Mathematics

Personal Study Notes and Formal Derivations

Victor Alves

2026-04

Contents

Disclaimer

- **Incomplete Document:** This file constitutes supporting material under active development. Several sections and asymptotic derivations are still being revised, expanded, and supplemented.
- **AI-Assisted Construction:** This material was structured, reviewed, and expanded with the assistance of Artificial Intelligence models for didactic organization and Markdown/LaTeX formatting rigor.
- **Margin of Error:** Due to the technical nature of the matrix and asymptotic proofs, the text may contain typographical errors, algebraic omissions, or theoretical inaccuracies not yet reviewed by the author. It should not be used as the sole definitive bibliographic source.

Chapter 1: Sets and Functions

1.1 Basic Set Theory

1.1.1 Motivation

The fundamental problem that set theory solves is the **formalization of mathematical objects as collections**. Without a precise definition of sets, operations, and functions, much of modern mathematics—from calculus to probability theory—would lack a rigorous foundation. Sets allow us to group objects, define relationships, and build the structures upon which all other mathematical concepts rest.

Geometric Interpretation in $\mathbb{R}^2$: Visualize a set as a collection of points in the plane. The union of two sets corresponds to the total area covered by both regions, while the intersection corresponds to the overlapping area. The complement of a set corresponds to the area outside the region. This geometric intuition helps understand logical operations through spatial relationships.

1.1.2 Formal Definitions

Definition 1 (Power Set): The **power set** of a set X, denoted $\mathcal{P}(X)$ or 2^X, is the set of all subsets of X:

$$\mathcal{P}(X) = \{A : A \subseteq X\} \quad (1)$$

Example: If $X = \{a, b\}$, then $\mathcal{P}(X) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$.

Definition 2 (Cartesian Product): The **Cartesian product** of sets A and B is the set of all ordered pairs:

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\} \quad (2)$$

If $A = B$, the set $A \times A = A^2$ is the Cartesian square of A. The **diagonal** of A^2 is the subset $\Delta_A = \{(a, a) : a \in A\}$.

1.1.3 Key Theorems

Theorem 1 (Properties of Set Operations): For any sets $A, B, C \subseteq X$, with complements taken relative to X, the following hold:

Property	Formula
Commutativity	$A \cup B = B \cup A$, $A \cap B = B \cap A$
Associativity	$A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$
Distributivity	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
De Morgan's Laws	$(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$
Double Complement	$(A^c)^c = A$
Subset Equivalence	$A \subset B \iff B^c \subset A^c$

Theorem 2 (De Morgan's Laws): For any family of sets $\{A_\lambda\}_{\lambda \in L}$:

$$\boxed{\left(\bigcup_{\lambda \in L} A_\lambda\right)^c = \bigcap_{\lambda \in L} A_\lambda^c}$$

$$\boxed{\left(\bigcap_{\lambda \in L} A_\lambda\right)^c = \bigcup_{\lambda \in L} A_\lambda^c}$$

Proof:

1. For the first identity: $x \in \left(\bigcup_{\lambda \in L} A_\lambda\right)^c \iff x \notin \bigcup_{\lambda \in L} A_\lambda$.
2. By definition of union: $x \notin \bigcup_{\lambda \in L} A_\lambda \iff x \notin A_\lambda$ for all $\lambda \in L$.
3. By definition of complement: $x \notin A_\lambda \iff x \in A_\lambda^c$ for all $\lambda \in L$.
4. Therefore: $x \in A_\lambda^c$ for all $\lambda \in L \iff x \in \bigcap_{\lambda \in L} A_\lambda^c$.
5. Thus: $\left(\bigcup_{\lambda \in L} A_\lambda\right)^c = \bigcap_{\lambda \in L} A_\lambda^c$.
6. The second identity follows by applying the first to the complements.

$\square$

1.2 Functions and Their Properties

1.2.1 Motivation

The problem that functions solve is the **mathematical description of relationships between objects**. Functions allow us to map elements from one set to another, providing a rigorous way to model dependencies, transformations, and correspondences. In economics, functions model production, utility, and demand; in analysis, they are the objects of differentiation and integration.

Geometric Interpretation in $\mathbb{R}^2$: Visualize a function $f : A \to B$ as a rule that assigns to each point in the domain A a point in the codomain B. The graph of the function is a curve in $\mathbb{R}^2$. Injectivity means the curve never repeats a height; surjectivity means the curve reaches every point in the codomain.

1.2.2 Formal Definitions

Definition 3 (Function): A **function** $f : A \to B$ is composed of: 1. A set A, called the **domain**. 2. A set B, called the **codomain**. 3. A rule that assigns to every $x \in A$ a unique element $f(x) \in B$.

The notation $x \mapsto f(x)$ means f associates with each $x \in A$ the value $f(x)$. The **graph** of f is:

$$\text{graph}(f) = \{(x, y) \in A \times B : y = f(x)\} \quad (3)$$

Definition 4 (Injective Function): A function $f : A \to B$ is **injective** (one-to-one) if:

$$\forall x, y \in A, \quad x \neq y \implies f(x) \neq f(y) \quad (4)$$

Equivalently: $f(x) = f(y) \implies x = y$.

Definition 5 (Surjective Function): A function $f : A \to B$ is **surjective** (onto) if:

$$\forall y \in B, \exists x \in A \text{ such that } f(x) = y \quad (5)$$

Definition 6 (Bijective Function): A function is **bijective** if it is both injective and surjective.

1.2.3 Key Theorems

Theorem 3 (Properties of Image): For a function $f : A \to B$ and sets $X, Y \subseteq A$:

$$\boxed{f(X \cup Y) = f(X) \cup f(Y)}$$

$$\boxed{f(X \cap Y) \subseteq f(X) \cap f(Y)}$$

If f is injective, then $f(X \cap Y) = f(X) \cap f(Y)$.

Proof:

1. For the union:
   - $z \in f(X \cup Y) \iff \exists x \in X \cup Y$ such that $f(x) = z$.
   - This means $x \in X$ or $x \in Y$, so $z \in f(X)$ or $z \in f(Y)$.
   - Thus $z \in f(X) \cup f(Y)$.
2. For the intersection:
   - $z \in f(X \cap Y) \implies \exists x \in X \cap Y$ with $f(x) = z$.
   - Then $x \in X$ and $x \in Y$, so $z \in f(X)$ and $z \in f(Y)$.
   - Thus $z \in f(X) \cap f(Y)$.
3. The reverse inclusion for intersection requires injectivity. If $z \in f(X) \cap f(Y)$, then $\exists x \in X$ and $\exists y \in Y$ with $f(x) = f(y) = z$. By injectivity, $x = y \in X \cap Y$, so $z \in f(X \cap Y)$.

