel $\sigma$-algebra and contains $C_1$. Since $\sigma(C_1)$ is the intersection of all $\sigma$-algebras containing $C_1$, then $\sigma(C_1) \subset \sigma(G)$.

To get the reverse inclusion, if $G$ is open, it is the countable union of open intervals. So $G \in \sigma(C_1)$, and hence $G \subset \sigma(C_1)$. $\sigma(G)$ is the intersection of all $\sigma$-algebras containing $G$; $\sigma(C_1)$ is one such, so $\sigma(G) \subset \sigma(C_1)$.

(2) If $[a, b] \in C_2$, then $[a, b] = \bigcap_{i=1}^{n} \left[a - \frac{1}{n}, b + \frac{1}{n}\right] \in \sigma(G)$. So $C_2 \subset \sigma(G)$. 
and by an argument similar to that in (1), we conclude $\sigma(C_2) \subseteq \sigma(G)$.

If $(a, b) \in C_1$, choose $n_0 \geq 2/(b - a)$ and note $\bigcup_{n=0}^{\infty} \left[ a + \frac{1}{n}, b - \frac{1}{n} \right] \in \sigma(c_1)$
So the Borel $\sigma$-algebra, which is equal to $\sigma(C_1)$ by part (1), is contained in $\sigma(C_2)$.

(3) The proof here is similar to (2), using $(a, b) = \bigcap_{n=1}^{\infty} \left( a, b + \frac{1}{n} \right)$ and

$$(a, b) = \bigcup_{n=n_0}^{\infty} \left[ a, b - \frac{1}{n} \right],$$
provided $n_0$ is taken large enough.

(4) The proof of this comes from (3), using that

$$(a, \infty) = \bigcup_{n=1}^{\infty} (a, a + n].$$

### 4.1.15 Definition: A measure on $(X, A)$ is a function $\mu : A \rightarrow [0, \infty]$ such that

(a) $\mu(A) \geq 0$ for all $A \in A$;

(b) $\mu(\emptyset) = 0$;

(c) if $A_i \in A$ are disjoint, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

**Example 10** $X$ is any set, $A$ is the collection of all subsets, and $\mu(A)$ is the number of elements in $A$. This is called counting measure.

**Example 11** $X = \mathbb{R}$, $A$ the collection of all subsets, $x_1, x_2, \ldots \in \mathbb{R}$, and $a_1, a_2, \ldots > 0$. Set $\mu(A) = \sum_{x \in A} a_i$. A particular case of this is if $x_i = i$ and all the $a_i = 1$. We will see later that this allows us to view infinite series as functions on this space.

**Example 12** $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. This measure is called point mass at $x$. We will construct Lebesgue measure on $\mathbb{R}$; this is an extension of the notion of length. However, the construction is a bit lengthy. We will also construct Lebesgue measure on $\mathbb{R}^n$; when $n = 2$, this is an extension of the notion of area, when $n = 3$, of volume.

### 4.1.16 Proposition: The following hold:

(a) If $A, B \in A$ with $A \subset B$, then $\mu(A) \leq \mu(B)$.

(b) If $A_i \in A$, then $\mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i)$.

(c) If $A_i \in A$, $A_1 \subset A_2 \subset \cdots$, and $A = \bigcup_{i=1}^{\infty} A_i$, then $\mu(A) = \lim_{n \to \infty} \mu(A_n)$.

(d) $A_1 \in A$, $A_1 \supset A_2 \supset \cdots$, $\mu(A_1)$ and $A = \bigcap_{i=1}^{\infty} A_i$, then $\mu(A) = \lim_{n \to \infty} \mu(A_n)$.

**Proof.** (a) Let $A_1 = A$, $A_2 = B - A$, and $A_3 = A_4 = \cdots = \emptyset$. Now from part (c) of the definition of measure

if $A_i \in A$ are disjoint, then $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$. Hence $\mu(A) \leq \mu(B)$ as $A_1 \subset A_2$. 
(b) Let $B_1 = A_1$, $B_2 = A_2 - B_1$, $B_3 = A_3 - (B_1 \cup B_2)$, and so on. The $B_i$ are disjoint and $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$. So $\mu(A) = \sum_{i=1}^{\infty} \mu(B_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$. 

(c) Define the $B_i$ as in (b). Since $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$, then $\mu(A) = \mu(\bigcup_{i=1}^{\infty} A_i) = \mu(\bigcup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i)$.

(d) Apply (c) to the sets $A_1 - A_i$, $i = 1, 2, \ldots$. 

Example 2.6: To see that $\mu(A_1) < \infty$ is necessary, let $X$ be the positive integers, $\mu$ counting measure, and $A_i = \{i, i + 1, \ldots\}$. Then the $A_i$ decrease, $\mu(A_i) = \infty$ for all $i$, but $\mu(\bigcap_{i=1}^{\infty} A_i) = \mu(\emptyset) = 0$.

4.1.17 Definition: A probability or probability measure is a measure such that $\mu(X) = 1$. In this case we usually write $(\Omega, F, \mu)$ instead of $(X, A, \mu)$.

Define $m((a, b)) = b - a$. If $G$ is an open set and $G \subset \mathbb{R}$, then $G = \bigcup_{i=1}^{\infty} (a_i, b_i)$ with the intervals disjoint. Define $m(G) = \sum_{i=1}^{\infty} (b_i - a_i)$. If $A \subset \mathbb{R}$, define

$$m^*(A) = \inf\{m(G) : G \text{ open}, A \subset G\}. \quad (4.2)$$

We will show the following.

(A) $m^*$ is not a measure on the collection of all subsets of $\mathbb{R}$.

(B) $m^*$ is a measure on a strictly smaller $\sigma$-algebra that strictly contains the Borel $\sigma$-algebra.

We will prove these two facts (and a bit more) in a moment, but let’s first make some remarks.

A set $N$ is a null set with respect to $m^*$ if $m^*(N) = 0$. Let $\mathcal{B}$ be the smallest $\sigma$-algebra containing $\mathcal{B}$ and all the null sets. More precisely, let $\mathcal{N}$ be the collection of all sets that are null sets with respect to $m^*$ and let $\mathcal{L} = \sigma(\mathcal{B} \cup \mathcal{N})$. $\mathcal{L}$ is called the Lebesgue $\sigma$-algebra, and sets in $\mathcal{L}$ are called Lebesgue measurable.

As part of our proof of (B) we will show that $m^*$ is a measure on $\mathcal{L}$. Lebesgue measure is the measure $m^*$ on $\mathcal{L}$. (A) shows that $\mathcal{L}$ is strictly smaller than the collection of all subsets of $\mathbb{R}$. It is easy to get lost in the construction of Lebesgue measure, so let us summarize our steps.

Proof (A): given in Proposition 4.1.18.

We then turn to the construction of Lebesgue measure. It is more convenient for technical reasons to define

$$m^*(A) = \inf\{\sum_{i=1}^{\infty} (b_i - a_i) : A \subset \bigcup_{i=1}^{\infty} (a_i, b_i)\}. \quad (4.3)$$

There is no real difference between this and (3.1) since it is clear that the difference between an open set and a set of the form $\bigcup_{i=1}^{\infty} (a_i, b_i]$ is countable, and with either definition of $m^*$, the measure of a point is 0, so the measure of a set consisting of countably many points is 0. However, when we talk about Lebesgue-Stieltjes measure, then there is a real difference.

We define what it means to be an outer measure (Definition 4.1.2) and prove that $m^*$ is
an outer measure (Proposition 4.1.3). We then define what it means for a set to be $m^*$-
measurable (Definition 4.1.4) and prove that the collection of $m^*$-measurable sets is an $\sigma$-algebra
and that $m^*$ restricted to this $\sigma$-algebra is a measure.

This looks promising, but we do not yet know that enough sets are $m^*$-
measurable. That
takes one more step. We show below in Proposition 4.1.19 that the collection of $m^*$-
measurable
sets contains the Borel $\sigma$-algebra.

4.1.18 Proposition: $m^*$ is not a measure on the collection of all subsets of $\mathbb{R}$.

Proof: Suppose $m^*$ is a measure. Define $x \sim y$ if $x - y$ is rational. This is an equivalence
relationship on $[0, 1]$. For each equivalence class, pick an element out of that class (by the axiom
of choice) Call the collection of such points $A$. Given a set $B$,

Define $B + x = \{y + x : y \in B\}$. Note $m^*(A + q) = m^*(A)$ since this translation invariance holds
for intervals, hence for open sets, hence for all sets.
Moreover, the sets $A + q$ are disjoint for different rationales $q$.
Now $[0, 1] \subset \bigcup_{q \in [-2, 2]} (A + q)$,
Where, the sum is only over rational $q$.
So $1 \leq \sum_{q \in [-2, 2]} m^*(A + q)$, and therefore $m^*(A) > 0$. But
$\bigcup_{q \in [-2, 2]} (A + q) \subset [-6, 6]$, where again the sum is only over rational $q$, so if $m^*$ is a measure,
then $12 \geq \sum_{q \in [-2, 2]} m^*(A + q)$, which implies $m^*(A) = 0$, a contradiction.

Define $m(\bigcup_{i=1}^n (a_i, b_i]) = \sum_{i=1}^n (b_i - a_i)$ if the $(a_i, b_i]$ are disjoint. Note $m$ is well-defined
(a set might be expressible as a union of such intervals in more than one way). The last step in
the construction is the following.

4.1.19 Proposition: Every set in the Borel $\sigma$-algebra is $m^*$-
measurable.

Proof: Since the collection of $m^*$-measurable sets is a $\sigma$-algebra, it suffices to show that every
interval $J$ of the form $(a, b]$ is $m^*$-measurable. Let $E$ be any set with $m^*(E) < \infty$; we need to show
$m^*(E) \geq m^*(E \cap J) + m^*(E \cap J^c)$. \hspace{1cm} (4.4)

Choose $I_1, I_2, \ldots$ of the form $(a_i, b_i]$ such that $E \subset \bigcup_i I_i$ and
$m^*(E) \geq \sum_i (b_i - a_i) - \varepsilon$.

Since $E \subset \bigcup_i I_i$, we have $m^*(E \cap J) \leq m^*(I_i \cap J)$ and $m^*(E \cap J^c) \leq m^*(I_i \cap J^c)$. Hence we have
$m^*(E \cap J) + m^*(E \cap J^c) \leq \sum_i [m^*(I_i \cap J) + m^*(I_i \cap J^c)]$.

Now $m^*(I_i \cap J)$ is the length of an interval and $m^*(I_i \cap J^c)$ is the length of two intervals, so
$m^*(I_i \cap J) + m^*(I_i \cap J^c) = m^*(I_i)$.
Thus
\[ m^* (E \cap J) + m^* (E \cap Jc) \leq m^* (Ii) \leq m^* (E) + \varepsilon. \]

Since \( \varepsilon \) is arbitrary, this proves (3.4).

We now drop the asterisks from \( m^* \) and call \( m \) Lebesgue measure.

**Examples and related results**

**Example 13** We know that the Cantor set is constructed by taking the interval \([0, 1]\), removing the middle third, removing the middle thirds of each of the two remaining subintervals, and continuing. The Cantor set is what remains; it is closed, uncountable, and every point is a limit point. Moreover, it contains no intervals. After one stage, the measure of the two intervals is \( 2(\frac{1}{3}) \), after two stages \( 4(\frac{1}{9}) \) and after \( n \) stages \( (2/3)^n \). Since the Cantor set \( C \) is the intersection of all these sets, the Lebesgue measure of \( C \) is 0.

Suppose we define \( f_0 \) to be

1/2 on the interval \((1/3, 2/3)\),
1/4 on the interval \((1/9, 2/9)\),
3/4 on the interval \((7/9, 8/9)\), and so on.

Define \( f(x) = \inf\{f_0(y) : y \geq x\} \) for \( x < 1 \). Define \( f(1) = 1 \). Notice \( f = f_0 \) on the complement of the Cantor set. \( f \) is monotone, so it has only jump discontinuities. But if it has a jump continuity, there is a rational of the form \( k/2^n \) with \( k \leq 2^n \) that is not in the range of \( f \).

On the other hand, by the construction, each of these values is taken by \( f_0 \) for some point in the complement of \( C \), and so is taken by \( f \). The only way this can happen is if \( f \) is continuous. This function \( f \) is called the Cantor-Lebesgue function. We will use it in examples later on. For now, we can see that it a function that increases only on the Cantor set, which is of Lebesgue measure 0, yet is continuous.

**Example 14** Let \( q_1, q_2, \ldots \) be an enumeration of the rationals, let \( \varepsilon > 0 \), and let \( I_i \) be the interval \((q_i - \varepsilon/2^i, q_i + \varepsilon/2^i)\). Then the measure of \( I_i \) is \( \varepsilon/2^{i-1} \), so the measure of \( \bigcup_i I_i \) is at most \( 2\varepsilon \). (It is not equal to that because there is a lot of overlap.) So the measure of \( A = [0, 1] - \bigcup_i I_i \) is larger than \( 1 - 2\varepsilon \). But \( A \) contains no rational numbers.

**Example 15** Let us follow the construction of the Cantor set, with this difference. Instead of removing the middle third at the first stage, remove the middle fourth, i.e., remove \((3/8, 5/8)\). On each of the two intervals that remain, remove the middle sixteenths. On each of the four intervals that remain, remove the middle interval of length \( 1/64 \), and so on. The total that we removed is

\[ 1/4 + 2(1/16) + 4(1/64) + \cdots = \frac{1}{2} \]

The set that remains contains no intervals, is closed, every point is a limit point, is uncountable, and has measure 1/2. Such a set is called a fat Cantor set or generalized Cantor set. Of course, other choices that \( 1/4, 1/16, \text{etc.} \) are possible.

Let \( A \subset [0, 1] \) be a Borel measurable set. We will show that \( A \) is almost equal to the countable intersection of open sets and almost equal to the countable union of closed sets.. (A similar argument to what follows is possible for sets that have infinite measure.)

**4.1.20 Proposition:** Suppose \( A \subset [0, 1] \) is a Borel measurable set.

(a) There exists a set \( H \) that is the countable intersection of open sets which contains \( A \) and \( m(H - A) = 0 \).

(b) There exists a set \( F \) that is the countable union of closed sets which is contained in \( A \) and \( m(A - F) = 0 \).
Proof.
(a) For each $i$, there is an open set $G_i$ that contains $A$ and such that 
$m(G_i - A) < 2^{-i}$.
This follows from the fact that $m(A) = m^*(A)$ and the definition of $m^*$. Then $H_i = \bigcap_{j\leq i} G_i$ will contain $A$, is open, and since it is contained in $G_i$, then $m(H_i - A) < 2^{-i}$.

Let $H = \bigcap_{i=1}^{\infty} H_i$. $H$ need not be open, but it is the intersection of countably many open sets. The set $H$ is a Borel set, contains $A$, and 
$m(H - A) \leq m(H_i - A) < 2^{-i}$ for each $i$, hence $m(H - A) = 0$.

(b) If $A \subset [0, 1]$, let $F_i = [0, 1] - H_i$, where $H_i$ is a decreasing sequence of open sets containing $A^c$ such that $m(H_i - A^c) < 2^{-i}$. (The $H_i$ are constructed as in the proof of (a), but in terms of $A^c$ instead of $A$.) Then $F_i$ is an increasing sequence of closed sets, $F_i \subset A$ for each $i$, and $m(A - F_i) < 2^{-i}$ for each $i$. Our result follows from letting $F = \bigcup_i F_i$ since $m(A - F) \leq m(A - F_i) < 2^{-i}$ for each $i$, hence $m(A - F) = 0$.

The countable intersections of open sets are sometimes called $G_\delta$ sets; the $G$ is for "geöffnet", the German word for "open" and the $\delta$ for Durchschnitt, the German word for "intersection." The countable unions of closed sets are called $F_\sigma$ sets, the $F$ coming from fermé, the French word for "closed," and the $\sigma$ coming from Summe, the German word for "union." Therefore, when trying to understand Lebesgue measure, we can look at $G_\delta$ or $F_\sigma$ sets, which are not so bad, and at null sets, which can be quite bad but don't have positive measure.

Next we prove the Carathéodory extension theorem. We say that a measure $\mu$ is $\sigma$-finite if there exist $E_1, E_2, \ldots$, such that $\mu(E_i) < \infty$ for all $i$ and $X \subset \bigcup_{i=1}^{\infty} E_i$.

4.1.21 Theorem: Suppose $A_0$ is algebra and $m$ restricted to $A_0$ is a measure. Define $m^*(E) = \inf \{ \sum_{i=1}^{\infty} m(A_i) : A_i \in A_0, E \subset \bigcup_{i=1}^{\infty} A_i \}$.

Then
(a) $m^*(A) = m(A)$ if $A \in A_0$;
(b) every set in $A_0$ is $m^*$-measurable;
(c) if $m$ is $\sigma$-finite, then there is a unique extension to the smallest $\sigma$-algebra containing $A_0$.

Proof. We start with (a). Suppose $E \in A_0$.

We know $m^*(E) \leq m(E)$ since we can take $A_1 = E$ and $A_2, A_3, \ldots$ empty in the definition of $m^*$. If $E \subset \bigcup_{i=1}^{\infty} A_i$ with $A_i \in A_0$, let $B_n = E \cap (A_n - \bigcup_{i=1}^{n-1} A_i)$. Then the $B_n$ are disjoint, they are each in $A_0$, and their union is $E$. Therefore 
$m(E) = \sum_{i=1}^{\infty} m(B_i) \leq \sum_{i=1}^{\infty} m(A_i)$.

Thus $m(E) \leq m^*(E)$.