□

Theorem 4 (Properties of Pre-image): For a function $f : A \to B$ and sets $Y, Z \subseteq B$:

$$\boxed{f^{-1}(Y \cup Z) = f^{-1}(Y) \cup f^{-1}(Z)}$$

$$\boxed{f^{-1}(Y \cap Z) = f^{-1}(Y) \cap f^{-1}(Z)}$$

$$\boxed{f^{-1}(Y^c) = [f^{-1}(Y)]^c}$$

Proof:

1. For the union:
   - $x \in f^{-1}(Y \cup Z) \iff f(x) \in Y \cup Z \iff f(x) \in Y$ or $f(x) \in Z$.
   - This is equivalent to $x \in f^{-1}(Y)$ or $x \in f^{-1}(Z)$.
   - Thus $x \in f^{-1}(Y) \cup f^{-1}(Z)$.
2. For the intersection:
   - $x \in f^{-1}(Y \cap Z) \iff f(x) \in Y \cap Z \iff f(x) \in Y$ and $f(x) \in Z$.
   - This is equivalent to $x \in f^{-1}(Y)$ and $x \in f^{-1}(Z)$.
   - Thus $x \in f^{-1}(Y) \cap f^{-1}(Z)$.
3. For the complement:
   - $x \in f^{-1}(Y^c) \iff f(x) \in Y^c \iff f(x) \notin Y \iff x \notin f^{-1}(Y)$.
   - Thus $x \in [f^{-1}(Y)]^c$.

□

1.2.4 Inverses of Functions

Definition 7 (Left Inverse): A function $g : B \to A$ is a **left inverse** of $f : A \to B$ if:

$$g \circ f = \mathrm{id}_A \quad (6)$$

That is, $g(f(x)) = x$ for all $x \in A$.

Theorem 5: A function $f : A \to B$ has a left inverse iff it is **injective**.

Proof:

1. ($\Rightarrow$) Suppose f has a left inverse g. If $f(x_1) = f(x_2)$, then:

   $$x_1 = g(f(x_1)) = g(f(x_2)) = x_2$$

   Thus f is injective.

2. ($\Leftarrow$) Suppose f is injective. Choose a fixed $a_0 \in A$. Define $g : B \to A$ by:

   $$g(y) = \begin{cases} x, & \text{if } y = f(x) \text{ for some } x \in A \\ a_0, & \text{otherwise} \end{cases}$$

   Then $g(f(x)) = x$ for all $x \in A$, so g is a left inverse.

□

Definition 8 (Right Inverse): A function $g : B \to A$ is a **right inverse** of $f : A \to B$ if:

$$f \circ g = \mathrm{id}_B \quad (7)$$

That is, $f(g(y)) = y$ for all $y \in B$.

Theorem 6: A function $f : A \to B$ has a right inverse iff it is **surjective**.

Proof:

1. ($\Rightarrow$) Suppose f has a right inverse g. For any $y \in B$, $f(g(y)) = y$, so f is surjective.
2. ($\Leftarrow$) Suppose f is surjective. For each $y \in B$, choose $x_y \in A$ such that $f(x_y) = y$. Define $g(y) = x_y$. Then $f(g(y)) = y$ for all $y \in B$, so g is a right inverse.

□

1.3 Families and Projections

1.3.1 Formal Definitions

Definition 9 (Family): A **family** is a function $x : L \to X$. The set L is called the **index set**. We often write $(x_\lambda)_{\lambda \in L}$.

When $L = \mathbb{N}$, the family $(x_n)_{n \in \mathbb{N}}$ is a **sequence**.

Definition 10 (Product of a Family): Given a family of sets $(A_\lambda)_{\lambda \in L}$, the **product** is:

$$\prod_{\lambda \in L} A_\lambda = \{(x_\lambda)_{\lambda \in L} : x_\lambda \in A_\lambda \text{ for all } \lambda \in L\} \quad (8)$$

Definition 11 (Projection): For $X = \prod_{\lambda \in L} A_\lambda$, the **$\lambda$-projection** is:

$$\pi_\lambda : X \to A_\lambda, \quad \pi_\lambda((x_{\lambda'})_{\lambda' \in L}) = x_\lambda \quad (9)$$

1.3.2 Key Theorems

Theorem 7 (Properties of Families): For any family of sets $(A_\lambda)_{\lambda \in L}$ and any function f:

$$f\left(\bigcup_{\lambda \in L} A_\lambda\right) = \bigcup_{\lambda \in L} f(A_\lambda)$$

$$f\left(\bigcap_{\lambda \in L} A_\lambda\right) \subseteq \bigcap_{\lambda \in L} f(A_\lambda)$$

$$f^{-1}\left(\bigcup_{\lambda \in L} A_\lambda\right) = \bigcup_{\lambda \in L} f^{-1}(A_\lambda)$$

$$f^{-1}\left(\bigcap_{\lambda \in L} A_\lambda\right) = \bigcap_{\lambda \in L} f^{-1}(A_\lambda)$$

Proof: (Follows directly from the set operations proofs by extending to arbitrary unions/intersections.)

1.4 Summary of Key Results

Concept	Definition	Key Formula
Power Set	Set of all subsets	$\mathcal{P}(X) = \{A : A \subseteq X\}$
Cartesian Product	Set of ordered pairs	$A \times B = \{(a, b) : a \in A, b \in B\}$
De Morgan's Laws	Complements of unions/intersections	$(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$
Injective Function	One-to-one	$f(x) = f(y) \implies x = y$
Surjective Function	Onto	$\forall y \in B, \exists x \in A : f(x) = y$
Image	Image of a set	$f(X) = \{f(x) : x \in X\}$
Pre-image	Inverse image	$f^{-1}(Y) = \{x \in A : f(x) \in Y\}$

Concept	Definition	Key Formula
Left Inverse	Exists iff injective	$g \circ f = \mathrm{id}_A$
Right Inverse	Exists iff surjective	$f \circ g = \mathrm{id}_B$

Chapter 2: Finite, Countable and Uncountable Sets

2.1 Natural Numbers and Induction

2.1.1 Motivation

The natural numbers $\mathbb{N}$ form the foundation upon which all of mathematics is built. The Peano axioms provide a rigorous definition of $\mathbb{N}$, establishing the principles of induction and recursion. These principles allow us to prove properties for infinitely many objects by checking just two conditions: a base case and an inductive step.

Geometric Interpretation: Think of the natural numbers as points on a ray starting at 1 and moving infinitely to the right. The Induction Principle states that if a property holds at the first point and, whenever it holds at a point, it holds at the next point, then it holds at every point on the ray.

2.1.2 Peano Axioms

Definition 1 (Peano Axioms): The natural numbers $\mathbb{N}$ are defined by a set and a function $s : \mathbb{N} \to \mathbb{N}$ (the **successor function**) satisfying:

1. **(P1)** $s : \mathbb{N} \to \mathbb{N}$ is injective.
2. **(P2)** $\mathbb{N} \setminus s(\mathbb{N})$ contains a single element, denoted by 1.
3. **(P3) Induction Principle:** For all $X \subseteq \mathbb{N}$, if $1 \in X$ and $s(X) \subseteq X$, then $X = \mathbb{N}$.

Definition 2 (Addition): Since this convention starts $\mathbb{N}$ at 1, addition is defined recursively by:

$$m + 1 := s(m), \qquad m + s(n) := s(m + n) \quad (1)$$

Equivalently, $m + n$ is obtained by applying the successor operation n times if one first introduces 0; under the present convention, the recursive definition above avoids that ambiguity.

Definition 3 (Inequality): For $m, n \in \mathbb{N}$, we define:

$$m < n \iff \exists p \in \mathbb{N} \text{ such that } n = m + p \quad (2)$$

2.1.3 Key Theorems

Theorem 1 (Properties of Addition): For all $m, n, p \in \mathbb{N}$:

Property	Formula
Commutativity	$m + n = n + m$
Associativity	$m + (n + p) = (m + n) + p$
Cancellation	$m + n = m + p \iff n = p$
Trichotomy	Either $m < n$, $m > n$, or $m = n$

Theorem 2 (Well-Ordering Principle): Every nonempty subset $A \subseteq \mathbb{N}$ has a minimum element.

Proof:

1. If $1 \in A$, then 1 is the minimum.
2. If $1 \notin A$, let $X = \{n \in \mathbb{N} : I_n \subseteq \mathbb{N} \setminus A\}$, where $I_n = \{1, 2, \ldots, n\}$.
3. Then $1 \in X$. Moreover, $X \neq \mathbb{N}$, since $A \neq \emptyset$.
4. By the Induction Principle, $s(X) \not\subset X$. That is, $\exists n \in \mathbb{N}$ such that $I_n \subseteq \mathbb{N} \setminus A$ and $I_{n+1} \not\subset \mathbb{N} \setminus A$.
5. In other words, $n + 1 \in A$ and $n \notin A$.
6. Then $n + 1$ is the minimum element of A.

□

2.2 Finite Sets

2.2.1 Formal Definitions

Definition 4 (Finite Set): A set X is **finite** if there exists a bijection $\phi : I_n \to X$ for some $n \in \mathbb{N}$, where $I_n = \{1, 2, \ldots, n\}$. The number n is called the **cardinality** of X, denoted $\#X$.

Definition 5 (Infinite Set): A set is **infinite** if it is not finite.

Definition 6 (Bounded Set): A set $X \subset \mathbb{N}$ is **bounded** if:

$$\exists m \in \mathbb{N} \text{ such that } m \geq n \text{ for all } n \in X \quad (3)$$

2.2.2 Key Theorems

Theorem 3 (Characterization of Finite Sets in $\mathbb{N}$): For $X \subset \mathbb{N}$, the following are equivalent: 1. X is finite. 2. X is bounded. 3. X has a maximum element.

Theorem 4 (Properties of Finite Sets):

Property	Statement
Subset	If X is finite and $Y \subseteq X$, then Y is finite and $\#Y \leq \#X$.
Equality	If $\#Y = \#X$, then $Y = X$.
Union	If X and Y are finite and disjoint, then $\#(X \cup Y) = \#X + \#Y$.
Union (general)	$\#(X \cup Y) \leq \#X + \#Y$.
Injective Image	If $f : Y \to X$ is injective and X is finite, then Y is finite.

2.3 Countable Sets

2.3.1 Motivation

Countable sets are those that can be "listed" or "enumerated" in a sequence. The problem that countability solves is distinguishing between different sizes of infinity. While $\mathbb{N}$ is infinite, so are $\mathbb{Z}$ and $\mathbb{Q}$, but they are the same "size" of infinity. This concept is fundamental for understanding which sets can be indexed by the natural numbers.

Geometric Interpretation: Imagine writing the elements of a set in an infinite list: $a_1, a_2, a_3, \ldots$. If every element of the set appears somewhere in this list, the set is countable. Countability means we can "count" the elements using the natural numbers.

2.3.2 Formal Definitions

Definition 7 (Countable Set): A set X is **countable** if either: 1. X is finite, or 2. There exists a bijection $\mathbb{N} \to X$.

A countable set that is not finite is called **countably infinite**.

2.3.3 Key Theorems

Theorem 5: Every infinite set contains an infinite countable subset.

Proof:

1. For each nonempty $A \subset X$, choose $x_A \in A$ (using the Axiom of Choice).
2. Define $f : \mathbb{N} \to X$ recursively:
   - $f(1) = x_X$
   - $f(n) = x_{X \setminus \{f(1),\ldots,f(n-1)\}}$
3. We show f is injective. Suppose $f(m) = f(n)$ with $m < n$. Then $f(n) \in X \setminus \{f(1), \ldots, f(n-1)\}$, but $f(m)$ was removed, so $f(n) \neq f(m)$, contradiction.
4. Therefore, $f(\mathbb{N})$ is a countable infinite subset of X.

□

Theorem 6: A subset of a countable set is countable.

Proof: If X is countable, there is a bijection $\phi : X \to \mathbb{N}$ or $\phi : X \to I_n$. For $Y \subseteq X$, the restriction $\phi|_Y$ gives an injection from Y into $\mathbb{N}$ or I_n, which implies Y is countable.

□

Theorem 7: If X is countable and $f : X \to Y$ is surjective, then Y is countable.