Next we look at (b). Suppose $A \in A_0$. Let $\varepsilon > 0$ and let $E \subset X$. Pick $B_i \in A_0$ such that $E$
...and
\[ \sum_{i=1}^{\infty} m(B_i) \leq m^*(E) + \varepsilon. \]
Then
\[ m^*(E) + \varepsilon \geq \sum_{i=1}^{\infty} m(B_i) = \sum_{i=1}^{\infty} m(B_i \cap A) + \sum_{i=1}^{\infty} m(B_i \cap A^c) \]
\[ \geq m^*(E \cap A) + m^*(E \cap A^c). \]
Since \( \varepsilon \) is arbitrary,
\[ m^*(E) \geq m^*(E \cap A) + m^*(E \cap A^c). \]
So \( A \) is \( m^* \)-measurable.

Finally, suppose we have two extensions to the smallest \( \sigma \)-algebra containing \( A_0 \); let the other extension be called \( n \). We will show that if \( E \) is in this smallest \( \sigma \)-algebra, then \( m^*(E) = n(E) \).

Since \( m \) is \( \sigma \)-finite, we can reduce to the case where \( m \) is a finite measure:

if \( X = \bigcup_i K_i \) with \( m(K_i) < \infty \) and we prove uniqueness for the measure \( m_i \) defined by \( m_i(A) = m(A \cap K_i) \), then uniqueness for \( m \) follows. So we suppose \( m(X) < \infty \).

Since \( E \) must be \( m^* \)-measurable,
\[ m^*(E) = \inf \{ \sum_{i=1}^{\infty} m(A_i) : A_i \in A_0, E \subset \bigcup_{i=1}^{\infty} A_i \} . \]
But \( m = n \) on \( A_0 \), so
\[ \sum_i m(A_i) = \sum_i n(A_i). \]
Therefore
\[ n(E) \leq \sum_i n(A_i), \]
which implies \( n(E) \leq m^*(E) \).

Since we do not know that \( n \) is constructed via an outer measure, we must use a different argument to get the reverse inequality.

Let \( \varepsilon > 0 \) and choose \( A_i \in A_0 \) such that
\[ m^*(E) + \varepsilon \geq m(A_i) \text{ and } E \subset \bigcup_i A_i. \]
Let \( A = \bigcup_i A_i \) and \( B_k = \bigcup_{i=1}^{k} A_i \).

Observe
\[ m^*(E) + \varepsilon \geq m^*(A), \] hence
\[ m^*(A - E) < \varepsilon. \]
We have
\[ m^*(A) = \lim_{k \to \infty} m^*(B_k) = \lim_{k \to \infty} n(B_k) = n(A). \]
Then
\[ m^*(E) \leq m^*(A) = n(A) \]
\[ = n(E) + n(A - E) \]
\[ \leq n(E) + m(A - E) \]
\[ \leq n(E) + \varepsilon. \]
Since \( \varepsilon \) is arbitrary, this completes the proof.
Points to remember
(1) Uniqueness implies there is only one possible Lebesgue measure.
(2) We will use the Carathéodory extension theorem in the study of product measures.
   It is also used in the Riesz representation theorem and in the Daniell-Kolmogorov extension
   theorem.
(3) We now define Lebesgue-Stieltjes measures.
   Let $\alpha : \mathbb{R} \rightarrow \mathbb{R}$ be nondecreasing and right continuous (i.e., $\alpha(x+) = \alpha(x)$ for all $x$, where $\alpha(x+) = \lim_{y \rightarrow x, y > x} \alpha(y)$). Suppose
   $m_\alpha([a, b]) = \alpha(b) - \alpha(a)$, define
   $m_\alpha(\bigcup_{i=1}^\infty (a_i, b_i]) = \sum_i (\alpha(b_i) - \alpha(a_i))$
   when the intervals $(a_i, b_i]$ are disjoint, and define
   $m_\alpha^*(A) = \inf\{ \sum_i (\alpha(b_i) - \alpha(a_i)) : A \subset \bigcup_i (a_i, b_i]\}. We can show that $m_\alpha^*$ is a measure on the Borel $\sigma$-algebra.
   The $m_\alpha$ measure of a point $x$ is $\alpha(x) - \alpha(x^-)$, where
   $\alpha(x^-) = \lim_{y \rightarrow x, y \searrow x} \alpha(y)$.
   So $m_\alpha(\{x\})$ is equal to the size of the jump (if any) of $\alpha$ at $x$.
(4) Lebesgue measure is the special case of $m_\alpha$ when $\alpha(x) = x$.
(5) Given a measure $\mu$ on $\mathbb{R}$ such that $\mu(K) < \infty$ whenever $K$ is compact, define
   $\alpha(x) = \mu((0, x])$ if $x \geq 0$ and $\alpha(x) = -\mu((x, 0])$ if $x < 0$. Then $\alpha$ is nondecreasing, right continuous, and it is not hard to see that $\mu = m_\alpha$.

Non measurable Sets
We are going to show the existence of a nonmeasurable set. If $x$ and $y$ are real numbers in $[0, 1)$, we define the sum modulo 1 of $x$ and $y$ to be $x + y$ if $x + y < 1$, and to be $x + y - 1$ if $x + y \geq 1$. Let us denote the sum modulo 1 of $x$ and $y$ by $x^* + y$. Then $^*$ is a commutative and associative operation taking pairs of numbers in $[0, 1)$ into numbers in $[0, 1)$. The following lemma shows that Lebesgue measure is invariant under translation modulo 1.

4.1.21 Lemma: Let $E \subset [0, 1)$ be a measurable set. Then for each $y \in [0, 1)$ the set $E^* + y$ is measurable and $m(E^* + y) = mE$.

Proof: Proof of theorem is left as an exercise for reader.
by the rational number \( r \); then \( x \in p_i \). Thus \( \cup P_i = [0, 1) \). Since each \( P_i \) is a translation modulo 1 of \( P \), each \( P_i \) will be measurable if \( P \) is and will have the same measure. But if this were the case,

\[
m(0,1) = \sum_{i=1}^{\infty} mP_i = \sum_{i=1}^{\infty} mP
\]

right side is either zero or infinite, depending on whether \( mP \) zero or positive. But this is impossible since \( m(0,1) = 1 \), and consequently \( P \) cannot be measurable.

**Problems**
1. Show that if \( E \) is a measurable set, then each translate \( E + y \) of \( E \) is also measurable.
2. Show that if \( E_1 \) and \( E_2 \) are measurable, then \( m(E_1 \cup E_2) + m(E_1 \cap E_2) = mE_1 + mE_2 \).

### 4.2 Integration of Nonnegative Functions

**The General Integral**

In this section we introduce the Lebesgue integral.

**4.2.1 Definition:** If \( E \subset X \), define the characteristic function of \( E \) by

\[
\chi_E(x) = \begin{cases} 
1 & x \in E; \\
0 & x \notin E.
\end{cases}
\]

A simple function \( s \) is one of the form

\[
S(x) = \sum_{i=1}^{n} a_i \chi_{E_i}(x)
\]

For real \( a_i \) and measurable set \( E_i \).

**4.2.2 Proposition:** Suppose \( f \geq 0 \) is measurable. Then there exists a sequence of nonnegative measurable simple functions increasing to \( f \).

**Proof:** Let

\[
E_n = \{x : (i - 1)/2^n \leq f(x) < i/2^n \}
\]

and \( F_n = \{x : f(x) \geq n \} \)

for \( n = 1, 2, \ldots \), and \( i = 1, 2, \ldots, n^2 \).

Then define

\[
s_n = \sum_{i=1}^{n^2} \frac{i - 1}{2^n} \chi_{E_i} + n \chi_{F_n}.
\]

It is easy to see that \( s_n \) has the desired properties.

**4.2.3 Definition:** If

\[
s = \sum_{i=1}^{n} a_i \chi_{E_i}
\]

is a nonnegative measurable simple function, define the Lebesgue integral of \( s \) to be

\[
\int s \, d\mu = \sum_{i=1}^{n} a_i \mu(E_i).
\]  \hfill (4.5)

If \( f \geq 0 \) is a measurable function, define

\[
\int f \, d\mu = \sup\{\int s \, d\mu : 0 \leq s \leq f, s \text{ simple}\}.
\]  \hfill (4.6)

If \( f \) is measurable and at least one of the integrals

\[
\int f^+ d\mu, \int f^- d\mu
\]

is finite,

where \( f^+ = \max(f, 0) \) and \( f^- = -\min(f, 0) \),
\[
\int f \, d\mu = \int f^+ d\mu - \int f^- d\mu.
\]

(4.7)

Finally, if \( f = u + iv \) and \( \int (|u| + |v|) \, d\mu \) is finite,

\[
\int f \, d\mu = \int u \, d\mu + i \int v \, d\mu.
\]

(4.8)

A few remarks are in order. A function \( s \) might be written as a simple function in more than one way. For example \( \chi_{A \cup B} = \chi_A + \chi_B \) is \( A \) and \( B \) are disjoint. It is clear that the definition of \( \int s \, d\mu \) is unaffected by how \( s \) is written. Secondly, if \( s \) is a simple function, one has to think a moment to verify that the definition of \( \int s \, d\mu \) by means of (4.5) agrees with its definition by means of (4.6).

4.2.4 Definition: If \( \int |f| \, d\mu < \infty \), we say \( f \) is integrable.

The proof of the next proposition follows from the definitions.

4.2.5 Proposition: (a) If \( f \) is measurable, \( a \leq f(x) \leq b \) for all \( x \), and \( \mu(X) < \infty \), then

\[
a\mu(X) \leq \int f \, d\mu \leq b\mu(X);
\]

(b) If \( f(x) \leq g(x) \) for all \( x \) and \( f \) and \( g \) are measurable and integrable, then

\[
\int f \, d\mu \leq \int g \, d\mu.
\]

(c) If \( f \) is integrable, then \( \int cf \, d\mu = c \int f \, d\mu \) for all real \( c \).

(d) If \( \mu(A) = 0 \) and \( f \) is measurable, then \( \int fd\mu = 0 \).

The integral \( \int f\chi_A \, d\mu \) is often written \( \int f\chi_A \, d\mu \). Other notation for the integral is to omit the \( \mu \) if it is clear which measure is being used, to write \( \int f(x) \, \mu(dx) \), or to write \( \int f(x) \, d\mu(x) \).

4.2.6 Proposition: If \( f \) is integrable, then \( \int |f| \leq \int |f| \).

Proof. For the real case, this is easy.

\[
f \leq |f|, \text{ so } \int f \leq \int |f|.
\]

Also

\[
-f \leq |f|, \text{ so } -f \leq |f|.
\]

Now combine these two facts. For the complex case, \( f \) is a complex number. If it is 0, the inequality is trivial. If it is not, then \( f = re^{i\theta} \) for some \( r \) and \( \theta \). Then

\[
|\int f| = r = e^{-i\theta} \int f = \int e^{-i\theta}f.
\]

From the definition of \( f \) when \( f \) is complex, we have \( \text{Re} (\int f) = \int \text{Re} (f) \).

Since \( \int |f| \) is real, we have,

\[
\int |f| = \text{Re} (\int e^{-i\theta}f) = \int \text{Re} (e^{-i\theta}f) \leq \int |f|.
\]

We do not yet have that \( \int (f + g) = \int f + \int g \).

**Integration of series**
One of the most important results concerning Lebesgue integration is the **monotone convergence theorem**.

### 4.2.7 Theorem:
Suppose \( f_n \) is a sequence of nonnegative measurable functions with 
\[ f_1(x) \leq f_2(x) \leq \cdots \text{ for all } x \] 
and with \( \lim_{n \to \infty} f_n(x) = f(x) \) for all \( x \).

Then \( \int f_n \, d\mu \to \int f \, d\mu \).

**Proof.** We know, 
If \( f(x) \leq g(x) \) for all \( x \) and \( f \) and \( g \) are measurable and integrable, then 
\[ \int f \, d\mu \leq \int g \, d\mu. \]

hence
\[ \int f_n \text{ is an increasing sequence of real numbers. Let } L \text{ be the limit.} \]

Since \( f_n \leq f \) for all \( n \), then \( L \leq \int f \). We must show \( L \geq f \).

Let \( s = \sum_{i=1}^{m} a_i \chi_{E_i} \) be any nonnegative simple function less than \( f \) and let \( c \in (0, 1) \). Let \( A_n = \{ x : f_n(x) \geq cs(x) \} \). Since the \( f_n(x) \) increases to \( f(x) \) for each \( x \) and \( c < 1 \), then \( A_1 \subset A_2 \subset \cdots \), and 
the union of the \( A_n \) is all of \( X \).

For each \( n \),
\[
\int f_n \geq \int_{A_n} f_n \geq c \int_{A_n} s_n \\
= c \int_{A_n} \sum_{i=1}^{m} a_i \chi_{E_i} \\
= c \sum_{i=1}^{m} a_i \mu(E_i \cap A_n).
\]

If we let \( n \to \infty \), we have if \( A_i \in A, A_1 \subset A_2 \subset \cdots \), and \( A = \bigcup_{i=1}^{\infty} A_i \), then 
\[ \mu(A) = \lim_{n \to \infty} \mu(A_n). \]

The right hand side converges to \( c \sum_{i=1}^{m} a_i \mu(E_i) = c \int s \)
Therefore \( L \geq c \int s \). Since \( c \) is arbitrary in the interval \( (0, 1) \), then \( L \geq \int s \).

Taking the supremum over all simple \( s \leq f \), we obtain \( L \geq f \).

**Example 16** Let \( X = [0, \infty) \) and \( f_n(x) = -1/n \) for all \( x \). Then \( \int f_n = -\infty \), but \( f_n \uparrow f \) where \( f = 0 \) and 
\[ \int f = 0. \] The problem here is that the \( f_n \) are not nonnegative.

**Example 17** Suppose \( f_n = n\chi_{(0,1/n)} \). Then \( f_n \geq 0, f_n \to 0 \) for each \( x \), but \( \int f_n = 1 \) does not converge to \( \int 0 = 0 \). The trouble here is that the \( f_n \) do not increase for each \( x \).

**Reimann and Lebesgue Integral:** We are familiar with the definition of Riemann Integral in earlier chapter. Once we have the monotone convergence theorem. Now we can prove that the Lebesgue integral is linear and compare Riemann and Lebesgue integral.

### 4.2.8 Theorem:
If \( f_1 \) and \( f_2 \) are integrable, then 
\[ \int (f_1 + f_2) = \int f_1 + \int f_2. \]

**Proof.** First suppose \( f_1 \) and \( f_2 \) are nonnegative and simple. Then it is clear from the definition that 
the theorem holds in this case.

Next suppose \( f_1 \) and \( f_2 \) are nonnegative.

Take \( s_n \) simple and increasing to \( f_1 \) and \( t_n \) simple and increasing to \( f_2 \).
Then $s_n + t_n$ increases to $f_1 + f_2$.

So from the monotone convergence theorem if $f_n$ is a sequence of nonnegative measurable functions with $f_1(x) \leq f_2(x) \leq \cdots$ for all $x$ with $\lim_{n \to \infty} f_n(x) = f(x)$ for all $x$.

Then $\int f_n d\mu \to f d\mu$, and the result for simple functions.

Finally in the general case, write $f_1 = f^+ - f^-$ and similarly for $f_2$, and use the definitions and the result for nonnegative functions.

**4.2.9 Proposition:** Suppose $f_n$ are nonnegative measurable functions. Then

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$  \hfill (4.9)

**Proof.** Let $F_N = \sum_{n=1}^{N} f_n$ and write

$$\int \sum_{n=1}^{\infty} f_n = \int \lim_{N \to \infty} \sum_{n=1}^{N} f_n = \lim_{N \to \infty} \int F_N = \lim_{N \to \infty} \sum_{n=1}^{N} \int f_n = \sum_{n=1}^{\infty} \int f_n,$$

using the monotone convergence theorem and the linearity of the integral.

The next theorem is known as **Fatou’s lemma.**

**4.2.10 Theorem:** Suppose the $f_n$ are nonnegative and measurable. Then

$$\int \liminf_{n \to \infty} f_n \leq \liminf_{n \to \infty} \int f_n.$$  \hfill (4.10)

**Proof.** Let $g_n = \inf_{i \geq n} f_i$. Then $g_n$ are nonnegative and $g_n$ increases to $\liminf f_n$.

Clearly $g_n \leq f_i$ for each $i \geq n$, so $\int g_n \leq \int f_i$.

Therefore

$$\int g_n \leq \inf_{i \geq n} \int f_i.$$  \hfill (4.11)

If we take the limit as $n \to \infty$, on the left hand side we obtain $\int \liminf f_n$ by the monotone convergence theorem, while on the right hand side we obtain $\liminf \int f_n$.

A typical use of Fatou’s lemma is the following.

Suppose we have $f_n \to f$ and $\sup_n \int |f_n| \leq K < \infty$.

Then $|f_n| \to |f|$, and by Fatou’s lemma, $\int |f| \leq K$.

Another very important theorem is the **dominated convergence theorem.**

**4.2.11 Theorem:** Suppose $f_n$ are measurable functions and $f_n(x) \to f(x)$. Suppose there exists an integrable function $g$ such that $|f_n(x)| \leq g(x)$ for all $x$. Then $\int f_n d\mu \to \int f d\mu$.

**Proof.** Since $f_n + g \geq 0$, by Fatou’s lemma,
\[ \int (f + g) \leq \lim \inf \int (f_n + g). \]
Since \( g \) is integrable,
\[ \int f \leq \lim \inf \int f_n. \]
Similarly, \( g - f_n \geq 0 \), so
\[ \int (g - f) \leq \lim \inf \int (g - f_n), \]
and hence
\[ -\int f \leq \lim \inf \int (-f_n) = -\lim \sup \int f_n \]
Therefore
\[ \int f \geq \lim \sup \int f_n, \] which with the above proves the theorem.