Proof:

1. Let $g : Y \to X$ be the right inverse of f, so $f(g(y)) = y$ for all $y \in Y$.
2. If $g(y) = g(y')$, then $y = f(g(y)) = f(g(y')) = y'$, so g is injective.
3. Thus g is a bijection between Y and $g(Y) \subset X$.
4. Since $g(Y)$ is a subset of a countable set, it is countable.
5. Therefore, Y is countable.

□

Theorem 8: If X and Y are countable, then $X \times Y$ is countable.

Proof:

1. It suffices to prove that $\mathbb{N} \times \mathbb{N}$ is countable.
2. Define $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by $f(m, n) = 2^m 3^n$.
3. By unique prime factorization, f is injective.
4. Therefore, $\mathbb{N} \times \mathbb{N}$ is in bijection with its image, a subset of $\mathbb{N}$, so it is countable.
5. Since any countable sets are in bijection with subsets of $\mathbb{N}$, their product is countable.

□

Theorem 9: $\mathbb{Q} = \left\{ \frac{p}{q} : p, q \in \mathbb{Z}, q \neq 0 \right\}$ is countable.

Proof:

1. Consider the map $f : \mathbb{Z} \times (\mathbb{Z} \setminus \{0\}) \to \mathbb{Q}$ given by $f(p, q) = \frac{p}{q}$.
2. By Theorem 8, $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ is countable (since $\mathbb{Z}$ is countable).
3. Since f is surjective, by Theorem 7, $\mathbb{Q}$ is countable.

□

Theorem 10: Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of countable sets. Then $\bigcup_{i=1}^{\infty} X_i$ is countable.

Proof:

1. For each $m \in \mathbb{N}$, let $f_m : \mathbb{N} \to X_m$ be a surjective function (exists since X_m is countable).
2. Define $f : \mathbb{N} \times \mathbb{N} \to \bigcup_{i=1}^{\infty} X_i$ by $f(m, n) = f_m(n)$.
3. This function is surjective, since each f_m is onto.
4. Since $\mathbb{N} \times \mathbb{N}$ is countable, the union is countable.

□

2.4 Uncountable Sets

2.4.1 Motivation

The existence of uncountable sets—sets that cannot be listed in a sequence—is one of the most profound results in mathematics. Cantor's diagonal argument shows that $\mathbb{R}$ is strictly larger than $\mathbb{N}$. This has deep implications: not all infinities are equal, and there are mathematical objects that cannot be "enumerated."

Geometric Interpretation: Imagine trying to list all real numbers between 0 and 1. Cantor's diagonal argument shows that any such list will always miss at least one number. The set of real numbers is like a continuous line, while countable sets are like isolated dots.

2.4.2 Cantor's Theorem

Theorem 11 (Cantor's Theorem): Suppose Y contains at least 2 elements and X is any set. There does not exist a surjection $X \to \mathcal{F}(X, Y)$, where $\mathcal{F}(X, Y)$ is the set of all functions from X to Y.

Proof:

1. Suppose $\phi : X \to \mathcal{F}(X, Y)$ is any function.
2. For each $x \in X$, denote by $\phi_x : X \to Y$ the function $\phi(x)$.
3. Define $f : X \to Y$ by $f(x) \neq \phi_x(x)$ for all $x \in X$. Such a function exists because Y has at least 2 elements.
4. By construction, $f \notin \phi(X)$. Therefore, ϕ is not surjective.

□

Theorem 12: For every set X, there exists a set Y of higher cardinality, i.e., $\mathrm{card}(X) < \mathrm{card}(Y)$.

Proof:

1. Consider $\mathcal{F}(X, X)$.
2. There exists an injection $X \to \mathcal{F}(X, X)$, namely $x \mapsto \phi_x$, where $\phi_x(y) = x$ for all $y \in X$.
3. By Cantor's Theorem, there does not exist a surjection $X \to \mathcal{F}(X, X)$.

4. Thus, $\text{card}(X) < \text{card}(\mathcal{F}(X, X))$.

□

Theorem 13: $\mathbb{R}$ is uncountable.

Proof:

1. It suffices to show that $(0, 1)$ is uncountable.
2. Suppose $(0, 1)$ were countable. Then we could list its elements: $x_1, x_2, x_3, \ldots$.
3. Choose a closed interval I_1 such that $x_1 \in I_1$.
4. Use the lemma (every interval contains a subinterval avoiding a given point) to choose $I_2 \subset I_1$ such that $x_1 \notin I_2$.
5. Continue inductively: choose $I_{n+1} \subset I_n$ such that $x_n \notin I_{n+1}$.
6. By the Nested Interval Property, $\cap_{n=1}^{\infty} I_n \neq \emptyset$.
7. Let $x \in \cap_{n=1}^{\infty} I_n$. Then $x \neq x_n$ for all n, since $x_n \notin I_{n+1}$.
8. Therefore, $x \notin \{x_1, x_2, \ldots\}$, contradicting that the list contained all elements of $(0, 1)$.
9. Thus, $(0, 1)$ is uncountable, and so is $\mathbb{R}$.

□

Corollary: $\mathbb{R} \setminus \mathbb{Q}$ is uncountable.

Proof: If $\mathbb{R} \setminus \mathbb{Q}$ were countable, then $\mathbb{R} = \mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})$ would be countable, since $\mathbb{Q}$ is countable. But $\mathbb{R}$ is uncountable. Therefore, $\mathbb{R} \setminus \mathbb{Q}$ is uncountable.

□

2.5 Summary of Key Results

Concept	Definition	Key Formula/Result
Induction Principle	Base case + inductive step	$1 \in X,\ s(X) \subset X \implies X = \mathbb{N}$
Well-Ordering	Every nonempty subset of $\mathbb{N}$ has a minimum	$\min(A)$ exists for all $A \neq \emptyset$
Finite Set	Bijection with I_n	$\#X = n$
Countable Set	Bijection with $\mathbb{N}$ or finite	X can be listed as $x_1, x_2, \ldots$
Countable Union	Union of countably many countable sets	$\bigcup_{i=1}^{\infty} X_i$ is countable
$\mathbb{Q}$ is Countable	Diagonal enumeration	$\mathbb{Q}$ has the same cardinality as $\mathbb{N}$
Cantor's Theorem	No surjection $X \to \mathcal{F}(X, Y)$	$\text{card}(X) < \text{card}(\mathcal{F}(X, X))$
$\mathbb{R}$ is Uncountable	Cantor diagonal argument	$\mathbb{R}$ is strictly larger than $\mathbb{N}$

Chapter 3: Real Numbers

3.1 Fields and Ordered Fields

3.1.1 Motivation

The real numbers $\mathbb{R}$ are the foundation of mathematical analysis. The problem that the axiomatic construction of $\mathbb{R}$ solves is the **formalization of continuity and completeness**. Unlike $\mathbb{Q}$, which has "gaps" (e.g., no rational number satisfies $x^2 = 2$), the real numbers are complete: every nonempty bounded set has a least upper bound. This property is what makes calculus possible.