**Example 18** is an example where the limit of the integrals is not the integral of the limit because there is no dominating function \( g \). If in the monotone convergence theorem or dominated convergence theorem we have only \( f_n(x) \to f(x) \) almost everywhere, the conclusion still holds. For if the \( f_n \) and \( f \) are measurable and \( A = \{ x : f_n(x) \to f(x) \} \), then \( f\chi_A \to f\chi_A \) for each \( x \). And since \( A^c \) has measure 0, we see from Proposition 4.2.5(d) that \( f\chi_A = f \), and similarly with \( f \) replaced by \( f_n \).

**Properties of Lebesgue integrals**
We will need the following two propositions.

**4.2.12 Proposition:** Suppose \( f \) is measurable and for every measurable set \( A \) we have
\[ \int_A f \, d\mu = 0. \] Then \( f = 0 \) almost everywhere.

**Proof.** Let \( A = \{ x : f(x) > \varepsilon \} \). Then
\[ 0 = \int_A f \geq \int_A \varepsilon = \varepsilon \mu(A) \]
since \( f\chi_A \geq \varepsilon \chi_A \).
Hence \( \mu(A) = 0 \).

We use this argument for \( \varepsilon = 1/n \) and \( n = 1, 2, \ldots \), so \( \mu \{ x : f(x) > 0 \} = 0 \).
Similarly \( \mu \{ x : f(x) < 0 \} = 0 \).

**4.2.13 Proposition:** Suppose \( f \) is measurable and nonnegative and \( \int f \, d\mu = 0. \) Then \( f = 0 \) almost everywhere.

**Proof.** If \( f \) is not almost everywhere equal to 0, there exists an \( n \) such that \( \mu(\text{An}) > 0 \) where \( \text{An} = \{ x : f(x) > 1/n \} \).
But then since \( f \) is nonnegative,
\[ 0 = \int f \geq \int\text{An} f \geq 1/n \mu(\text{An}), \] a contradiction.

We give a result on approximating a function on \( R \) by continuous functions.

**4.2.14 Proposition:** Suppose \( f \) is a measurable function from \( R \) to \( R \) that is integrable. Let \( \varepsilon > 0. \) Then there exists a continuous function that is 0 outside some bounded interval such that
\[ \int |f - g| < \varepsilon. \]

**Proof.** If we have continuous functions \( g_1, g_2 \) such that
\[ \int |f^* - g_1| < \varepsilon/2 \text{ and } \int |f^- - g_2| < \varepsilon/2, \]

where 
\[ f^* = \max(f, 0) \text{ and } f^- = \max(-f, 0), \]

then taking 
\[ g = g_1 - g_2 \]

will prove our result.

Without loss of generality, we may assume \( f \geq 0. \)

By monotone convergence \( \int f \chi_{[-n,n]} \) increases to \( \int f \), so by taking \( n \) large enough, the difference of the integrals will be less than \( \varepsilon/2. \)

If we find \( g \) such that \( \int |f \chi_{[-n,n]} - g| < \varepsilon/2, \) then \( \int |f - g| < \varepsilon. \)

Therefore we may assume that \( f \) is 0 outside some bounded interval.

We can find simple functions increasing to \( f \) whose integrals increase to \( \int f. \)

Let \( s_m \) be a simple function such that \( s_m \leq f \) and \( \int s_m \geq f - \varepsilon/2. \)

If we find \( g \) such that \( \int |s_m - g| < \varepsilon/2, \) then \( \int |f - g| < \varepsilon. \)

So it suffices to consider the case where \( f \) is a simple function.

If \( f = \sum_{i=1}^p a_i \chi_{A_i} \) and we find \( g_i \) continuous such that \( \int |a_i \chi_{A_i} - g_i| < \varepsilon/p, \) then 
\[ \sum_{i=1}^p g_i \]

will be the desired function.

So we may assume \( f \) is a constant times a characteristic function, and by linearity, we may assume \( f \) is equal to \( \chi_A \) for some \( A \) contained in a bounded interval \([-n, n]. \)

We can choose \( G \) open and \( F \) closed such that \( F \subset A \subset G \) and \( m(G - F) < \varepsilon. \)

We can replace \( G \) by \( G \cap (-n -1, n +1). \) \( G^c \cap [-n -1, n +1] \) and \( F \) are compact sets, so there is a minimum distance between them, say, \( \delta. \)

Let \( g(x) = \max(0, 1 - \text{dist}(x, F)/\delta). \)

Then 
\( g \)

is continuous, \( 0 \leq g \leq 1, \)
\( g \)

is 1 on \( F, \)
\( g \)

is 0 on \( G^c, \) and
\( g \)

is 0 outside of \([-n -1, n +1]. \)

Therefore 
\[ |g - \chi_A| \leq \chi_G - \chi_F, \]

so 
\[ \int |g - \chi_A| \leq \int (\chi_G - \chi_F) = m(G - F) < \varepsilon. \]

**Remark:** The method of proof, where one proves a result for characteristic functions, then simple functions, then non-negative functions, and then finally integrable functions is very common.

We finish this section with a

**Comparison of the Lebesgue integral and the Riemann integral.**

Here we are only looking at bounded functions from \([a, b]\) into \( R. \)

If we are looking at the Lebesgue integral, we write \( \int f, \) while,

If we are looking at the Riemann integral, we write \( R(f). \)

Recall that the Riemann integral on \([a, b]\) is defined as follows:
If $P$ is a partition of $[a, b]$, then

$$U(P, f) = \sum_{i=1}^{n} \text{Sup}_{x_{i-1} \leq x \leq x_{i}} f(x) (x_{i} - x_{i-1})$$

And

$$L(P, f) = \sum_{i=1}^{n} \text{Inf}_{x_{i-1} \leq x \leq x_{i}} f(x) (x_{i} - x_{i-1})$$

Set

$$\bar{R} (f) = \inf \{ U(P, f) : P \text{ is a partition} \}$$

$$R (f) = \sup \{ U(P, f) : P \text{ is a partition} \}$$

Then the Riemann integral exists if $\bar{R} (f) = R (f)$, and the common value is the Riemann integral, which we denote $R(f)$.

**Theorem:** A bounded measurable function $f$ on $[a, b]$ is Riemann integrable if and only if the set of points at which $f$ is discontinuous has Lebesgue measure 0, and in that case, the Riemann integral is equal in value to the Lebesgue integral.

**Proof.** If $P$ is a partition, define

$$T_{P}(x) = \sum_{i=1}^{n} \text{Sup}_{x_{i-1} \leq x \leq x_{i}} f(y) \chi_{[x_{i-1}, x_{i})}(x),$$

And

$$S_{P}(x) = \sum_{i=1}^{n} \text{Inf}_{x_{i-1} \leq x \leq x_{i}} f(y) \chi_{[x_{i-1}, x_{i})}(x),$$

We see that $\int T_{P} = U(P, f)$ and $\int S_{P} = L(P, f)$.

If $f$ is Riemann integrable, there exists a sequence of partitions $Q_{i}$ such that $U(Q_{i}, f) \downarrow R(f)$ and a sequence $Q_{i}$ such that $L(Q_{i}, f) \uparrow R(f)$.

It is not hard to check that adding points to a partition increases $L$ and decreases $U$, so if we let $P_{i} = \bigcup_{j \leq i} (Q_{j} \cup Q'_{j})$, then $P_{i}$ is an increasing sequence of partitions, $U(P_{i}, f) \downarrow R(f)$, $L(P_{i}, f) \uparrow R(f)$.

We see also that $T_{P}(x)$ decreases at each point, say, to $T(x)$, and $S_{P}(x)$ increases at each point, say, to $S(x)$.

Also

$$T(x) \geq f(x) \geq S(x).$$

Then by dominated convergence (recall that $f$ is bounded)

$$\int (T - S) = \lim_{i \to \infty} \int (T_{P_{i}} - S_{P_{i}}) = \lim_{i \to \infty} (U(P_{i}, f) - L(P_{i}, f)) = 0.$$

We conclude $T = S = f$ a.e. If $x$ is not in the null set where $T(x) \neq S(x)$ nor in $\bigcup_{i} P_{i}$, which is countable and hence of lebesgue measure 0, then $T_{P_{i}}(x) \downarrow f(x)$ and $S_{P_{i}}(x) \uparrow f(x)$. This implies that $f$ is continuous at such $x$.

Since

$$R(f) = \lim_{i \to \infty} U(P_{i}, f) = \lim_{i \to \infty} \int P_{i} = \int f,$$

we see the Riemann integral and Lebesgue integral agree.

Now suppose that $f$ is continuous a.e. Let $\varepsilon > 0$. Let $P_{i}$ be the partition where we divide $[a, b]$ into $2^{i}$ equal parts. If $x$ is not in the null set where $f$ is discontinuous, nor in $\bigcup_{i} P_{i}$, then $T_{P_{i}}(x) \downarrow f(x)$ and $S_{P_{i}}(x) \uparrow f(x)$.
By dominated convergence,
\[ U(P_i, f) = \int T_{P_i} \to \int f \]
And
\[ L(P_i, f) = \int S_{P_i} \to \int f. \]
This does it.

4.3 Differentiation and Integration

Four Derivative

4.3.1 Definition:(Four Derivative) Given a function \( f \), we define the derivatives of \( f \) at \( x \) by

\[
\begin{align*}
D^+ f(x) &= \lim_{h \to 0^+} \frac{f(x+h) - f(x)}{h} \\
D^- f(x) &= \lim_{h \to 0^-} \frac{f(x+h) - f(x)}{h} \\
D_+ f(x) &= \lim_{h \to 0^+} \frac{f(x+h) - f(x)}{h} \\
D_- f(x) &= \lim_{h \to 0^-} \frac{f(x+h) - f(x)}{h}
\end{align*}
\]

When all the derivatives are equal we say that \( f \) is differentiable at \( x \) and define \( f'(x) \) to be the common value.

Functions of bounded Variation

In this section we want to look at when \( f : \mathbb{R} \to \mathbb{R} \) is differentiable and when the fundamental theorem of calculus holds. Briefly,

(1) Functions of bounded variation are differentiable;

(2) The derivative of \( \int_a^b f(y) \, dy \) is equal to \( f \) a.e. if \( f \) is integrable;

(3) \( \int_a^b f'(y) \, dy = f(b) - f(a) \) if \( f' \) is absolutely continuous.

Let \( E \subseteq \mathbb{R} \) be a measurable set and let \( O \) be a collection of intervals. We say \( O \) is a Vitali cover of \( E \) if for each \( x \in E \) and each \( \varepsilon > 0 \) there exists an interval \( G \in O \) containing \( x \) whose length is less than \( \varepsilon \). \( m \) will denote Lebesgue measure.

4.3.2 Lemma: Let \( E \) have finite measure and let \( O \) be a Vitali cover of \( E \). Given \( \varepsilon > 0 \) there exists a finite sub collection of disjoint intervals \( I_1, \ldots, I_n \) such that \( m(E - \bigcup_{i=1}^n I_i) < \varepsilon \).

Proof. We may replace each interval in \( O \) by a closed one, since the set of endpoints of a finite sub collection will have measure 0.

Let \( O \) be an open set of finite measure containing \( E \).
Since \( O \) is a Vitali cover, we may suppose without loss of generality that each set of \( O \) is contained in \( O \).
Let \( a_0 = \sup \{ m(I) : I \in O \} \).
Let \( I_1 \) be any element of \( O \) with \( m(I_1) \geq a_0/2 \).
Let \( a_1 = \sup \{ m(I) : I \in O \}, I \text{ disjoint from } I_1 \}, \text{ and choose } I_2 \in O \text{ disjoint from } I_1 \text{ such that } m(I_2) \geq a_1/2 \).
Continue in this way, choosing \( I_{n+1} \) disjoint from \( I_1, \ldots, I_n \) and in \( O \) with length at least one half as large as any other such interval in \( O \) that is disjoint from \( I_1, \ldots, I_n \).
If the process stops at some finite stage, we are done.
If not, we generate a sequence of disjoint intervals $I_1, I_2, \ldots$. Since they are disjoint and all contained in $O$, then $\sum_{i=1}^{\infty} m(I_i) \leq m(O) < \infty$.

So there exists $N$ such that $\sum_{i=N+1}^{\infty} m(I_i) < \epsilon/5$.

Let $R = E - \bigcup_{i=1}^{N} I_i$; we will show $m(R) < \epsilon$. Let $J_n$ be the interval with the same center as $I_n$ but five times the length. Let $x \in R$. There exists an interval $I \in O$ containing $x$ with $I$ disjoint from $I_1, \ldots, I_N$. Since $\sum_{i=1}^{n} m(I_i) < \infty$, then $\sum a_n \leq 2 \sum m(I_n) < \infty$, and $a_n \to 0$. So I must either be one of the $I_n$ for some $n > N$ or at least intersect it, for otherwise we would have chosen $I$ at some stage.

Let $n$ be the smallest integer such that $I$ intersects $I_n$; note $n > N$. We have $m(I) \leq a_{n-1} \leq 2m(I_n)$.

Therefore $x \in J_n$. Then $R \subseteq \bigcup_{i=N+1}^{\infty} J_i$, so

$$m(R) \leq \sum_{i=N+1}^{\infty} m(I_i) = 5 \sum_{i=N+1}^{\infty} m(I_i) < \epsilon.$$

**Lebesgue Differentiation Theorem**  

4.3.3 Theorem: Suppose $f$ is nondecreasing on $[a, b]$. Then $f$ is differentiable almost everywhere, $f'$ is integrable, and $\int_a^b f'(x)dx \leq f(b) - f(a)$.

**Proof.** We will show that the set where any two derivates are unequal has measure zero. We consider the set $E$ where $D^+ f(x) > D^- f(x)$, the other sets being similar.

Let $E_{u,v} = \{x : D^+ f(x) > u > v > D^- f(x)\}$.

If we show $m(E_{u,v}) = 0$, then taking the union of all pairs of rationals with $u > v$ shows $m(E) = 0$.

Let $s = m(E_{u,v})$, $\epsilon > 0$, and choose an open set $O$ such that $E_{u,v} \subseteq O$ and $m(O) < s + \epsilon$.

For each $x \in E_{u,v}$ there exists an arbitrarily small interval $[x - h, x]$ contained in $O$ such that $f(x) - f(x - h) < vh$.

We know if $E$ have finite measure and $O$ be a Vitali cover of $E$. Then Given $\epsilon > 0$ there exists a finite sub collection of disjoint intervals $I_1, \ldots, I_n$ such that $m(E - \bigcup_{i=1}^{n} I_i) < \epsilon$. So consider $I_1, \ldots, I_n$ which are disjoint and whose interiors cover a subset $A$ of $E_{u,v}$ of measure greater than $s - \epsilon$. Suppose $I_n = [x_n - h_n, x_n]$. Summing over these intervals,

$$\sum_{n=1}^{N} [f(x_n) - f(x_n - h_n)] < s \sum_{n=1}^{N} h_n < vm(O) < v(s + \epsilon).$$

Each point $y \in A$ is the left endpoint of an arbitrarily small interval $(y, y + k)$ that is contained in some $I_n$ and for which $f(y + k) - f(y) > uk$.

Using Lemma 9.1 again, we pick out a finite collection $J_1, \ldots, J_M$ whose union contains a subset of $A$ of measure larger than $s - 2\epsilon$. Summing over these intervals yields

$$\sum_{i=1}^{M} [f(y_i + k_i) - f(y_i)] > u \sum k_i > u(s - 2\epsilon).$$

Each interval $J_i$ is contained in some interval $I_n$, and if we sum over those $i$ for which $J_i \subseteq I_n$ we find
\[
\sum_{i=1}^{k} |f(y_i + k_i) - f(y_i)| \leq f(x_n) - f(x_n - h_n),
\]
since \( f \) is increasing. Thus
\[
\sum_{n=1}^{N} [f(x_n) - f(x_n - h_n)] \geq \sum_{i=1}^{M} [f(y_i + k_i) - f(y_i)]
\]
and so \( v(s + \varepsilon) > u(s - 2\varepsilon) \). This is true for each \( \varepsilon \), so \( vs \geq us \). Since \( u > v \), this implies \( s = 0 \). This shows that
\[
g(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}
\]
is defined almost everywhere and that \( f \) is differentiable wherever \( g \) is finite.
Define \( \bar{f}(x) = f(b) \) if \( x \geq b \).
Let \( g_n(x) = n[f(x + 1/n) - f(x)] \).
Then \( g_n(x) \to g(x) \) for almost all \( x \), and so \( g \) is measurable.
Since \( f \) is increasing, \( g_n \geq 0 \). By Fatou’s lemma
\[
\int_a^b g \leq \lim \inf \int_a^b g_n = \lim \inf n \int_a^b [f(x + 1/n) - f(x)] \text{dx}
\]
\[
= \lim \inf \left[ n \int_{a+1/n}^{b+1/n} f - n \int_a^{a+1/n} f \right]
\]
\[
= \lim \inf \left[ f(b) - n \int_a^{a+1/n} f \right]
\]
\[
\leq f(b) - f(a).
\]
For the last inequality, we use the fact that \( f \) is increasing. This shows that \( g \) is integrable and hence finite almost everywhere.

A function is of bounded variation if
\[
\sup \left\{ \sum_{i=1}^{k} k |f(x_i) - f(x_{i-1})| \right\} \text{ is finite, where the supremum is over all partitions}
\]
a = \( x_0 < x_1 < \cdots < x_k = b \) of \( [a, b] \).