Geometric Interpretation: Visualize the real numbers as a continuous line with no gaps. The rational numbers are like dots on this line—dense, but not filling the entire line. The completeness axiom fills in all the gaps, ensuring that every cut or boundary point corresponds to a real number.

3.1.2 Field Axioms

Definition 1 (Field): A **field** is a set K equipped with two operations $+$ and $\cdot$ satisfying:

Axiom	Name	Formula
A_1	Commutativity of addition	$x + y = y + x$
A_2	Associativity of addition	$(x + y) + z = x + (y + z)$
A_3	Additive identity	$\exists 0 \in K : x + 0 = x$
A_4	Additive inverse	$\forall x, \exists(-x) : x + (-x) = 0$
M_1	Commutativity of multiplication	$x \cdot y = y \cdot x$
M_2	Associativity of multiplication	$(x \cdot y) \cdot z = x \cdot (y \cdot z)$
M_3	Multiplicative identity	$\exists 1 \in K : x \cdot 1 = x$
M_4	Multiplicative inverse	$\forall x \neq 0, \exists x^{-1} : x \cdot x^{-1} = 1$
D	Distributivity	$x \cdot (y + z) = x \cdot y + x \cdot z$

3.1.3 Ordered Fields

Definition 2 (Ordered Field): An **ordered field** is a field $(K, +, \cdot)$ together with a binary relation $<$ satisfying:

1. **Transitivity:** $(x < y) \wedge (y < z) \implies x < z$.
2. **Trichotomy:** Exactly one of $x = y$, $x < y$, or $y < x$ holds.
3. **Translation invariance:** $x < y \implies x + z < y + z$ for all z.
4. **Multiplication by positive:** $(x < y) \wedge (z > 0) \implies x \cdot z < y \cdot z$.
5. **Multiplication by negative:** $(x < y) \wedge (z < 0) \implies x \cdot z > y \cdot z$.

3.1.4 Key Theorems

Theorem 1 (Properties of Ordered Fields): In an ordered field K:

1. $x > 0 \iff -x < 0$.
2. If $x > 0$ and $y > 0$, then $x + y > 0$ and $x \cdot y > 0$.
3. If $x < y$ and $z < 0$, then $x \cdot z > y \cdot z$.
4. $x^2 \geq 0$ for all x, with equality iff $x = 0$.
5. $0 < 1$.

3.2 Archimedean Property

3.2.1 Motivation

The Archimedean property states that for any positive real number a, there exists a natural number n such that $n \cdot a > 1$. This means that no matter how small a is, by adding it to itself enough times, we can exceed any bound. This property prevents the existence of "infinitely small" positive numbers.

Geometric Interpretation: Imagine a ruler with markings at every integer. The Archimedean property says that if we take a very small step of length a, after finitely many steps we will have traveled a distance greater than any given bound b.

3.2.2 Formal Definitions

Definition 3 (Bounded Above): A subset X of an ordered field K is **bounded above** if:

$$\exists b \in K \text{ such that } x \leq b \text{ for all } x \in X \quad (1)$$

Definition 4 (Bounded Below): A subset X is **bounded below** if:

$$\exists b \in K \text{ such that } x \geq b \text{ for all } x \in X \quad (2)$$

Definition 5 (Archimedean Field): An ordered field K is **Archimedean** if any of the following equivalent properties hold:

1. $\mathbb{N} \subset K$ is unbounded above.
2. $\forall a > 0, \forall b \in K, \exists n \in \mathbb{N}$ such that $n \cdot a > b$.
3. $\forall a > 0, \exists n \in \mathbb{N}$ such that $0 < \frac{1}{n} < a$.

3.2.3 Key Theorems

Theorem 2 (Archimedean Property of $\mathbb{R}$): The real numbers are Archimedean.

Proof:

1. Suppose $\mathbb{N} \subset \mathbb{R}$ were bounded above.
2. By completeness, $\mathbb{N}$ would have a least upper bound: $x = \sup(\mathbb{N})$.
3. Since x is an upper bound, $\forall n \in \mathbb{N}, n \leq x$.
4. Since $1 \in \mathbb{N}$, $x - 1$ is not an upper bound, so $\exists n \in \mathbb{N}$ such that $x - 1 < n$.
5. Then $n + 1 > x$, but $n + 1 \in \mathbb{N}$, contradicting that x is an upper bound.
6. Therefore, $\mathbb{N}$ is not bounded above.

□

Theorem 3 (Density of $\mathbb{Q}$): $\mathbb{Q}$ is dense in $\mathbb{R}$. That is, for all $a, b \in \mathbb{R}$ with $a < b$, there exists $q \in \mathbb{Q}$ such that $a < q < b$.

Proof:

1. Let $a, b \in \mathbb{R}$ with $a < b$. By the Archimedean property, choose $p \in \mathbb{N}$ such that $\frac{1}{p} < b - a$.
2. Consider the set $A = \left\{ m \in \mathbb{Z} : \frac{m}{p} \geq b \right\}$.
3. By the well-ordering property of $\mathbb{Z}$, A has a least element m_0.
4. Then $\frac{m_0 - 1}{p} < b$.
5. Claim: $\frac{m_0 - 1}{p} > a$. If not, then $\frac{m_0 - 1}{p} \leq a < b < \frac{m_0}{p}$.
6. Then $b - a \leq \frac{m_0}{p} - \frac{m_0 - 1}{p} = \frac{1}{p}$, a contradiction.
7. Thus, $a < \frac{m_0 - 1}{p} < b$, so $(a, b) \cap \mathbb{Q} \neq \emptyset$.

□

Theorem 4 (Density of Irrationals): $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$.

Proof: Similar to the proof for $\mathbb{Q}$, using intervals of the form $\frac{m}{p\sqrt{2}}$.

□

3.3 Completeness and Suprema

3.3.1 Motivation

The completeness axiom is the key property that distinguishes $\mathbb{R}$ from $\mathbb{Q}$. It states that every nonempty subset of $\mathbb{R}$ that is bounded above has a least upper bound (supremum) in $\mathbb{R}$. This property ensures that there are no "gaps" in the real line and is the foundation for limits, continuity, and the existence of solutions to equations.

Geometric Interpretation: Visualize a set of points on the real line bounded above. The supremum is the point that marks the "frontier" of the set—the smallest number that is greater than or equal to every point in the set. In $\mathbb{R}$, this frontier always exists.

3.3.2 Formal Definitions

Definition 6 (Supremum): For a set $X \subset K$ that is bounded above, the **supremum** of X, denoted $\sup X$, is the least upper bound of X:

1. $x \leq \sup X$ for all $x \in X$.
2. If $b \in K$ is such that $x \leq b$ for all $x \in X$, then $\sup X \leq b$.

Equivalently: If $b < \sup X$, then $\exists x \in X$ such that $b < x \leq \sup X$.

Definition 7 (Infimum): For a set $X \subset K$ that is bounded below, the **infimum** of X, denoted $\inf X$, is the greatest lower bound of X.

3.3.3 Key Theorems

Theorem 5 (Fundamental Axiom of Analysis): There exists a complete ordered field, called the field of real numbers $\mathbb{R}$. Completeness means: every nonempty subset of $\mathbb{R}$ that is bounded above has a supremum in $\mathbb{R}$.

Theorem 6 (Nested Interval Property): Let $I_n = [a_n, b_n]$, with $a_n < b_n$, be a sequence of closed intervals such that $I_1 \supset I_2 \supset \cdots$. Then:

$$\bigcap_{n=1}^{\infty} I_n = [a, b]$$

where $a = \sup_{n\in\mathbb{N}} a_n$ and $b = \inf_{n\in\mathbb{N}} b_n$.

Proof:

1. Since $I_1 \supset I_2 \supset \cdots$, the sequences (a_n) and (b_n) are monotone: a_n is nondecreasing, b_n is nonincreasing.
2. Since $a_n \leq b_1$ for all n, (a_n) is bounded above, so $a = \sup a_n$ exists.
3. Similarly, $b = \inf b_n$ exists.
4. Since $a_n \leq b_n$ for all n, and the sequences are monotone, $a \leq b$.
5. $[a, b] \subset I_n$ for all n, so $[a, b] \subset \bigcap_{n=1}^{\infty} I_n$.

6. Conversely, if $x \in \bigcap_{n=1}^{\infty} I_n$, then $a_n \leq x \leq b_n$ for all n.
7. Taking suprema and infima: $a \leq x \leq b$, so $x \in [a, b]$.
8. Thus, $\bigcap_{n=1}^{\infty} I_n = [a, b]$.

□

Theorem 7 (Existence of nth Roots): For all $n \in \mathbb{N}$ and all $a > 0 \in \mathbb{R}$, there exists a unique $x > 0 \in \mathbb{R}$ such that $x^n = a$.

Proof: (Existence follows from completeness; uniqueness follows from monotonicity of $x \mapsto x^n$.)

□

3.4 Irrationality of $\sqrt{2}$

3.4.1 Motivation

The discovery that $\sqrt{2}$ is irrational was a major milestone in mathematics. It showed that $\mathbb{Q}$ has gaps—there are lengths that cannot be expressed as ratios of integers. This motivated the development of the real numbers.

3.4.2 Key Theorems

Theorem 8: There does not exist $x \in \mathbb{Q}$ with $x^2 = 2$.

Proof:

1. Suppose, towards contradiction, that there exist $p, q \in \mathbb{Z}$, $q \neq 0$, with $\left(\frac{p}{q}\right)^2 = 2$.
2. Then $p^2 = 2q^2$.
3. Consider the prime factorization of p^2 and q^2.
4. In p^2, the factor 2 appears an even number of times (since p^2 is a square).
5. In $2q^2$, the factor 2 appears an odd number of times (one from the factor 2, plus an even number from q^2).
6. By unique prime factorization, the number of times 2 appears in the prime decomposition must be the same on both sides.
7. This is a contradiction. Therefore, no rational x satisfies $x^2 = 2$.

□

3.5 Summary of Key Results

Concept	Definition	Key Formula/Result
Field	Set with $+$ and $\cdot$ satisfying field axioms	$\mathbb{Q}, \mathbb{R}$ are fields
Ordered Field	Field with compatible order relation	$\mathbb{Q}, \mathbb{R}$ are ordered fields
Archimedean Property	$\forall a > 0, \exists n \in \mathbb{N} : n \cdot a > 1$	$\mathbb{R}$ is Archimedean
Completeness	Every bounded set has a supremum	$\mathbb{R}$ is complete; $\mathbb{Q}$ is not
Supremum	Least upper bound	$\sup X = \min\{b : x \leq b \text{ for all } x \in X\}$

Concept	Definition	Key Formula/Result
Infimum	Greatest lower bound	$\inf X = \max\{b : b \leq x \text{ for all } x \in X\}$
Density of $\mathbb{Q}$	Between any two reals, there is a rational	$a < b \implies \exists q \in \mathbb{Q} : a < q < b$
Nested Interval Property	Nested closed intervals have nonempty intersection	$\cap I_n = [\sup a_n, \inf b_n]$
Irrationality of $\sqrt{2}$	$\sqrt{2} \notin \mathbb{Q}$	Unique prime factorization contradiction

---

Chapter 4: Sequences

4.1 Basic Definitions and Properties

4.1.1 Motivation

The problem that sequences solve is the **formalization of limiting processes**. Many mathematical concepts—such as the value of infinite sums, the slope of a curve, or the area under a curve—depend on understanding what happens as we take more and more terms or finer and finer partitions. Sequences provide the language for describing this "limit" behavior.

Geometric Interpretation in $\mathbb{R}$: Visualize a sequence as a point moving along the real line, with each term x_n representing the position at time n. Convergence means the point eventually settles down near a fixed location a, staying within any prescribed distance ϵ of a for all sufficiently large times.

4.1.2 Formal Definitions

Definition 1 (Sequence): A **sequence** of real numbers is a function $x : \mathbb{N} \to \mathbb{R}$. We often write $(x_n)_{n\in\mathbb{N}}$ or simply (x_n).

Definition 2 (Bounded Sequence): A sequence (x_n) is **bounded** if the set $\{x_n : n \in \mathbb{N}\}$ is bounded.

Definition 3 (Monotone Sequence): A sequence is:

- **Non-decreasing (increasing)** if $x_n \leq x_{n+1}$ for all n.
- **Non-increasing (decreasing)** if $x_n \geq x_{n+1}$ for all n.