4.3.4 Lemma: If \( f \) is of bounded variation on \( [a, b] \), then \( f \) can be written as the difference of two nondecreasing functions on \( [a, b] \).

Proof. Define
\[
P(y) = \sup \left\{ \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \right\}, \quad N(y) = \sup \left\{ \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \right\},
\]
where the supremum is over all partitions \( a = x_0 < x_1 < \cdots < x_k = y \) for \( y \in [a, b] \).
\( P \) and \( N \) are measurable since they are both increasing. Since
\[
\sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| = \sup \left\{ \sum_{i=1}^{k} |f(x_i) - f(x_{i-1})| \right\} + f(y) - f(a),
\]
taking the supremum over all partitions of \( [a, y] \) yields
\[
P(y) = N(y) + f(y) - f(a).
\]
Clearly \( P \) and \( N \) are nondecreasing in \( y \), and by solving for \( f(y) \).
\( P(y) - N(y) + f(a) = f(y) \)
Result follows.

From this lemma, we see that functions of bounded variation are differentiable a.e. But the function \( \sin(1/x) \) defined on \( (0, 1] \) is differentiable everywhere, but is not of bounded variation.

Next we look at when the derivative of \( \int_a^x f(t) \text{ dt} \) is equal to \( f(x) \) a.e.
Define the indefinite integral of an integrable function \( f \) by
\[
F(x) = \int_a^x f(t) \, dt.
\]

**4.3.4 Lemma:** If \( f \) is integrable, then \( F \) is continuous and of bounded variation.

**Proof.** The continuity follows from the dominated convergence theorem. The bounded variation follows from
\[
\sum_{i=1}^k |F(x_i) - F(x_{i-1})| = \sum_{i=1}^k \left| \int_{x_{i-1}}^{x_i} f(t) \, dt \right| \leq |f(t)| \, dt
\]
\[
\leq \sum_{i=1}^k \int_{x_{i-1}}^{x_i} |f(t)| \, dt
\]
\[
\leq \int_a^b |f(t)| \, dt
\]
for all partitions.

**4.3.5 Lemma:** If \( f \) is integrable and \( F(x) = 0 \) for all \( x \), then \( f = 0 \) a.e.

**Proof.** For any interval,
\[
\int_a^d f = \int_a^c f - \int_a^c f = 0.
\]
By dominated convergence and the fact that any open set is the countable union of disjoint open intervals, \( \int_a f = 0 \) for any open set \( O \).

If \( E \) is any measurable set, take \( O_n \) open that such that \( \chi_{O_n} \) decreases to \( \chi_E \) a.e.

By dominated convergence,
\[
\int_E f = \int \chi_E f = \lim \int \chi_{O_n} f = \lim \int_{O_n} f = 0.
\]
This with Proposition 4.2.12 implies \( f \) is zero a.e.

**4.3.6 Proposition:** If \( f \) is bounded and measurable, then \( F'(x) = f(x) \) for almost every \( x \).

**Proof.** We know if \( f \) is integrable, then \( F \) is continuous and of bounded variation so \( F' \) exists a.e.

Let \( K \) be a bound for \( |f| \). If
\[
f_n(x) = \lim_{n \to 0} \frac{F(x+1/n) - F(x)}{1/n}
\]
Then
\[
f_n(x) = n \int_x^{x+1/n} f(t) \, dt
\]
so \( |f_n| \) is also bounded by \( K \). Since \( f_n \to F' \) a.e., then by dominated convergence,
\[
\int_a^c F'(x) \, dx = \lim \int_a^c f_n(x) \, dx
\]
\[
= \lim n \int_a^c [F(x+1/n) - F(x)] \, dx
\]
\[
= \lim [n \int_c^{c+1/n} F(x) \, dx - n \int_a^{a+c} F(x) \, dx]
\]
\[
= F(c) - F(a)
\]
\[
= \int_a^c f(x) \, dx,
\]
using the fact that \( F \) is continuous. So \( \int_a^c \left[ F'(x) - f(x) \right] \, dx = 0 \) for all \( c \),
which implies \( F' = f \) a.e.

**4.3.7 Theorem:** If \( f \) is integrable, then \( F' = f \) almost everywhere.

**Proof.** Without loss of generality we may assume \( f \geq 0 \).

Let \( f_n(x) = f(x) \) if \( f(x) \leq n \) and let \( f_n(x) = n \) if \( f(x) > n \). Then \( f - f_n \geq 0 \).

If \( G_n(x) = \int_a^x \left[ f - f_n \right] \), then \( G_n \) is nondecreasing, and hence has a derivative almost everywhere.

We know the derivative of \( \int_a^x f_n \) is equal to \( f_n \) almost everywhere. Therefore

\[
F'(x) = G'_n(x) + \left[ \int_a^x f_n \right] \geq f_n(x) \text{ a.e.}
\]

Since \( n \) is arbitrary, \( F' \geq f \) a.e.

So \( \int_a^b F' \geq \int_a^b f = F(b) - F(a) \).

On the other hand,

\[
\int_a^b F'(x) \, dx \leq F(b) - F(a) = \int_a^b f.
\]

We conclude that

\[
\int_a^b \left[ F' - f \right] = 0
\]

Since \( F' - f \geq 0 \) this tells us \( F' = f \) a.e.

---

### 4.4 Assignment

1. Prove that completeness of \( \mu \).
2. Let \( X \) be the set of rational numbers and \( \hat{\mathcal{A}} \) the algebra of finite unions of intervals of the form \( (a, b) \) with \( \mu(a, b) = \infty \) and \( \mu_{\phi} = 0 \). Then show that, extension of \( \mu \) to the smallest \( \sigma \)-algebra containing \( \hat{\mathcal{A}} \) is not unique.
3. Let \( C \) be a semi algebra of sets and \( \hat{\mathcal{A}} \) the smallest algebra of sets containing \( C \). Then, Show that \( \hat{\mathcal{A}} \) is comprised of sets of the form \( A = \bigcup_{i=1}^n C_i \) with \( C_i \in C \)
4. Let \( \hat{\mathcal{A}} \) be a collection of sets which is closed under finite unions and finite intersections; then, Show that \( \hat{\mathcal{A}}_c \) is closed under countable unions and finite intersections.
5. Let $\mu$ be a finite measure on an algebra $\mathcal{A}$, and $\mu^*$ the induced outer measure. Show that a set $E$ is measurable if and only if for each $\epsilon > 0$ there is a set $A \in \mathcal{A}$ such that $\mu^*(E \sim A) < \epsilon$.

6. Consider an outer measure $\mu^*$ on $X$ and form the induced measure $\mu$ on the $\mu^*$-measurable sets, also use $\hat{\mu}$ to induce an outer measure $\mu^+$.
   a. Show that for each set $E$, $\mu^+ E \geq \mu^* E$.
   b. For a given set prove that, $\mu^+ E = \mu^* E$ if and only if there is $\mu^*$-measurable sets $A \supsetneq \mu^* A = \mu^* E$.
   c. Show that $\mu^+ E = \mu^* E$ for every $E$ if and only if $\mu^*$ is regular.
   d. Show that an outer measure $\mu^*$ is regular if and only if it is induced by a measure on an algebra.
   e. Let $X$ be a set consisting of two points. Construct an outer measure on $X$ which is not regular.

4.5 Check your progress

1. Let $\mu^*$ be a regular outer measure.
   a. Show that the measure $\mu$ induced by $\mu^*$ is complete and saturated.
   b. Let $(X, \mathcal{A}, \mu)$ be a complete measure space. Let $\mu$ be the extension of $\mu$ obtained by the Carathéodory process. Then show that $\mu$ is the extension given $\mu^+ E = \mu^* E$.

2. Let $\mu$ be a measure on an algebra $\mathcal{A}$ and $\mu$ the extension of it given by the Carathéodory process. Let $E$ be measurable with respect to $\mu$ and $\mu E < \infty$. Then given $\epsilon > 0$, there is an $A \in \mathcal{A}$, with $\mu (A \vee E) < \epsilon$.

3. Let $\mu$ be a measure on an algebra $\mathcal{A}$, and let $E$ be a $\mu^*$ measurable set.
   Show that the measure $\mu$ defined by $\mu E = \mu^* E$ has the property that $\mu B = \mu^* B$ for all $B \in \mathcal{B}$ & Thus $\mu$ agrees on $\mathcal{B}$ with the measure given by the Carathéodory extension process.

4. Let $\mathcal{A}$ be the algebra of finite unions of half-open intervals of $\mathbb{R}$ and let $\mu_\phi = 0$ and $\mu A = \infty$ for $A \neq \phi$. The class (of Borel sets is the smallest $\sigma$-algebra containing $\mathcal{A}$.
   a. Show that $\mu^* E = \infty$ if $E \neq \phi$.
   b. Show that $\mu^* E = 0$ if $E$ contains no interval and that $\mu^* E = \infty$ if $E$ contains an interval.
   c. The restriction of $\mu^*$ to $\mathcal{B}$ is not a measure.
   Hence there is no smallest extension of $\mu$ to $\mathcal{B}$.
   d. The counting measure on $\mathcal{B}$ is an extension of $\mu$ to $\mathcal{B}$.
5. If $X$ is a metric space, $A$ contains all the open sets, and $f$ is continuous, then show that $f$ is measurable.

4.6 POINTS FOR DISCUSSION / CLARIFICATION

POINTS FOR DISCUSSION
4.7 Suggested Study material
International student edition.
5- Lecture Notes on Real analysis Richard F. Bass
Measures and outer measures

Introduction: In this Unit we first consider some of the ways in which a measure can be defined on a σ-algebra. In the case of Lebesgue measure we defined measure for open sets and used this to define outer measure, from which we obtain the notion of measurable set and Lebesgue measure. Such a procedure is feasible in general. First we discuss the process of deriving a measure from an outer measure, and later we derive an outer measure from a measure that is defined only on algebra of sets. The remainder of the unit is devoted to some applications of this process.

Objectives: The main aim of this unit is to understand the concept of Measures and outer measures, Extension of a measure, Uniqueness of Extension, Completion of a measure, Measure spaces, Integration with respect to a measure, Convergence in Measure, Almost uniform convergence, The \( L^p \)-spaces, Convex functions, Jensen's inequality, Holder and Minkowski inequalities, Completeness of \( L^p \).

5.1 Measure and outer measure

Extension of Measure

5.1.1 Definition: By an outer measure \( \mu^* \) we mean a nonnegative extended real-valued set function defined on all subsets of a space \( X \) and having the following properties:

(i) \( \mu^* \emptyset = 0 \).

(ii) \( A \subseteq B \Rightarrow \mu^* A \leq \mu^* B \)

(iii) \( E \subseteq \bigcup_{i=1}^{\infty} E_i \Rightarrow \mu^* E \leq \sum_{i=1}^{\infty} \mu^* E_i \)

The second property is called monotonicity and the third countable subadditivity.

In view of (i) finite subadditivity follows from (iii). Because of (ii), property (iii) can be replaced by

\[
E = \bigcup_{i=1}^{\infty} E_i, \text{disjo int } \Rightarrow \mu^* E \leq \sum_{i=1}^{\infty} \mu^* E_i
\]

The outer measure \( \mu^* \) is called finite if \( \mu^* X < \infty \).

By analogy with the case of Lebesgue measure we define a set \( E \) to be measurable with respect to \( \mu^* \) if for every set \( A \) we have

\[
\mu^* (A) = \mu^* (A \cap E) + \mu^* (A \cap \bar{E})
\]

Since \( \mu^* \) is subadditive, it is only necessary to show that

\[
\mu^* (A) \geq \mu^* (A \cap E) + \mu^* (A \cap \bar{E})
\]

for every \( A \) in order to prove that \( E \) is measurable. This inequality is trivially true when \( \mu^* A = \infty \), and so we need only establish it for sets \( A \) with \( \mu^* A \) finite.

5.1.2 Theorem: The class \( \beta \) of \( \mu^* \)-measurable sets is a σ-algebra. If \( \bar{\mu} \) is \( \mu^* \) restricted to \( \beta \), then \( \bar{\mu} \) is a complete measure on \( \beta \).

Proof: Trivially, the empty set is measurable. The symmetry of the definition of measurability in \( E \) and \( \bar{E} \) shows that \( \bar{E} \) is measurable whenever \( E \) is.

Let \( E_1 \) and \( E_2 \) be measurable sets. From the measurability of \( E_2 \),
\[ \mu^*(A) = \mu^*(A \cap E_2) + \mu^*(A \cap \tilde{E}_2) \] and
\[ \mu^*(A) = \mu^*(A \cap E_2) + \mu^*(A \cap \tilde{E}_2 \cap E_1) \mu^*(A \cap \tilde{E}_1 \cap \tilde{E}_2) \]
by the measurability of \( E_1 \). Since
\[ A \cap [E_1 \cup E_2] = [A \cap E_2] \cup [A \cap E_1 \cap \tilde{E}_2] \]
We have
\[ \mu^*(A \cap [E_1 \cup E_2]) \leq \mu^*[A \cap E_2] + \mu^*[A \cap \tilde{E}_2 \cap E_1] \]
by subadditivity, and so
\[ \mu^* A \geq \mu^*(A \cap [E_1 \cup E_2] + \mu^*[A \cap \tilde{E}_1 \cap \tilde{E}_2] \]
This means that \( E_1 \cup E_2 \) is measurable, since
\[ \sim [E_1 \cup E_2] = \tilde{E}_1 \cap \tilde{E}_2 \]
Thus the union of two measurable sets is measurable, and by induction the union of any finite number of measurable sets is measurable, so that \( \beta \) are algebra of sets.
Assume that \( E = \cup E_i \), where \( <E_i> \) is a disjoint sequence of measurable sets, and set
\[ G_n = \bigcup_{i=1}^{n} E_i \]
Then \( G_n \) is measurable, and
\[ \mu^*(A) = \mu^*(A \cap G_n) + \mu^*(A \cap \tilde{G}_n) \]
\[ \geq \mu^*(A \cap G_n) + \mu^*(A \cap \tilde{E}) \]
since \( E \subset G_n \). Now \( G_n \cap E_n = E_n \), and \( G_n \cap \tilde{E}_n = G_{n-1} \) and by measurability of \( \tilde{E}_n \) we have
\[ \mu^*(A \cap G_n) = \mu^*(A \cap \tilde{E}_n) + \mu^*(A \cap \tilde{G}_{n-1}) \]
By induction
\[ \mu^*(A \cap G_n) = \sum_{i=1}^{n} \mu^*(A \cap E_i) \]
and so
\[ \mu^*(A) \geq \mu^*(A \cap \tilde{E}) + \sum_{i=1}^{\infty} \mu^*(A \cap E_i) \]
\[ \mu^*(A) \geq \mu^*(A \cap \tilde{E}) + \mu^*(A \cap E) \]
since
\[ A \cap E \subset \bigcup_{i=1}^{\infty} (A \cap E_i) \]
Thus \( E \) is measurable. Since the union of any sequence of sets in algebra can be replaced by a disjoint union of sets in the algebra, it follows that \( \beta \) is a \( \sigma \)-algebra.
We next demonstrate the finite additivity of \( \mu \). Let \( E_1 \) and \( E_2 \) be disjoint measurable sets. Then the measurability of \( E_2 \) implies that
\[ \mu^*(E_1 \cup E_2) = \mu^*(E_1) + \mu^*(E_2) \]
\[= \mu^*([E_1 \cup E_2] \cap E_2) + \mu^*([E_1 \cup E_2] \cap \tilde{E}_2)\]
\[= \mu^* E_1 + \mu^* E_2\]

Finite additivity follows by induction.
If \(E\) is the disjoint union of the measurable sets \(\{E_i\}\), then
\[\bar{\mu}(E) \geq \sum_{i=1}^{\infty} \mu(E_i)\]
and so
\[\bar{\mu}(E) \geq \sum_{i=1}^{n} \mu(E_i)\]

But
\[\bar{\mu}(E) \leq \bar{\mu} \sum_{i=1}^{\infty} \mu(E_i)\]
\[1 = 1\]
by the subadditivity of \(\mu^*\). Hence \(\bar{\mu}\) is countably additive and thus a measure since it is nonnegative and \(\bar{\mu} \phi = \mu^* \phi = 0\).

**Uniqueness of Extension**

5.1.3 Definition: By a measure on an algebra we mean a nonnegative extended real-valued set function \(\mu\) defined on an algebra \(\mathfrak{A}\) of sets such that:

(i) \(\mu(\phi) = 0\).
(ii) If \(\langle A_i \rangle\) is a disjoint sequence of sets in \(\mathfrak{A}\) whose union is also in \(\mathfrak{A}\) then
\[\bar{\mu}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu A_i\]

Thus a measure on an algebra \(\mathfrak{A}\) is a measure if and only if \(\mathfrak{A}\) is \(\sigma\)-algebra.

The purpose of this section is to show that, if we start with a measure on an algebra \(\mathfrak{A}\) of sets, we may extend it to a Measure defined on a \(\sigma\)-algebra \(\beta\) containing \(\mathfrak{A}\). We shall do this by using the measure on the algebra to construct an outer measure \(\mu^*\) and show that the measure \(\bar{\mu}\) induced by \(\mu^*\) is the desired extension of \(\mu\). The process by which we construct \(\mu^*\) from \(\mu\) is analogous to that by which we constructed Lebesgue outer measure from the lengths of intervals: We define
\[\mu^* E = \inf \sum_{i=1}^{\infty} \mu A_i\]
where \(\langle A_i \rangle\) ranges over all sequences from \(\mathfrak{A}\) such that \(E \subseteq \bigcup_{i=1}^{\infty} A_i\).