Definition 4 (Subsequence): A **subsequence** of $(x_n)_{n\in\mathbb{N}}$ is $(x_n)_{n\in\mathbb{N}'}$, where $\mathbb{N}'$ is an infinite subset of $\mathbb{N}$. We write $\mathbb{N}' = \{n_1 < n_2 < \cdots\}$ and denote the subsequence by $(x_{n_k})_{k\in\mathbb{N}}$.

4.1.3 Key Theorems

Theorem 1 (Bounded Monotone Sequences): A monotone sequence is bounded iff it has a bounded subsequence.

Proof for non-decreasing case:

1. ($\Rightarrow$) Let (x_n) be a bounded non-decreasing sequence. Then $\exists b \in \mathbb{R}$ such that $x_n \leq b$ for all n. Let (x_{k_n}) be any subsequence. Since $x_{k_n} \in (x_n)$, $x_{k_n} \leq b$. Also, since (x_n) is non-decreasing, $x_1 \leq x_{k_n}$. Thus, the subsequence is bounded.

2. ($\Leftarrow$) Let (x_{k_n}) be a bounded subsequence of a non-decreasing sequence (x_n). Let b be an upper bound of (x_{k_n}). Since (x_{k_n}) is a subsequence, for every n there exists $k_n \geq n$. Since (x_n) is non-decreasing, $x_n \leq x_{k_n} \leq b$. Thus, b is an upper bound for (x_n).

$\square$

4.2 Convergence of Sequences

4.2.1 Motivation

The definition of convergence is the heart of analysis. It formalizes the intuitive idea that a sequence "approaches" a limit a. The ϵ-n_0 definition provides a rigorous way to express this without relying on vague notions of "closeness."

Geometric Interpretation: The sequence converges to a if, for every ϵ-neighborhood around a, all sufficiently far terms of the sequence lie inside that neighborhood. Think of a as a target: eventually, the sequence stays within any desired distance ϵ of the target.

4.2.2 Formal Definitions

Definition 5 (Convergence): A sequence (x_n) **converges** to $a \in \mathbb{R}$ if:

$$\forall \epsilon > 0, \exists n_0 \in \mathbb{N} \text{ such that } \forall n \geq n_0, |x_n - a| < \epsilon \quad (1)$$

We write $\lim x_n = a$ or $x_n \to a$.

4.2.3 Key Theorems

Theorem 2 (Uniqueness of Limits): If $x_n \to a$ and $x_n \to b$, then $a = b$.

Proof:

1. Fix $\epsilon > 0$.
2. Since $x_n \to a$, $\exists n_0$ such that $|x_n - a| < \frac{\epsilon}{2}$ for all $n \geq n_0$.
3. Since $x_n \to b$, $\exists n_1$ such that $|x_n - b| < \frac{\epsilon}{2}$ for all $n \geq n_1$.
4. Let $N = n_0 + n_1$. Then:

$$|a - b| = |a - x_N + x_N - b| \leq |a - x_N| + |x_N - b| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

5. Since this holds for all $\epsilon > 0$, $|a - b| = 0$, so $a = b$.

□

Theorem 3 (Bounded Monotone Convergence): A bounded non-decreasing sequence converges to its supremum.

Proof:

1. Let (x_n) be bounded and non-decreasing. Let $a = \sup_{n \in \mathbb{N}} x_n$.
2. Fix $\epsilon > 0$. Since $a - \epsilon < a$, $a - \epsilon$ is not an upper bound.
3. Therefore, $\exists n_0 \in \mathbb{N}$ such that $x_{n_0} > a - \epsilon$.
4. Since (x_n) is non-decreasing, for all $n \geq n_0$:

$$a \geq x_n \geq x_{n_0} > a - \epsilon$$

5. Thus, $x_n \in (a - \epsilon, a + \epsilon)$ for all $n \geq n_0$.
6. Therefore, $x_n \to a$.

□

4.3 Properties of Limits

4.3.1 Key Theorems

Theorem 4 (Algebra of Limits): Let (x_n) and (y_n) be convergent sequences. Then:

1. **Sum:** $\lim(x_n + y_n) = \lim x_n + \lim y_n$.
2. **Product:** $\lim(x_n y_n) = (\lim x_n)(\lim y_n)$.
3. **Quotient:** If $\lim y_n \neq 0$, then $\lim \left(\frac{x_n}{y_n}\right) = \frac{\lim x_n}{\lim y_n}$.
4. **Bounded times zero:** If $\lim x_n = 0$ and (y_n) is bounded, then $\lim x_n y_n = 0$.

Proof of (4):

1. Let b satisfy $|y_n| \leq b$ for all n.
2. Fix $\epsilon > 0$. Since $x_n \to 0$, $\exists n_0$ such that $|x_n| < \frac{\epsilon}{b}$ for all $n \geq n_0$.
3. Then $|x_n y_n| = |x_n||y_n| \leq |x_n|b < \epsilon$ for all $n \geq n_0$.
4. Therefore, $x_n y_n \to 0$.

□

Theorem 5 (Order Properties):

1. If $\lim x_n > a$, then $\exists n_0$ such that $x_n > a$ for all $n \geq n_0$.
2. If $x_n \leq y_n$ for all n, then $\lim x_n \leq \lim y_n$.
3. **Sandwich Theorem:** If $x_n \leq z_n \leq y_n$ and $\lim x_n = \lim y_n = a$, then $\lim z_n = a$.

Proof of (1):

1. Let $b = \lim x_n > a$.
2. Choose $\epsilon > 0$ such that $0 < \epsilon < b - a$.
3. Since $x_n \to b$, $\exists n_0$ such that $|x_n - b| < \epsilon$ for all $n \geq n_0$.
4. Then $x_n > b - \epsilon > a$ for all $n \geq n_0$.

□

4.4 Cluster Points and Subsequences

4.4.1 Motivation

A cluster point is a value that the sequence approaches infinitely often. Even if a sequence does not converge, it may still have subsequences that converge to different limits. Understanding cluster points is essential for describing the behavior of non-convergent sequences.

Geometric Interpretation: A cluster point is a value around which infinitely many terms of the sequence accumulate. Think of it as a point that the sequence visits infinitely many times, or gets arbitrarily close to infinitely many times.