We first state without proof some lemmas, propositions concerning \(\mu^*\) to prove extension theorem

5.1.4 Lemma: If \(A \in \mathfrak{A}\) and if \(\langle A_i \rangle\) is any sequence of sets in \(\mathfrak{A}\) such that \(A \subseteq \bigcup_{i=1}^{\infty} A_i\), then
\[\mu A \leq \inf \sum_{i=1}^{\infty} \mu A_i\]

5.1.5 Corollary: If \(A \in \mathfrak{A}\) then \(\mu^* A = \mu A\)
5.1.6 Lemma: The set function $\mu^*$ is an outer measure.

5.1.7 Lemma: If $A \in \mathcal{A}$ then $A$ is measurable with respect to $\mu^*$.

5.1.8 Proposition: Let $\mu$ be a measure on an algebra $\mathcal{A}$, $\mu^*$ the outer Measure induced by $\mu$, and $E$ any set. Then for $\epsilon > 0$, there is a set $A \in \mathcal{A}$ with $E \subset A$ and $\mu^*A < \mu^*E + \epsilon$.

There is also a set $B \in \mathcal{A}_{\sigma\delta}$ with $E \subset B$ and $\mu^*E = \mu^*B$.

5.1.9 Proposition: Let $\mu$ be a $\sigma$-finite measure on an algebra $\mathcal{A}$, and let $\mu^*$ be the outer measure generated by $\mu$. A set $E$ is $\mu^*$ measurable if and only if $E$ is the proper difference $A \sim B$ of a set $A$ in $\mathcal{A}_{\sigma\delta}$ and a set $B$ with $\mu^*B = 0$. Each set $B$ with $\mu^*B = 0$ is contained in a set $C$ in $\mathcal{A}_{\sigma\delta}$ with $\mu^*C = 0$.

Proof: The "if" part of the proposition follows from the fact that each set in $\mathcal{A}_{\sigma\delta}$ must be measurable, since the measurable sets form $\sigma$-algebra, while each set of $\mu^*$-measure zero must be measurable, since $\overline{\mu}$ is complete.

To prove the "only if" part of the proposition, let $\{X_i\}$ be a countable disjoint collection of sets in $\mathcal{A}$ with $\mu X_i$ finite and $X = \bigcup X_i$. If $E$ is measurable, then $E$ is the disjoint union of the measurable sets $E_i = X_i \cap E$. By Proposition 5.1.8 we can find for each positive integer $n$ a set $A_{ni}$ in $\mathcal{A}_{\sigma\delta}$ such that $E_i \subset A_{ni}$ and

$$\overline{\mu}(A_{ni}) \leq \mu E_i + \frac{1}{n2^i}.$$ 

Set

$$A_n = \bigcup_{i=1}^{\infty} A_{ni}.$$ 

Then

$$E \subset A_n$$

and $A_n \sim E \subset \bigcup_{i=1}^{\infty} A_{ni} \sim E_i$. Hence

$$\overline{\mu}(A_n \sim E) \leq \sum_{i=1}^{\infty} \overline{\mu}(A_{ni} \sim E_i) \leq \sum_{i=1}^{\infty} \frac{1}{n2^i} = \frac{1}{n}.$$ 

Since $A_n \in \mathcal{A}_{\sigma}$ the set $A = \bigcap_{n=1}^{\infty} A_n$ is in $\mathcal{A}_{\sigma\delta}$ and for each $n$

$$A \sim E \subset A_n \sim E$$

Hence

$$\overline{\mu}(A \sim E) \subset \overline{\mu}(A_n \sim E) \leq \frac{1}{n}.$$ 

Since this holds for each positive integer $n$, we must have $\overline{\mu}(A \sim E) = 0$.

We summarize the results of this section in the following theorem.

5.1.10 Theorem: (Caratheodory): Let $\mu$ be a measure on an algebra $\mathcal{A}$, and $\mu^*$ the outer measure induced by $\mu$. Then the restriction $\overline{\mu}$ of $\mu^*$ to the $\mu^*$-measurable sets is an extension of $\mu$ to a $\sigma$-
algebra containing $\hat{a}$. If $\mu$ is finite (or $\sigma$-finite) so is $\overline{\mu}$. If $\mu$ is $\sigma$-finite, then $\overline{\mu}$ is the only measure on the smallest $\sigma$-algebra containing $\hat{a}$, which is an extension of $\mu$.

**Proof:** The fact that $\overline{\mu}$ is an extension of $\mu$ from $\hat{a}$ to be a measure on a $\sigma$-algebra containing $\hat{a}$ follows directly from Corollary 5.1.5, lemma 5.1.7 and Theorem 5.1.2. Unicity of $\overline{\mu}$ is left as an exercise for learner.

**Completeness of a Measure**

5.1.11 **Definition:** If $\mu$ is a measure, we say a sequence of measurable functions $f_n$ converges to $f$ almost everywhere (written $f_n \rightarrow f$ a.e.) if there is a set of measure $0$ and for $x$ not in this set we have $f_n(x) \rightarrow f(x)$.

We say $f_n$ converges to $f$ in measure if for each $\epsilon > 0$

$$\mu(\{x : |f_n(x) - f(x)| > \epsilon\}) \rightarrow 0$$

as $n \rightarrow \infty$.

5.1.12 **Proposition:** Suppose $\mu$ is a finite measure.

(a) If $f_n \rightarrow f$, a.e., then $f_n$ converges to $f$ in measure.

(b) If $f_n \rightarrow f$ in measure, there is a subsequence $n_j$ such that $f_{n_j} \rightarrow f$, a.e.

**Proof.** Let $\epsilon > 0$. If $A_n = \{x : |f_n(x) - f(x)| > \epsilon\}$, then $\chi_{A_n} \rightarrow 0$ a.e., and by dominated convergence,

$$\mu(A_n) = \int \chi_{A_n}(x) \, \mu(dx) \rightarrow 0.$$

This proves (a). To prove (b), let $n_1 = 1$ and choose $n_j > n_{j-1}$ inductively so that

$$\mu(\{x : |f_{n_j}(x) - f(x)| > 1/j\}) \leq 2^{-j}.$$

Let $A_j = \{x : |f_{n_j}(x) - f(x)| > 1/j\}$. Then $\mu(A_j) \leq 2^{-j}$, and $A = \bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_j$ has measure less than $\bigcup_{j=k}^{\infty} A_j$ for every $k$, hence less than $2^{-k+1}$ for every $k$. Therefore $A$ has measure $0$.

If $x \not\in A$, then $x \not\in \bigcup_{j=k}^{\infty} A_j$ for some $k$, so

$$|f_{n_j}(x) - f(x)| \leq 1/j$$

for $j \geq k$, which means $f_{n_j} \rightarrow f$ a.e. on $A^c$.

**Example1.** Part (a) of the above proposition is not true if $\mu(X) = \infty$.

Let $X = \mathbb{R}$ and let $f_n = \chi_{[n,n+1]}$. 

**Example2.** (An example where $f_n \rightarrow f$ in measure but not almost everywhere)

Let $X = [0, 1]$, let $\mu$ be Lebesgue measure, and let $f_n(x) = \chi_{F_n}(x)$, where

$$F_n = \{y : (\sum_{j=1}^{n} 1/j)(\mod 1) \leq y \leq (\sum_{j=1}^{n} 1/j)(\mod 1)\}.$$ 

$z(\mod 1)$ is defined as the fractional part of $z$ (where the largest integer less than $z$ is subtracted from $z$).

Let $f(x) = 0$ for all $x$.

Then $\mu(F_n) \leq 1/n \rightarrow 0$, so $f_n \rightarrow f$ in measure.

But any $x$ will be in infinitely many $F_n$’s, so $f_n$ does not converge to $f(x)$ at any point.

**Egoroff’s theorem.**

5.1.13 **Theorem:** If $\mu$ is a finite measure, $\epsilon > 0$, and $f_n \rightarrow f$ a.e., then there exists a measurable set $A$ such that $\mu(A) < \epsilon$ and $f_n \rightarrow f$ uniformly on $A^c$.

This type of convergence is sometimes known as almost uniform convergence.

**Proof.** Let
$E_{nk} = \bigcup_{m=n}^{\infty} \{ x : |f_m(x) - f(x)| > 1/k \}$.

For fixed $k$, $E_{nk}$ decreases as $n$ increases, and the intersection $\cap_n E_{nk}$ has measure 0.

So $\mu(E_{nk}) \to 0$.

Then there exists an integer $n_k$ such that $\mu(E_{nk}) < \varepsilon 2^{-k}$.

Let $E = \bigcup_{k=1}^{\infty} E_{nk}$. Then $\mu(E) < \varepsilon$, and if $x \not\in E$ and $n > n_k$, then $|f_n(x) - f(x)| \leq 1/k$.

Thus $f_n \to f$ uniformly on $E$.

**Product measures**

If $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_{i=1}^{\infty} A_i$, we write $A_i \uparrow A$.

If $A_1 \supset A_2 \supset \cdots$ and $A = \bigcap_{i=1}^{\infty} A_i$, we write $A_i \downarrow A$.

5.1.14 **Definition:** $M$ is a monotone class if $M$ is a collection of subsets of $X$ such that
(a) if $A_i \uparrow A$ and each $A_i \in M$, then $A \in M$;
(b) if $A_i \downarrow A$ and each $A_i \in M$, then $A \in M$.

The intersection of monotone classes is a monotone class, and the intersection of all monotone classes containing a given collection of sets is the smallest monotone class containing that collection. The next theorem, the monotone class lemma, is rather technical, but very useful.

5.1.15 **Theorem:** Suppose $A_0$ is an algebra, $A$ is the smallest $\sigma$-algebra containing $A_0$, and $M$ is the smallest monotone class containing $A_0$. Then $M = A$.

**Proof.** $A$ is a $\sigma$-algebra, so $M \subseteq A$. We must show $A \subseteq M$.

Let $N_1 = \{ A \in M : A^c \in M \}$. Note $N_1$ is contained in $M$, contains $A_0$, and is a monotone class. So $N_1 = M$, and therefore $M$ is closed under the operation of taking complements.

Let $N_2 = \{ A \in M : A \cap B \in M \text{ for all } B \in A_0 \}$. $N_2$ is contained in $M$; $N_2$ contains $A_0$ because $A_0$ is an algebra; $N_2$ is a monotone class because

$(\bigcup_{i=1}^{\infty} A_i) \cap B = \bigcup_{i=1}^{\infty} (A_i \cap B)$, and similarly for intersections.

$(\bigcap_{i=1}^{\infty} A_i) \cap B = \bigcap_{i=1}^{\infty} (A_i \cap B)$.

Therefore $N_2 = M$; in other words, if $B \in A_0$ and $A \in M$, then $A \cap B \in M$.

Let $N_3 = \{ A \in M : A \cap B \in M \text{ for all } B \in M \}$. $N_3$ is a monotone class contained in $M$. By the last sentence of the preceding paragraph, $N_3$ contains $A_0$. Hence $N_3 = M$.

We thus have that $M$ is a monotone class closed under the operations of taking complements and taking intersections. This shows $M$ is a $\sigma$-algebra, and so $A \subseteq M$.

Suppose $(X, A, \mu)$ and $(Y, B, \nu)$ are two measure spaces, i.e., $A$ and $B$ are $\sigma$-algebras on $X$ and $Y$, respectively, and $\mu$ and $\nu$ are measures on $A$ and $B$, respectively. Then $A \times B$ is a set of the form $A \times B$, where $A \in A$ and $B \in B$. Define a set function $\mu \times \nu$ on rectangles by

$\mu \times \nu(A \times B) = \mu(A)\nu(B)$.

5.1.16 **Lemma:** Suppose

$A \times B = \bigcup_{i=1}^{\infty} A_i \times B_i$,

where $A, A_i \in A$ and $B, B_i \in B$ and the $A_i \times B_i$ are disjoint.
Then
\[ \mu \times \nu(A \times B) = \sum_{i=1}^{\infty} \mu \times \nu(A_i \times B_i). \]

**Proof.** We have
\[ \chi_{A \times B}(x, y) = \sum_{i=1}^{\infty} \chi_{A_i \times B_i}(x, y), \]
and so
\[ \chi_A(x)\chi_B(y) = \sum_{i=1}^{\infty} \chi_{A_i}(x)\chi_{B_i}(y). \]
We know, if \( f_n \) is a sequence of nonnegative measurable functions with
\[ f_1(x) \leq f_2(x) \leq \cdots \]
for all \( x \) and with \( \lim_{n \to \infty} f_n(x) = f(x) \) for all \( x \).

Then \( \int f_n \, d\mu \to \int f \, d\mu. \)

Holding \( x \) fixed and integrating over \( y \) with respect to \( \nu \) Hence
\[ \chi_A(x)\nu(B) = \sum_{i=1}^{\infty} \chi_{A_i}(x)\nu(B_i) \]
Now use as above again and integrate over \( x \) with respect to \( \mu \) we obtain the result.

**Remark:** Let \( C_0 = \{ \text{finite unions of rectangles} \} \). It is clear that \( C_0 \) is algebra. By linearity, we see that \( \mu \times \nu \) is a measure on \( C_0 \).

Let \( A \times B \) be the smallest \( \sigma \)-algebra containing \( C_0 \); this is called the product \( \sigma \)-algebra. By the Carathéodory extension theorem, \( \mu \times \nu \) can be extended to a measure on \( A \times B \).

We will need the following observation.
Suppose a measure \( \mu \) is \( \sigma \)-finite. So there exist \( E_i \) which have finite \( \mu \) measure and whose union is \( X \). If we let
\[ F = \bigcup_{i=1}^{n} E_i \], then \( F_i \uparrow X \) and \( \mu(F_n) \) is finite for each \( n \).
If \( \mu \) and \( \nu \) are both \( \sigma \)-finite, say with \( F_i \uparrow X \) and \( G_i \uparrow Y \), then \( \mu \times \nu \) will be \( \sigma \)-finite, using the sets \( F_i \times G_i \).

The main result of this section is Fubini’s theorem, which allows one to interchange the order of integration.

**5.1.17 Theorem:** Suppose \( f : X \times Y \to \mathbb{R} \) is measurable with respect to \( A \times B \). If \( f \) is nonnegative or

\[ \int |f(x, y)| \, d(\mu \times \nu)(x, y) < \infty, \]

then
(a) the function \( \int g(x) = f(x, y)\nu(dy) \) is measurable with respect to \( A \);
(b) the function \( h(y) = f(x, y)\mu(dx) \) is measurable with respect to \( B \);
(c) we have
\[ \int f(x, y) \, d(\mu \times \nu)(x, y) = \int \left( \int f(x, y) \, d\mu(x) \right) \nu(dy) \]
\[ = \int \left( \int f(x, y) \, dv(y) \right) \mu(dx). \]

**Proof.** First suppose \( \mu \) and \( \nu \) are finite measures. If \( f \) is the characteristic function of a rectangle, then proof of (a) – (c) are obvious. By linearity, (a) – (c) hold if \( f \) is the characteristic function of a set in \( C_0 \), the set of finite unions of rectangles.

Let \( M \) be the collection of sets \( C \) such that (a) – (c) hold for \( \chi_C \).
If \( C_i \uparrow C \) and \( C_i \in M \), then (c) holds for \( \chi_C \), by monotone convergence.
If \( C_i \downarrow C \), then (c) holds for \( \chi_C \) by dominated convergence. (a) and (b) are easy.
So $M$ is a monotone class containing $A_0$, so $M = A \times B$.

If $\mu$ and $\nu$ are $\sigma$-finite, applying monotone convergence to $C \cap (F_n \times G_n)$ for suitable $F_n$ and $G_n$ and monotone convergence, we see that (a) – (c) holds for the characteristic functions of sets in $A \times B$ in this case as well.

By linearity, (a) – (c) hold for nonnegative simple functions.

By monotone convergence, (a) – (c) hold for nonnegative functions.

In the case $\int |f| < \infty$, writing $f = f^+ - f^-$ and using linearity proves (a) – (c) for this case, too.

5.2 Measure Spaces

The purpose of the section is to abstract the most important properties of Lebesgue measure and Lebesgue integration. We shall do this by giving certain axioms which Lebesgue measure satisfies and base our integration theory on these axioms. As a consequence our theory will be valid for every system satisfying the given axioms. We begin by recalling that a $\sigma$-algebra $\beta$ is a family of subsets of a given set $X$ which contains 0 and is closed with respect to complements and with respect to countable unions. By a set function $\mu$ we mean a function which assigns an extended real number to certain sets. With this in mind we make the following definitions:

5.2.1 Definition: By a measurable space we mean a couple $(X, \beta)$ consisting of a set $X$ and $\sigma$-algebra $\beta$ of subsets of $X$. A subset $A$ of $X$ is called measurable (or measurable with respect to $\beta$) if $A \in \beta$.

5.2.2 Definition: By a measure $\mu$ on a measurable space $(X, \beta)$ we mean a nonnegative set function defined for all sets of $\beta$ and satisfying $\mu(\emptyset) = 0$ and $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu E_i$ for any sequence $E_i$, of disjoint measurable sets.

By a measure space $(X, \beta, \mu)$ we mean a measurable space $(X, \beta)$ together with a measure $\mu$ defined on $\beta$.

5.2.3 The Radon-Nikodym Theorem: Suppose $f$ is nonnegative, measurable, and integrable with respect to $\mu$. If we define $\nu$ by

$$\nu(A) = \int_A f \, d\mu, \quad (5.1)$$

then $\nu$ is a measure. The only part that needs thought is the countable additivity, and this follows from (7.1) applied to the functions $\int \chi_{A_i}$. Moreover, $\nu(A)$ is zero whenever $\mu(A)$ is. We sometimes write $f = d\nu/d\mu$ for (5.1).

5.2.4 Definition: A measure $\nu$ is called absolutely continuous with respect to a measure $\mu$ if $\nu(A) = 0$ whenever $\mu(A) = 0$. This is frequently written $\nu \ll \mu$.

5.2.5 Proposition: A finite measure $\nu$ is absolutely continuous with respect to $\mu$ if and only if for all $\varepsilon$ there exists $\delta$ such that $\mu(A) < \delta$ implies $\nu(A) < \varepsilon$.

Proof. If the condition given in the statement of the proposition holds, it is clear that $\nu \ll \mu$. Suppose now that $\nu \ll \mu$. If the condition does not hold, there exists $E_k$ such that $\mu(E_k) < 2^{-k}$ but $\nu(E_k) \geq \varepsilon$. Let $F = \bigcup_{n=1}^{\infty} \bigcap_{k=1}^{n} E_k$. Then
\[
\mu(F) = \lim_{n \to \infty} \mu\left(\bigcup_{k=1}^{n} E_k\right) \leq \lim_{n \to \infty} \sum_{k=n}^{\infty} 2^{-k} = 0,
\]

but
\[
\nu(F) = \lim_{n \to \infty} \nu\left(\bigcup_{k=1}^{n} E_k\right) \geq \varepsilon
\]

This contradicts the absolute continuity.

**5.2.6 Definition:** A function \( \mu : A \to (\mathbb{R}, \infty] \) is called a signed measure if \( \mu(\emptyset) = 0 \) and \( \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \) whenever the \( A_i \) are disjoint and all the \( A_i \) are in \( A \).

**5.2.7 Definition:** Let \( \mu \) be a signed measure. A set \( A \in A \) is called a positive set for \( \mu \) if \( \mu(B) \geq 0 \) whenever \( B \subset A \) and \( B \in A \). We define a negative set \( A \in A \) is called a positive set for \( \mu \) if \( \mu(B) \leq 0 \) whenever \( B \subset A \) and \( B \in A \). A null set \( A \) is one where \( \mu(B) = 0 \) whenever \( B \subset A \).

**Example 3.** Suppose \( m \) is Lebesgue measure and \( \mu(A) = \int f \, dm \) for some integrable \( f \). If we let \( P = \{x : f(x) \geq 0\} \), then \( P \) is easily seen to be a positive set, and if \( N = \{x : f(x) < 0\} \), then \( N \) is a negative one. The Hahn decomposition which we give below is a decomposition of our space (in this case\( \mathbb{R} \)) into positive and negative sets. This decomposition is unique, except that \( C = \{x : f(x) = 0\} \) could be included in \( N \) instead of \( P \), or apportioned partially to \( P \) and partially to \( N \).

**Note:** however, that \( C \) is a null set. The Jordan decomposition below is a decomposition of \( \mu \) into \( \mu^+ \) and \( \mu^- \), where \( \mu^+(A) = \int_A f^+ \, dm \), and similarly \( \mu^-(A) = \int_A f^- \, dm \). If \( \mu \) is a signed measure, then
\[
\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \to \infty} \mu\left(\bigcup_{i=1}^{n} A_i\right). \text{ The proof is the same as in the case of positive measures.}
\]

**5.2.8 Proposition:** Let \( \mu \) be a signed measure taking values in \((\mathbb{R}, \infty]\). Let \( E \) be measurable with \( \mu(E) < 0 \). Then there exist a subset \( F \) of \( E \) that is negative set with \( \mu(F) < 0 \).

**Proof.** If \( E \) is a negative set, we are done. If not, there exists a subset with positive measure. Let \( n_1 \) be the smallest positive integer such that there exists \( E_1 \subset E \) with \( \mu(E_1) \geq 1/n_1 \). Let \( k \geq 2 \). If \( F_k = E - (E_1 \cup \cdots \cup E_{k-1}) \) is negative, we are done. If not, let \( n_k \) be the smallest positive integer such that there exists \( E_k \subset F_k \) with \( \mu(E_k) \geq 1/n_k \). We continue. If the construction stops after a finite number of sets, we are done. If not, let \( F = \cap_k F_k = E - (\cup_k E_k) \). Since \( 0 > \mu(E) > -\infty \) and \( \mu(E_k) \geq 0 \), then
\[
\mu(E) = \mu(F) + \sum_{k=1}^{\infty} \mu(E_k).
\]

Then \( \mu(F) \leq \mu(E) < 0 \), so the sum converges. If \( G \subset F \) is measurable with \( \mu(G) > 0 \), then \( \mu(G) \geq 1/N \) for some \( N \), which contradicts the construction. Therefore \( F \) must be a negative set.

We write \( A \Delta B \) for \( (A - B) \cup (B - A) \). The following is known as the Hahn decomposition theorem.

**5.2.9 Theorem:** Let \( \mu \) be a signed measure taking values in \((\mathbb{R}, \infty]\). There exist sets \( E \) and \( F \) in \( A \) that are disjoint whose union is \( X \) and such that \( E \) is a negative set and \( F \) is a positive set. If \( E \) and \( F \) are another such pair, then \( E \Delta E' = F \Delta F' \) is a null set with respect to \( \mu \).

**Proof.** Let \( L = \inf\{\mu(A) : A \text{ is a negative set}\} \). Choose negative sets \( A_n \) such that \( \mu(A_n) \to L \). Let \( E = \bigcup_{n=1}^{\infty} A_n \).
Let $B_n = A_n - (B_1 \cup \cdots \cup B_{n-1})$ for $n = 1$ each $n$. Since $A_n$ is a negative set, so is each $B_n$. Also, the $B_n$ are disjoint. If $C \subset E$, then

$$
\mu(C) = \lim_{n \to \infty} \mu(C \cap (\bigcup_{i=1}^{n} B_i)) = \lim_{n \to \infty} \sum_{i=1}^{\infty} \mu(C \cap B_i) \leq 0.
$$

So $E$ is a negative set.

Since $E$ is negative,

$$
\mu(E) = \mu(A_n) + \mu(E - A_n) \leq \mu(A_n).
$$

Letting $n \to \infty$, we obtain $\mu(E) = L$.

Let $F = E^c$. If $F$ were not a positive set, there would exist $B \subset F$ with $\mu(B) < 0$. By Proposition 11.6 there exists a negative set $C$ contained in $B$ with $\mu(C) < 0$. But then $E \cup C$ would be a negative set with $\mu(E \cup C) < \mu(E) = L$, a contradiction.

To prove uniqueness, if $E'$, $F'$ are another such pair of sets and $A \subset E - E' \subset E$, then $\mu(A) \leq 0$. But $A \subset E - E' = F' - F \subset F'$, so $\mu(A) \leq 0$. Therefore $\mu(A) = 0$. The same argument works if $A \subset E - E'$, and any subset of $E \Delta E'$ can be written as the union of $A_1$ and $A_2$, where $A_1 \subset E - E'$ and $A_2 \subset E' - E$.

Let us say two measures $\mu$ and $\nu$ are mutually singular if there exist two disjoint sets $E$ and $F$ in $A$ whose union is $X$ with $\mu(E) = \nu(F) = 0$. This is often written $\mu \perp \nu$.

**Example 4.** If $\mu$ is Lebesgue measure restricted to $[0, 1/2]$, that is, $\mu(A) = m(A \cap [0, 1/2])$, and $\nu$ is Lebesgue measure restricted to $[1/2, 1]$, then $\mu$ and $\nu$ are mutually singular.

We let $E = [0, 1/2]$ and $F = (1/2, 1]$. This example works because the Lebesgue measure of $\{1/2\}$ is 0.

**Example 5.** A more interesting example is the following. Let $f$ be the Cantor-Lebesgue function and let $\nu$ be the Lebesgue-Stieltjes measure associated with $f$. Let $\mu$ be Lebesgue measure restricted to $[0, 1]$. Then $\mu \perp \nu$. To see this, we let $E = C$, where $C$ is the Cantor set, and $F = [0, 1] - C$. We already know that $m(E) = 0$ and we need to show $\nu(F) = 0$.

To do that, we need to show $\nu(I) = 0$ for every open interval contained in $F$. This will follow if we show $\nu(J) = 0$ for every interval of the form $J = (a, b]$ contained in $F$. But $f$ is constant on every such interval, so $f(b) = f(a)$, and therefore $\nu(J) = f(b) - f(a) = 0$. The following is known as the Jordan decomposition theorem.

**5.2.10 Theorem:** If $\mu$ is a signed measure, there exist measures $\mu^+$ and $\mu^-$ such that $\mu = \mu^+ - \mu^-$ and $\mu^+$ and $\mu^-$ are mutually singular. This decomposition is unique.

**Proof.** Let $E$ and $F$ be positive and negative sets for $\mu$ and let $\mu^+(A) = \mu(E \cap A)$, $\mu^-(A) = -\mu(A \cap F)$.

This gives the desired decomposition.

If $\mu = \nu^+ - \nu^-$ is another such decomposition with $\nu^+$, $\nu^-$ mutually singular, let $E'$ and $F'$ be the sets in the definition of mutually singular. Then $X = E' \cup F'$ gives another Hahn decomposition, hence $E \Delta E'$ is a null set with respect to $\mu$. Then for any $A \in A$, $\nu^+(A) = \mu(A \cap E') = \mu^+(A)$, and similarly for $\nu^-, \mu^-$. 

The measure $\mu^+ + \mu^-$ is called the total variation measure and is written $|\mu|$.

**Convergence in Measure**

We now are ready for the **Radon-Nikodym theorem**.

5.2.11 Theorem: Suppose $\mu$ is a $\sigma$-finite measure and $\nu$ is a finite measure such that $\nu$ is absolutely continuous with respect to $\mu$. There exists a $\mu$-integrable nonnegative function $f$ such that $\nu(A) = \int_A f \, d\mu$ for all $A \in \mathcal{A}$. Moreover, if $g$ is another such function, then $f = g$ almost everywhere with respect to $\mu$.

**Proof.** Let us first prove the uniqueness assertion. For every set $A$ we have

$$\int_A (f - g) \, d\mu = \nu(A) - \nu(A) = 0.$$  

A By Proposition 8.1 we have $f - g = 0$ a.e. with respect to $\mu$.

Since $\mu$ is $\sigma$-finite, there exist $F_i \uparrow X$ such that $\mu(F_i) < \infty$ for each $i$. Let $\mu_i$ be the restriction of $\mu$ to $F_i$, that is, $\mu_i(A) = \mu(A \cap F_i)$. Define $\nu_i$, the restriction of $\nu$ to $F_i$, similarly. If $f_i$ is a function such that $\nu_i(A) = \int_A f_i \, d\mu_i$ for all $A$, the argument of the first paragraph shows that $f_i = f_j$ on $F_i$ if $i \leq j$. If we define $f$ by $f(x) = f_i(x)$ if $x \in F_i$, we see that $f$ will be the desired function.

So it suffices to restrict attention to cases where $\mu$ is finite.

Let $F = \{ g : 0 \leq g, \int_A g \, d\mu \leq \nu(A) \text{ for all } A \in \mathcal{A} \}$.

$F$ is not empty because $0 \in F$.

Let $L = \sup \{ \int_A g \, d\mu : g \in F \}$, and let $g_n$ be a sequence in $F$ such that $\int_A g_n \, d\mu \to L$.

Let $h_n = \max(g_1, \ldots, g_n)$.

If $g_1$ and $g_2$ are in $F$, then $h_2 = \max(g_1, g_2)$ is also in $F$.

To see this, let $B = \{ x : g_1(x) \geq g_2(x) \}$, and write

$$\int_A h_2 \, d\mu = \int_{A \cap B} h_2 \, d\mu + \int_{A \setminus B} h_2 \, d\mu$$

$$= \int_{A \cap B} g_1 \, d\mu + \int_{A \setminus B} g_2 \, d\mu$$

$$\leq \nu(A \cap B) + \nu(A \cap B^c) = \nu(A).$$

By an induction argument, $h_n$ is in $F$.

The $h_n$ increase, say to $f$. By monotone convergence $\int_A f \, d\mu = L$ and

$$\int_A f \, d\mu \leq \nu(A) \quad \text{(5.2)}$$

for all $A$.

Let $A$ be a set where there is strict inequality in (11.2); let $\varepsilon$ be chosen sufficiently small so that if $\pi$ is defined by

$$\pi(B) = \nu(B) - \int_B f \, d\mu - \varepsilon \mu(B),$$

then $\pi(A) > 0$.

$\pi$ is a signed measure; let $F$ be the positive. In particular, $\pi(F) > 0$.

So for every $B$
\[
\int_{B \cap F} f \, d\mu + \varepsilon \mu(B \cap F) \leq \nu(B \cap F).\]
We then have, using (11.2), that
\[
\int_B (f + \varepsilon \chi_F) \, d\mu = \int_B f \, d\mu + \varepsilon \mu(B \cap F)
\]
\[
= \int_{B \cap F} f \, d\mu + \int_{B \cap F} f \, d\mu + \varepsilon \mu(B \cap F) 
\]
\[
\leq \nu(B \cap F^c) + \nu(B \cap F) = \nu(B).
\]
This says that \( f + \varepsilon \chi_F \in F \). However,
\[
L \geq \int (f + \varepsilon \chi_F) \, d\mu = \int f \, d\mu + \varepsilon \mu(F) = L + \varepsilon \mu(F),
\]
which implies \( \mu(F) = 0 \). But then \( \nu(F) = 0 \), and hence \( \pi(F) = 0 \), contradicting the fact that \( F \) is a positive set for \( F \) with \( \pi(F) > 0 \).

The proof of \textbf{Lebesgue decomposition theorem} is the almost same.

\textbf{5.2.12 Theorem:} Suppose \( \mu \) and \( \nu \) are two finite measures. There exist measures \( \lambda \), \( \rho \) such that \( \nu = \lambda + \rho \), \( \rho \) is absolutely continuous with respect to \( \mu \), and \( \lambda \) and \( \mu \) are mutually singular.

\section*{5.3 \textbf{L}^p \textbf{S}paces}

\textbf{Convex Functions}

\textbf{5.3.1 Definition:} A function \( \varphi \) defined on an open interval \( (a, b) \) is said to be convex if for each \( x, y \in (a, b) \) and each \( \lambda, 0 \leq \lambda \leq 1 \) we have
\[
\varphi(\lambda x + (1-\lambda)y) \leq \lambda \varphi(x) + (1-\lambda) \varphi(y).
\]
If we look at the graph of \( \varphi \) in \( \mathbb{R}^2 \), this condition can be formulated geometrically by saying that each point on the chord between \( (x, \varphi(x)) \) and \( (y, \varphi(y)) \) is above the graph of \( \varphi \). An important property of the chords of a convex function is given by the following lemma, whose proof is left to the reader.

\textbf{5.3.2 Lemma:} If \( \varphi \) is convex on \( (a, b) \) and if \( x, y, x', y' \) are points of \( (a, b) \) with \( x \leq x' < y' < y \) then the chord over \( (x', y') \) has larger slope than the chord over \( (x, y) \); that is,
\[
\frac{\varphi(y) - \varphi(x)}{y-x} \leq \frac{\varphi(y') - \varphi(x')}{y'-x'}.
\]

If the upper and lower left-hand derivates \( D^-f \) and \( D^+f \) of a function \( f \) are equal and finite at a point \( x \), we say that \( f \) is differentiable on the left at \( x \) and call this common value the left-hand derivative at \( x \). Similarly, we say that \( f \) is differentiable on the right at \( x \) if \( D^+f \) and \( D^-f \) are equal there. Some of the continuity and differentiability properties of convex functions are given by the following proposition.

\textbf{5.3.3 Proposition:} If \( \varphi \) is convex on \( (a, b) \), then \( \varphi \) is absolutely continuous on each closed subinterval of \( (a, b) \). The right- and left-hand derivatives of \( \varphi \) exist at each point of \( (a, b) \) and are equal to each other except on a countable set. The right- and left-hand derivatives of \( \varphi \) are monotone increasing functions, and each point the left-hand derivatives of \( \varphi \) is less or equal to the right-hand derivatives of \( \varphi \).

\textbf{5.3.4 Corollary:} Let \( \varphi \) have a second derivative at each point of \( (a, b) \). Then \( \varphi \) is convex on \( (a, b) \) if and only if \( \varphi''(x) > 0 \) for each \( x \in (a, b) \).

\textbf{Jensen Inequality}

This notion enables us to give a short proof for the following proposition:

\textbf{5.3.5 Proposition:} (Jensen Inequality): Let \( \varphi \) be a convex function on \( (-\infty, \infty) \) and \( f \) an integrable function on \([0, 1]\). Then
Let $\alpha = \int f(t) \, dt$, and let $y = m( x - \alpha) + \varphi(\alpha)$ be the equation of a supporting line at $\alpha$. Then
\[
\varphi( f(t)) \geq m( f(t) - \alpha) + \varphi(\alpha)
\]
Integrating both sides with respect to $t$ gives the proposition.

This inequality has a geometric interpretation worth mentioning. Since the point $\lambda x_1 + (1 - \lambda) x_2$ is the centroid of masses $\lambda$ and $(1 - \lambda)$ at $x_1$, and $x_2$, we can say that a function $f$ is convex if its value at the centroid of a two-point mass is less than the weighted average of its values at the two points. The Jensen inequality is a generalization of this fact: If we define a mass distribution $\mu$ on the line by setting
\[
\int m( x) \, d\mu(x) = m( \alpha)
\]
and $\int \varphi(x) \, d\mu(x) = \varphi(\alpha)$, then
\[
\int \varphi( f(t)) \, dt \leq \int \varphi(x) \, d\mu
\]
This inequality then becomes a generalization of the inequality between arithmetic and geometric mean.

**Holder’s Minkowski’s inequality**

We assume throughout this section that the measure is $\sigma$-finite. For $1 \leq p < \infty$, define the $L^p$ norm of $f$ by
\[
\|f\|_p = \left( \int |f(x)|^p \, d\mu \right)^{1/p}
\]
For $p = \infty$, define the $L^\infty$ norm of $f$ by
\[
\|f\|_\infty = \inf \{ M : \mu( \{ x : |f(x)| \geq M \}) = 0 \}
\]
For $1 \leq p \leq \infty$, the space $L^p$ is the set $\{f : \|f\|_p < \infty\}$. The $L^\infty$ norm of a function $f$ is the supremum of $f$ provided we disregard sets of measure 0.

It is clear that $\|f\|_p = 0$ if and only if $f = 0$ a.e.

**5.3.6 Proposition:** (Holder’s inequality) If $1 < p, q < \infty$ and $p^{-1} + q^{-1} = 1$ then
\[
\int |f(x)g(x)| \, d\mu \leq \|f\|_p \|g\|_q
\]
This also holds if $p = \infty$ and $g = 1$

**Proof:** If $M = \|f\|_\infty$, then $\int |f| \leq M \|g\|$ and the case $p = \infty$ and $q = 1$ follows. So let us assume $1 < p, q < \infty$. If $\|f\|_p = 0$, then $f = 0$ a.e. and $\int fg = 0$, so the result is clear if $\|f\|_p = 0$ and similar if $\|g\|_q = 0$. Let $F(x) = |f(x)|$, $\|f\|_p$ and $G(x) = |g(x)|$, $\|g\|_q$. Note $\|F\|_p = 1$, $\|G\|_q = 1$, and it suffices to show that $\int FG \leq 1$.

The second derivative of the function $e^x$ is again $e^x$, which is positive, and so $e^x$ is convex. Therefore if $0 \leq \lambda \leq 1$, we have
\[
e^{\lambda a + (1 - \lambda)b} \leq \lambda e^a + (1 - \lambda)e^b.
\]
If $F(x), G(x) \neq 0$, let $a = p \log F(x)$, $b = q \log G(x)$, $\lambda = 1/p$, and $1 - \lambda = 1/q$. We then obtain
\[ F(x) G(x) \leq \frac{F(x)^p}{p} + \frac{G(x)^q}{q} \]

Clear this inequality also holds of \( F(x) = 0 \) or \( G(x) = 0 \). Integrating,

\[
\int F G \leq \frac{\|F\|^p_p}{p} + \frac{\|G\|^q_q}{q} = \frac{1}{p} + \frac{1}{q} = 1
\]

One application of Holder’s inequality is to prove Minkowski’s inequality, which is simply the triangle inequality for \( L^p \).

We first need the following lemma:

5.3.7 Lemma: If \( a, b > 0 \) and \( 1 \leq p < \infty \), then

\[(a+b)^p \leq 2^{p-1} a^p + 2^{p-1} b^p.\]

Proof: To prove this, we may without loss of generality assume \( a \leq b \). the case \( a = 0 \) is obvious, so we assume \( a > 0 \). Dividing both sides by \( a \) and letting \( x = b/a \), the inequality we want is equivalent to

\[(1+x)^p \leq 2^{p-1} + 2^{p-1} x^p, \quad x \geq 1.\] (5.1)

Clearly this inequality is valid for \( x = 1 \). So to prove (5.1) it suffices to show that the derivative of

\[(1+x)^{p-1} \leq 2^{p-1} x^p, \quad x \geq 1.\]

5.3.8 Proposition: (Minkowski’s inequality) If \( 1 \leq p \leq \infty \) then

\[\|f + g\|_p \leq \|f\|_p + \|g\|_p.\]

Proof: Since \(|(f + g)(x)| \leq |f(x)| + |g(x)|\), integrating gives the case when \( p = 1 \). The case \( p = \infty \) is also easy. So let us suppose \( 1 < p < \infty \). If \( \|f\|_p \) or \( \|g\|_p \) is infinite, the result is obvious, so we may assume both are finite.

The inequality Lemma 5.3.7 with \( a = |f(x)| \) and \( b = |g(x)| \) yields, after an integration,

\[\int |f+g(x)|^p d\mu \leq 2^p \int |f(x)|^p d\mu + 2^p \int |g(x)|^p d\mu\]

So we have \( \|f+g\|_p < \infty \). Clearly we may assume \( \|f+g\|_p > 0 \).

Now write

\[|f+g|^p \leq |f| |f+g|^{p-1} + |g| |f+g|^{p-1}\]

and apply Holder’s inequality with \( q = \left(1 - \frac{1}{p}\right)^{-1} \). We obtain

\[\int |f+g|^p \leq \|f\|_p \left(\int |f+g|^{(p-1)q} d\mu\right)^{\frac{1}{q}} + \|g\|_p \left(\int |f+g|^{(p-1)q} d\mu\right)^{\frac{1}{q}}.\]
Since \( p^{-1} + q^{-1} = 1 \), then \((p-1) q = p\), so we have
\[
\|| f + g \||_p^p \leq (|| f ||_p + || g ||_p)(|| f + g ||_p^{p/q})
\]
Dividing both side by \( || f + g ||_p^{p/q} \) and using the fact that \( p - (p/q) = 1 \) gives us our result.

Minkowski’s inequality says that \( L^p \) is a normed linear space, provided we identify functions that are equal a.e.

**Completeness of \( L^p \)**
The notation of convergence for a sequence of real numbers generalizes to give us a notion of convergence for sequence in a normed linear space.
We say \( f_n \) converges to \( f \) in \( L^p \) if \( || f_n - f ||_p \to 0 \) as \( n \to \infty \). The next proposition compares convergence in \( L^p \) to convergence in measure. Before we prove this, we prove an easy preliminary result known as Chebyshev’s inequality.

**5.3.9 Lemma:** If \( 1 \leq p < \infty \),
\[
\mu (\{ x : |f(x)| \geq a \}) \leq \frac{|| f ||_p^p}{a^p}
\]
**Proof:** If \( A = \{ x : |f(x)| \geq a \} \), then
\[
\mu (A) \leq \frac{\int_A |f(x)|^p d\mu}{a^p} \leq \frac{1}{a^p} \int |f|^p d\mu
\]

**5.3.10 Proposition:** If \( f_n \) converges to \( f \) in \( L^p \), then it converges in measure.
**Proof** If \( \varepsilon > 0 \), by chebyshev’s inequality
\[
\mu (\{ x : |f_n(x) - f(x)| > \varepsilon \}) = \mu (\{ x : |f_n(x) - f(x)|^p > \varepsilon^p \}) \leq \frac{|| f_n - f ||_p^p}{\varepsilon^p} \to 0
\]
Letting \( f_n = n^2 X(0, 1/n) \) on \([0, 1]\) with the measure being Lebesgue measure gives an example where \( f_n \) converges to 0 a.e. and in measure, but does not converge in \( L^p \).

We next show that \( L^p \) is complete. This is often phrased as saying that \( L^p \) is a Banach space, i.e. a complete normed linear space.

**5.3.11 Proposition:** If \( 1 \leq p \leq \infty \), then \( L^p \) is complete.
**Proof.** We do only the case \( p < \infty \), the case \( p = \infty \) is easy. Suppose \( f_n \) is a Cauchy sequence in \( L^p \). Given \( \varepsilon = 2^{-(j+1)} \), there exists \( n_j \) such that if \( n, m \geq n_j \), then \( || f_n - f_m ||_p \leq 2^{-(j+1)} \). Without loss of generality we may assume \( n_j \geq n_{j-1} \) for each \( j \).

Set \( n_0 = 0 \) and define \( f_0 = 0 \). If \( A_j = \{ x : |f_{nj} - f_{nj+1}(x)| > 2^{-j/2} \} \),
Then from Lemma 13.4, \( \mu (A_j) \leq 2^{-jp/2} \). We have
\[
\mu \left( \bigcap_{j=1}^\infty \bigcup_{m=j}^\infty A_m \right) = \lim_{j \to \infty} \mu \left( \bigcup_{m=j}^\infty A_m \right) \leq \lim_{j \to \infty} \mu \left( \bigcup_{m=j}^\infty A_m \right) = 0
\]
So except for a set of measure 0, for each x there is a last \( j \) for which \( x \in U_{m=j}^\infty A_m \), hence a last \( j \) for which \( x \in A_j \). So for each \( x \) (except for the null set) there is an \( j_0 \) (depending on \( x \)) such that if \( j \geq j_0 \), then \( |f_{nj}(x) - f_{nj-1}(x)| \leq 2^{-j} \).

Set

\[
g_j(x) = \sum_{m=1}^j |f_{nm}(x) - f_{nm-1}(x)|
\]

\( g_j(x) \) increases for each \( x \), and the limit is finite for almost every \( x \) by the preceding paragraph. Let us call the limit \( g(x) \). We have

\[
\|g\|_p \leq 2 + \|f_{n1}\|_p < \infty.
\]

By Minkowski’s inequality, and so by Factou’s lemma, \( \|g\|_p \leq 2 + \|f_{n1}\|_p \). This follows by dominated convergence with the function \( g \) defined above as the dominating function.

We have thus shown that \( \|f - f_{nj}\|_p \to 0 \). Given \( \varepsilon = 2^{-(j+1)} \), if \( m \geq n_j \), then \( \|f - f_m\|_p \leq \|f - f_{nj}\|_p + \|f_{m} - f_{nj}\|_p \). This shows that \( f_m \) converges to \( f \) in \( L^p \) norm.

Next we show:

5.3.12 **Proposition:** The set of continuous functions with compact support is dense in \( L^p(\mathbb{R}^d) \).

**Proof.** Suppose \( f \in L^p \). By dominated convergence \( \int f - f_{[n,n]} \to 0 \) as \( n \to \infty \), the dominating function being \( |f|^p \). So we may suppose \( f \) has compact support. By writing \( f = f^+ - f^- \) we may suppose \( f \geq 0 \). By taking simple functions \( s_n \) increasing to \( f \), we have \( \int f - s_n \to 0 \) by dominated convergence, so it suffices to consider simple functions. By linearity, it suffices to consider characteristic functions with compact support. Given such a \( X \) and \( \varepsilon > 0 \) we showed that there exists \( g \) continuous with compact support and with values in \([0, 1]\) such that \( \int |g - X| < \varepsilon \). Since \( |g - X| \leq 1 \), then \( \int |g - X| \leq \int |g - X| < \varepsilon \).

The following is very useful.

5.3.13 **Proposition:** For \( 1 < p < \infty \) and \( p^{-1} + q^{-1} = 1 \),
\[ \|f\|_p = \text{Sup} \{ \int f g : \|g\|_q \leq 1 \} \quad (5.2) \]

When \( p = 1 \) (5.2) holds if we take \( q = \infty \), and if \( p = \infty \) (5.2) holds if we take \( q = 1 \).

**Proof:** The right hand side of (5.2) is less than the left hand side by Holder’s inequality. So we need only show that the right hand side is greater than the left hand side.

First suppose \( p = 1 \). Take \( g(x) = \text{sgn } f(x) \) where \( \text{sgn } a \) is 1 if \( a > 0 \), is 0 if \( a = 0 \), and is \(-1\) if \( a < 0 \). Then \( g \) is bounded by 1 and \( f g = |f| \). This takes care of the case \( p = 1 \).

Next suppose \( p = \infty \). Since \( \mu \) is \( \sigma \) finite, there exist sets \( F_n \) increasing up to \( X \) such that \( \mu(F_n) < \infty \) for each \( n \). If \( M = \|f\|_\infty \), let \( a \) be any finite real less than \( M \). By the definition of \( L_\infty \) norm, the measure of \( \{x \in F_n : |f(x)| > a\} \) must be positive if \( n \) is sufficiently large. Let \( g(x) = (\text{sgn } f(x)) X \alpha(x) / \mu(A) \). Then the \( L_1 \) norm of \( g \) is 1 and \( \int f g = \int |f| / \mu(A) \geq a \). Since \( a \) is arbitrary, the supremum on the right hand side must be \( M \).

Now suppose \( 1 < p < \infty \). We may suppose \( \|f\|_p > 0 \). Let \( q_n \) be a sequence of nonnegative simple functions increasing to \( f^+ \), \( r_n \) a sequence of nonnegative simple functions increasing to \( f^- \), and \( s_n(x) = q_n(x) - r_n(x) \) for \( x \neq X \). Then \( s_n(x) \rightarrow f(x) \) for each \( x \), \( |s_n(x)| \leq |f(x)| \) for each \( x \), \( s_n \) is a simple function and \( \|s_n\|_p < \infty \) for each \( n \). If \( f \in L^p \), then \( \|s_n\|_p \rightarrow \|f\|_p \) by dominated convergence. If \( \int |f|^p = \infty \), then \( \int |s_n|^p \rightarrow \infty \) by monotone convergence. For \( n \) sufficiently large, \( \|s_n\|_p > 0 \).

Let
\[
G_n(x) = (\text{sgn } f(x)) \frac{|s_n(x)|^{p-1}}{\|s_n\|_p^{p/q}}
\]

Since \( (p-1)q = p \), then
\[
\|g_n\|_q = \left( \int |s_n|^{(p-1)q/p} \right)^{1/q} = \|s_n\|_p^{p/q}
\]

On the other hand, since \( |f| \geq |s_n| \),
\[
\int f g_n = \int f |s_n|^{p-1} \geq \int |s_n|^{p} \geq \|s_n\|_p^{p-1} \frac{\|s_n\|_p^{p/q}}{\|s_n\|_p^{p/q}} = \|s_n\|_p^{p-1} \frac{\|s_n\|_p^{p/q}}{\|s_n\|_p^{p/q}}
\]

Since \( p - (p/q) = 1 \), then \( \int f g_n \geq \|s_n\|_p \), which tends to \( \|f\|_p \).

The above proof also establishes

**5.3.14 Corollary:** For \( 1 < p < \infty \) and \( p^{-1} + q^{-1} = 1 \),
\[ \|f\|_p = \sup \{ \int f g : \|g\|_q \leq 1, \ g \text{ simple} \} \]

5.3.15 Theorem: If \( 1 < p < \infty \) and \( p^{-1} + q^{-1} = 1 \), then \((L^p)^* = L^q\)

What this means is that if \( H \) is a bounded linear functional on \( L^p \), then there exists \( g \in L^q \), then \( H(f) = \int f g \) is a bounded linear functional on \( L^p \).

Proof. If \( g \in L^q \), then setting \( H(f) = \int f g \) for \( f \in L^p \) yields a bounded linear functional; the boundedness follows from Holder’s inequality. Moreover, from Holder’s inequality we see that \( \|H\| = \|g\|_q \).

Now suppose we are given a bounded linear functional \( H \) on \( L^p \) and we must show there exists \( g \in L^q \) such that \( H(f) = \int f g \). First suppose \( \mu(X) < \infty \). Define \( \nu(A) = H(X_A) \). If \( A \) and \( B \) are disjoint, then

\[ v(A \cup B) = H(X_{A \cup B}) = H(X_A + X_B) = H(X_A) + H(X_B) = v(A) + v(B) \]

To show \( v \) is countably additive, it suffices to show that if \( A_n \uparrow A \), then \( v(A_n) \to v(A) \). But if \( A_n \uparrow A \), then \( X_{A_n} \to X_A \) in \( L^p \), and so \( v(A_n) = H(X_{A_n}) \to H(X_A) = v(A) \); we use here the fact that \( \mu(X) < \infty \). Therefore \( v \) is a countably additive signed measure. Moreover, if \( \mu(A) = 0 \), then \( X_A = 0 \) a.e., hence \( v(A) = H(X_A) = 0 \) By writing \( v = v^+ - v^- \) and using the Radon-Nikodym theorem for both the positive and negative parts, we see there exists an integrable \( g \) such that \( v(A) = \int_A g \) for all sets \( A \). If \( s = \sum a_i X_{A_i} \) is a simple function, by linearity we have

\[ H(s) = \sum a_i H(X_{A_i}) = \sum a_i v(A_i) = \sum a_i \int g X_{A_i} = \int g s \]

By Corollary 5.3.14,

\[ \|g\|_q = \sup \{ \int g s : \|s\|_p \leq 1, \ s \text{ simple} \} \leq \sup \{ H(s) : \|s\|_p \leq 1 \} \leq \|H\| \]

If \( s_n \) are simple functions tending to \( f \) in \( L^p \), then \( H(s_n) \to H(f) \), while by Holder’s inequality \( \int s_n g \to \int f g \). We thus have \( H(f) = \int f g \) for all \( f \in L^p \), and \( \|g\|_p \leq \|H\| \). By Holder’s inequality, \( \|H\| \leq \|g\|_p \).

In the case where \( \mu \) is \( \sigma \) finite, but not finite, let \( F_n \uparrow X \) be such that \( \mu(F_n) < \infty \) for each \( n \). Define functionals \( H_n \) by \( H_n(f) = H(f X_{F_n}) \). Clearly each \( H_n \) is a bounded linear functional of \( L^p \). Applying the above argument, we see there exist \( g_n \) such that \( H_n(f) = \int g_n f \) and \( \|g_n\|_q = \|H_n\| \leq \|H\| \). It is easy to see that \( g_n \) is 0 if \( x \notin F_n \). Moreover, by the uniqueness part of the Radon-Nikodym theorem, if \( n > m \), then \( g_n = g_m \) on \( F_m \). Define \( g \) by setting \( g(x) = g_n(x) \) if \( x \in F_n \). Then \( g \) is well defined. By Fatou’s lemma, \( g \) is in \( L^q \) with a norm bounded by \( \|H\| \). Since \( f \in L^p \) by dominated convergence, then \( H_n(f) = H(f X_{F_n}) \to H(f) \), since \( H \) is a bounded linear
functional on $L^p$. On the other hand $H_n (f) = \int f \, g_n = \int f \, g \to \int f \, g$ by dominated convergence. So $H (f) = \int fg$. Again by Holder’s inequality $\|H\| \leq \|g\|_p$.

### Applications
Fourier transforms give a representation of a function in terms of frequencies. We give the basic properties here.

If $f \in L^1(\mathbb{R}^n)$, define the Fourier transform $\hat{f}$ by

$$\hat{f}(u) = \int_{\mathbb{R}^n} e^{iu \cdot x} f(x) \, dx \quad u \in \mathbb{R}^n \quad (5.3)$$

We are using $\cdot$, $\cdot$ for the standard inner product in $\mathbb{R}^n$. Various books have slightly different definitions. Some put a negative sign before the $iu \cdot x$, some have a $2\pi$ either in front of the integral or in the exponent. The basic theory is the same in any case.

Some basic properties of the Fourier transform are given by

#### 5.3.17 Proposition:

Suppose $f$ and $g$ are in $L^1$. Then

(a) $\hat{f}$ is bounded and continuous;

(b) $(\hat{f} + \hat{g})(u) = \hat{f}(u) + \hat{g}(u)$; $(af)(u) = a \hat{f}(u)$;

(c) if $f_a(x) = f(x+a)$, then $\hat{f}_a(u) = e^{-iu \cdot a} \hat{f}(u)$;

(d) if $g_a(x) = e^{iu \cdot x} g(x)$, then $\hat{g}_a(u) = \hat{f}(u+a)$;

(e) if $h_a(x) = f(ax)$, then $\hat{h}_a(u) = a^{-n} \hat{f}(u/a)$.

**Proof** (a) $\hat{f}$ is bounded because $f \in L^1$ and $|e^{iu \cdot x}| = 1$. We have

$$\hat{f}(u+h) - \hat{f}(u) = \int (e^{i(u+h) \cdot x} - e^{iu \cdot x}) f(x) \, dx$$

So

$$|\hat{f}(u+h) - \hat{f}(u)| \leq \int |e^{iu \cdot x}||e^{ih \cdot x} - 1||f(x)| \, dx$$

The integrand is bounded by $2 |f(x)|$, which is integrable, and $e^{ih \cdot x} - 1 \to 0$ as $h \to 0$, and thus the continuity follows by dominated convergence.
(b) is obvious (c) follows because

\[ \hat{f}_a(x) = \int e^{iux} f(x) dx = e^{iua} \hat{f}(u/a) \]

By a change of variables For (d),

\[ \hat{g}_a(u) = \int e^{iux} e^{iax} f(x) dx = \int e^{i(u+a)x} f(x) dx = \hat{f}(u+a) \]

Finally for (e), by a change of variables,

\[ \hat{h}_a(u) = \int e^{iux} f(ax) dx = a^{-n} \int e^{iux} f(y) dy \\
= a^{-n} \int e^{i(u/a)x} f(y) dy = a^{-n} \hat{f}(u/a) \]

One reason for the usefulness of Fourier transforms is that they relate derivatives and multiplication.

5.3.18 Proposition: Suppose \( f \in L^1 \) and \( x_j f(x) \in L^1 \), where \( x_j \) is the \( j^{th} \) coordinate of \( x \). Then

\[ \frac{\partial \hat{f}}{\partial u_j}(u) = i \int e^{iu.x} x_j f(x) dx. \]

Proof. Let \( e_j \) be the unit vector in the \( j^{th} \) direction. Then

\[ \frac{\hat{f}(u + he_j) - \hat{f}(u)}{h} = \frac{1}{h} \int (e^{i(u+he_j).x} - e^{iu.x}) f(x) dx \]

\[ = \int e^{iu.x} \left( \frac{e^{ihx_j}}{h} - 1 \right) f(x) dx. \]

Since

\[ \left| \frac{1}{h} \left( e^{ihx_j} - 1 \right) \right| \leq |x_j| \]

And \( x_j f(x) \in L^1 \), the right hand side converges to \( \int e^{iu.x} x_j f(x) dx \) by dominated convergence. Therefore the left hand side converges. Of course the limit is \( \frac{\partial \hat{f}}{\partial u_j} \).

The convolution of \( f \) and \( g \) is defined by
\[ f * g(x) = \int f(x - y)g(y)\,dy. \]

By a change of variables, this is the same as \( \int f(y)g(x - y)\,dy \), so \( f*g = g*f \).

**5.3.19 Proposition:** (a) If \( f, g \in L^1 \), then \( f*g \) is in \( L^1 \) and \( \|f*g\|_1 \leq \|f\|_1 \|g\|_1 \)

(b) The Fourier transform of \( f*g \) is \( \hat{f}(u) \hat{g}(u) \).

**Proof.** (a) We will show \( f*g \) is finite a.e by showing \( \int |f*g(x)|\,dx < \infty \).

We have \( \int |f*g(x)|\,dx \leq \int \int |f(x-y)||g(y)|\,dy\,dx \).

Since the integrand is nonnegative, we can apply Fubini and the right hand side is equal to

\[
\int \int |f(x-y)|\,dx\,|g(y)|\,dy = \int \int |f(x)|\,dx\,|g(y)|\,dy = \|f\|_1 \|g\|_1
\]

The first equality here follows by a change of variables. To verify that we can do a change of variables, we reduce to simple functions and then characteristic functions, and then use the translation invariance of Lebesgue measure.

(b) We have

\[
\int e^{iu.x} \int f(x-y)g(y)\,dy\,dx = \int e^{iu.(x-y)}f(x-y)\,dx\,e^{iu.y}g(y)\,dy = \int \hat{f}(u)\hat{g}(u)\,dy = \hat{f}(u)\hat{g}(u)
\]

We applied Fubini in the first equality; this is valid because as we saw in (a), the absolute value of the integrand is integrable.

We want to give a formula for recovering \( f \) from \( \hat{f} \). First we need to calculate the Fourier transform of a particular function.

**5.3.20 Proposition:** (a) Suppose \( f_1 : \mathbb{R} \to \mathbb{R} \) is defined by
\[ f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \]

Then \( \hat{f}_1(u) = e^{-u^2/2} \)

(b) Suppose \( f_n : \mathbb{R}^n \to \mathbb{R} \) is given by

\[ f_n(x) = \frac{1}{(2\pi)^{n/2}} e^{-|x|^{2/2}} \]

Then \( \hat{f}_n(u) = e^{-|u|^{2/2}} \)

**Proof.** (a) may be proved using contour integration, but let’s give a real variable proof. Let \( g(u) = \int e^{iur} e^{-x^2/2} dx \). Differentiate with respect to \( u \). We may differentiate under the integral sign because \( (e^{i(u+h)r} - e^{iur}) \) \( / \) \( h \) is bounded in absolute value by \( |x| \) and \( |x| e^{-x^2/2} \) is bounded in absolute value by \( |x| \) and \( |x| e^{-x^2/2} \) is integrable; therefore dominated convergence applies. We then obtain

\[ g(u) = i \int e^{iur} xe^{-x^2/2} dx. \]

By integration by parts this is equal to

\[ -u \int e^{iur} e^{-x^2/2} dx = -ug(u). \]

Solving the differential equation \( g(u) = -ug(u) \), we have

\[ [ \log g(u) ] = \frac{g'(u)}{g(u)} = -u, \]

So \( \log g(u) = -u^2/2 + c_1 \) and so then

\[ g(u) = c_2 e^{-u^2}. \]

Since \( g(0) = \int e^{-x^2/2} dx = \sqrt{2\pi}, c_2 = \sqrt{2\pi} \). Substituting this value of \( c_2 \) in (7.2) and dividing both sides by \( \sqrt{2\pi} \) proves (a).

For (b), since \( f_n(x) = f_1(x_1) \ldots f_1(x_n) \) if \( x = (x_1 \ldots, x_n) \),

\[ \hat{f}_n(u) = \int \ldots \int e^{i\sum_{j=1}^{n} u_j x_j} f_1(x_1)\ldots f_1(x_n) dx_1 \ldots dx_n \]

\[ = \hat{f}_1(u_1)\ldots \hat{f}_1(u_n) = e^{-|u|^{2/2}} \]

One more preliminary before proving the inversion theorem.
5.3.21 Proposition: Suppose \( \varphi \) is in \( L^1 \) and \( \int \varphi (x) \, dx = 1 \). Let \( \varphi_A (x) = A^{-n} \varphi (x/A) \).

(a) Then \( \|f^* \varphi_A - f\|_1 \to 0 \) as \( A \to 0 \)

(b) If \( f \) is continuous with compact support, then \( f^* \varphi_A \) converges to \( f \) point wise.

Proof. (a) Let \( \varepsilon > 0 \). Choose \( g \) continuous with compact support so that \( \|f - g\|_1 < \varepsilon \).

Let \( h = f - g \). A change of variables shows that \( \|\varphi_A\|_1 = \|\varphi\|_1 \).

Observe
\[
\|f^* \varphi_A - f\|_1 \leq \|g^* \varphi_A - h\|_1
\]

And
\[
\|h^* \varphi_A - h\|_1 \leq \|h\|_1 + \|h^* \varphi_A\|_1 \leq \|h\|_1 \|\varphi_A\|_1 < \varepsilon (1 + \|\varphi\|_1).
\]

So since \( \varepsilon \) is arbitrary, it suffices to show that \( g^* \varphi_A \to g \) in \( L^1 \).

We start by writing
\[
g^* \varphi_A(x) - g(x) = \int g(x - y) \varphi_A(y) \, dy - g(x) = \int g(x - Ay) \varphi(y) \, dy - g(x)
\]
\[
= \int [g(x - Ay) - g(x)] \varphi(y) \, dy.
\]

We used a change of variables and the fact that \( \int \varphi(y) \, dy = 1 \). Because \( g \) is continuous with compact support, then \( g \) is bounded, and the integral on the right goes to 0 by dominated convergence, the dominating function being \( \|g\|_{\infty} |\varphi(y)| \). Therefore \( g^* \varphi_A(x) \) converges to \( g(x) \) point wise.

To show the convergence in \( L^1 \), we have
\[
\int |g^* \varphi_A(x) - g(x)| \, dx \leq \int \int |g(x - Ay) - g(x)| \varphi(y) \, dy \, dx
\]
\[
= \int \int |g(x - Ay) - g(x)| \varphi(y) \, dx \, dy.
\]

Since \( g \) is continuous with compact support and hence bounded, for each \( y \)
\[ G_A(y) = \int |g(x - Ay) - g(x)| \, dx \]

Converges to 0 as \( A \to 0 \) by dominated convergence. Also
\[ G_A(y) \leq \int |g(x - Ay)| \, dx + \int |g(x)| \, dx \leq 2\|g\|_1 < \infty. \]

Then
\[ \int G_A(y) |\varphi(y)| \, dy \]

Converges to 0 as \( A \to 0 \) by dominated convergence, the dominating function being \( 2\|g\|_1 |\varphi(y)| \).

(c) This follows from the argument we used for \( g \) above.

Now we are ready to give the inversion formula. The proof seems longer than it might be, but there is no avoiding the introduction of the function \( H_a \) or some similar function.

5.3.22 Theorem: Suppose \( f, \hat{f} \in L^1 \). Then
\[ f(y) = \frac{1}{(2\pi)^n} \int e^{-iux} \hat{f}(u) \, du, \quad \text{a.e.} \]

Proof. If \( g(x) = a^{-n}f(x/a) \), then its Fourier transform is \( \hat{f}(au) \). So the Fourier transform of
\[ \frac{1}{a^n} \frac{1}{(2\pi)^{n/2}} e^{-\frac{\|u\|^2}{2a^2}} \]

Is \( e^{-a^2u^2/2} \). Therefore if we let
\[ H_a(x) = \frac{1}{(2\pi)^n} e^{-\|x\|^2/2a^2} \]

We have
\[ \hat{H}_a(u) = (2\pi)^{-n/2} a^{-n} e^{\|u\|^2/2a^2}. \]

We have
\[ \int \hat{f}(u) e^{-iu\cdot x} H_a(u) \, du = \int \int e^{iu\cdot x} f(x) e^{-iu\cdot x} H_a(u) \, dx \, du \]
\[ = \int \int e^{iux} H_a(u) \, du \, f(x) \, dx \]
\[ = \int \hat{H}_a(x - y) f(x) \, dx. \quad (5.4) \]
We can interchange the order of integration because
\[ \int \int |f(x)||H_a(u)|dxdu < \infty \]
The left hand side of the first line of (7.3) converges to \((2\pi)^{-n}\int \hat{f}(u)e^{-iu.y}dy\)
As a \(\to \infty\) by dominated convergence and the fact that \(\hat{f} \in L^1\).
The last line of (5.4) is equal to.
\[ \int \hat{H}_a(y-x)f(x)dx = f \ast \hat{H}_a(y) \quad (5.5) \]
Since \(\hat{H}_a\) is symmetric. But by Proposition 14.5, \(f \ast \hat{H}_a\) converges to \(f\) in \(L^1\) as \(a \to \infty\)

The last topic that we consider is the Plancherel theorem.

5.3.23 Theorem: (a) Suppose \(f\) is continuous with compact support. Then \(\hat{f} \in L^2\) and
\[ ||f||_2 = (2\pi)^{-n/2} ||\hat{f}||_2. \quad (5.6) \]

(b) We can use the result in (a) to define \(\hat{f}\) when \(f \in L^2\) and so that (5.6) holds.

Proof: (a) Let \(g(x) = \overline{f(-x)}\). Note
\[ \hat{g}(u) = \int e^{iu.x}f(-x)dx = \int e^{-iu.x}f(x)dx = \int e^{iu.x}f(x)dx = \hat{f}(u) \]
By (5.4) and (5.5) with \(y = 0\)
\[ F \ast g \ast \hat{H}_a(0) = \int \hat{f} \ast \overline{g}(u)H_a(u)du. \quad (5.7) \]
Since \(\hat{f} \ast \overline{g}(u) = \hat{f}(u)\hat{g}(u) = |\hat{f}(u)|^2\), the right hand side of (5.7) converges by monotone convergence to \((2\pi)^{-n}\int |\hat{f}(u)|^2 du\) as \(a \to \infty\). Since \(f\) and \(g\) are continuous with compact support, then it is easy to see that \(f \ast g\) is also, and so the left hand side of (5.7) converges to \(f \ast g(0) = \int f(y)g(-y)dy = \int |f(y)|^2 dy\) by proposition 5.3.23(b).
(b) The set of continuous functions with compact support is dense in $L^2$. Given a function $f$ in $L^2$, choose a sequence of continuous functions with compact support $\{f_m\}$ such that $f_m \to f$ in $L^2$. By the result in (a), $\{\hat{f}_m\}$ is a Cauchy sequence in $L^2$, and therefore converges to a function in $L^2$, which we call $\hat{f}$. If $\{f_m'\}$ is another sequence of continuous functions with compact support converging to $f$ in $L^2$, then $\{f_m' - f_m\}$ is a sequence of continuous functions with compact support converging to 0 in $L^2$; by the result in (a), $\hat{f}_m' - \hat{f}_m$ converges to 0 in $L^2$, and therefore $\hat{f}$ is defined uniquely up to almost everywhere equivalence. By passing to the limit in $L^2$ on both sides of (5.6), we see that (5.6) holds for $f \in L^2$.

**5.4 Assignment**

1. Prove the completeness of $f_1$.
2. Assume that $<E_i>$ is a sequence of disjoint measurable sets and $E = \bigcup E_i$. Then for any set $A$ prove that $\mu^*(A \cap E) = \sum \mu^*(A \cap E_i)$.
3. If, $(X, \beta, \mu)$ is a measure space, then find a complete measure space $(X, \beta_0, \mu_0)$ such that,
   (i) $\beta \subseteq \beta_0$
   (ii) $E \in \beta \Rightarrow \mu E = \mu_0 E$
   (iii) $E \in \beta_0 \Leftrightarrow E = A \cup B$ where $B \in \beta$ and $A \subseteq C$, $C \in \beta$, $\mu C = 0$
5.5 Check your progress
1. Prove that If $A \in \mathcal{B}$, $B \in \mathcal{B}$, and $A \subset B$, then $\mu A \leq \mu B$
2. If $E_i \in \mathcal{B}$, $\mu E_1 < \infty$ and $E_i \supseteq E_{i+1}$, then prove that, $\mu \left( \bigcap_{i=1}^{\infty} E_i \right) = \lim_{n \to \infty} \mu E_n$
3. If $E_i \in \mathcal{B}$ then prove that, $\mu \left( \bigcup_{i=1}^{\infty} E_i \right) \leq \sum_{i=1}^{\infty} \mu E_i$
4. Let $\{ (X_\alpha, \mathcal{B}_\alpha, \mu_\alpha) \}$ be a collection of measure spaces, and suppose that the sets $\{ X_\alpha \}$ are disjoint. Then we can form a new measure space (called their union) by letting $X = \bigcup X_\alpha$, $\mathcal{B} = \{ B: (\alpha)[B \cap X_\alpha \in \mathcal{B}_\alpha] \}$, and defining $\mu (B) = \sum \mu_\alpha (B \cap X_\alpha)$
   a. Show that $\mathcal{B}$ is a $\sigma$-algebra.
   b. Show that $\mu$ is a measure.
   c. Show that $\mu$ is $\sigma$-finite if and only if all but a countable number of the $\mu_\alpha$ are zero and the remainder is $\sigma$-finite.

5.6 POINTS FOR DISCUSSION / CLARIFICATION

5.7 Suggested Study material
5- Lecture Notes on Real analysis Richard F. Bass.