4.4.2 Formal Definitions

Definition 6 (Cluster Point): A point $a \in \mathbb{R}$ is a **cluster point** (or accumulation point) of a sequence (x_k) if any of the following equivalent statements hold:

1. There exists a subsequence (x_{k_j}) converging to a.
2. For all $\epsilon > 0$, $\{k \in \mathbb{N} : x_k \in (a - \epsilon, a + \epsilon)\}$ is infinite.

3. For all $\epsilon > 0$, the set $\{k \in \mathbb{N} : x_k \in (a - \epsilon, a + \epsilon)\}$ is nonempty.

4.4.3 Key Theorems

Theorem 6 (Cluster Point Characterization): For a bounded sequence (x_k):

1. $\limsup x_k$ is the greatest cluster point of (x_k).
2. $\liminf x_k$ is the least cluster point of (x_k).

Where:

$$\limsup x_k = \lim_{k\to\infty} \left(\sup_{j\geq k} x_j \right), \quad \liminf x_k = \lim_{k\to\infty} \left(\inf_{j\geq k} x_j \right)$$

4.5 Bolzano-Weierstrass Theorem

4.5.1 Motivation

The Bolzano-Weierstrass theorem states that every bounded sequence in $\mathbb{R}$ has a convergent subsequence. This is one of the most important results in analysis: it guarantees that bounded sequences cannot "escape to infinity" without having some accumulation behavior.

Geometric Interpretation: No matter how a bounded sequence behaves, there is always some point around which infinitely many terms cluster. The sequence may oscillate wildly, but it cannot do so without having at least one point of accumulation.

4.5.2 Key Theorems

Theorem 7 (Bolzano-Weierstrass): Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.

Proof:

1. Let (x_k) be a bounded sequence.
2. By the lemma, $\limsup x_k$ and $\liminf x_k$ are cluster points.
3. By definition of cluster point, there exists a subsequence converging to $\limsup x_k$.
4. Therefore, (x_k) has a convergent subsequence.

□

4.6 Cauchy Sequences

4.6.1 Motivation

Cauchy sequences provide a criterion for convergence that does not require knowing the limit in advance. A sequence is Cauchy if its terms get arbitrarily close to each other as the sequence progresses. In $\mathbb{R}$, this condition is equivalent to convergence—a property known as completeness.

Geometric Interpretation: Think of a Cauchy sequence as a sequence where the points are eventually all clustered together. They don't necessarily have to approach a pre-specified point, but they must be getting closer and closer to each other.

4.6.2 Formal Definitions

Definition 7 (Cauchy Sequence): A sequence (x_k) in $\mathbb{R}$ is **Cauchy** if:

$$\forall \epsilon > 0, \exists k_0 \in \mathbb{N} \text{ such that } k, j \geq k_0 \implies |x_k - x_j| < \epsilon \quad (2)$$

4.6.3 Key Theorems

Theorem 8: Every Cauchy sequence in $\mathbb{R}$ is convergent.

Proof:

1. Let (x_k) be a Cauchy sequence.
2. First, show (x_k) is bounded. Choose k_0 such that $|x_k - x_{k_0}| < 1$ for all $k \geq k_0$.
3. Then for $k \geq k_0$, $|x_k| \leq |x_{k_0}| + 1$.
4. Let $M = \max\{|x_1|, \dots, |x_{k_0-1}|, |x_{k_0}| + 1\}$. Then $|x_k| \leq M$ for all k, so the sequence is bounded.
5. By Bolzano-Weierstrass, (x_k) has a convergent subsequence (x_{k_j}) with limit x.
6. Show $x_k \to x$. Fix $\epsilon > 0$.
7. Since (x_k) is Cauchy, $\exists k_1$ such that $|x_k - x_j| < \frac{\epsilon}{2}$ for all $k, j \geq k_1$.
8. Since $x_{k_j} \to x$, choose j_0 such that $k_j \geq k_1$ and $|x_{k_j} - x| < \frac{\epsilon}{2}$.
9. For any $k \geq k_1$:
$$|x_k - x| \leq |x_k - x_{k_j}| + |x_{k_j} - x| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$
10. Therefore, $x_k \to x$.

□

4.7 Summary of Key Results

Concept	Definition	Key Formula/Result
Sequence	Function $x : \mathbb{N} \to \mathbb{R}$	$(x_n)_{n\in\mathbb{N}}$
Bounded Sequence	$\{x_n\}$ bounded	$\exists M : \lvert x_n\rvert \leq M \ \forall n$
Monotone Sequence	Non-decreasing or non-increasing	$x_n \leq x_{n+1}$ or $x_n \geq x_{n+1}$
Convergence	$x_n \to a$	$\forall \epsilon > 0, \exists n_0 : n \geq n_0 \implies \lvert x_n - a\rvert < \epsilon$
Subsequence	(x_{n_k}) with $n_k \to \infty$	Restriction to infinite subset
Cluster Point	Limit of subsequence	a is cluster point iff subsequence converges to a
Bolzano-Weierstrass	Every bounded sequence has convergent subsequence	Fundamental compactness result
Cauchy Sequence	Terms get arbitrarily close	$\forall \epsilon > 0, \exists k_0 : k, j \geq k_0 \implies \lvert x_k - x_j\rvert < \epsilon$
Cauchy Criterion	Cauchy iff convergent in $\mathbb{R}$	Completeness of $\mathbb{R}$
Limit Algebra	Limits preserve arithmetic operations	$\lim(x_n + y_n) = \lim x_n + \lim y_n$
Sandwich Theorem	Squeeze between two sequences	$x_n \leq z_n \leq y_n, \lim x_n = \lim y_n = a \implies \lim z_n = a$

Chapter 5: Vector Spaces

5.1 Definition and Basic Properties

5.1.1 Motivation

The fundamental problem that vector spaces solve is the **abstraction of linear structure**. Many mathematical objects—geometric vectors in $\mathbb{R}^n$, polynomials, functions, matrices—share common algebraic properties: they can be added together and multiplied by scalars. By formalizing these properties, we can study all such objects simultaneously using the same tools of linear algebra.

Geometric Interpretation in $\mathbb{R}^2$: Visualize a vector space as a plane where vectors are arrows starting at the origin. Addition corresponds to the parallelogram rule: placing vectors head-to-tail. Scalar multiplication stretches or shrinks a vector, or reverses its direction if the scalar is negative. The zero vector is the point at the origin.

5.1.2 Formal Definitions

Definition 1 (Vector Space): A **vector space** over a field K is a set V equipped with operations $+$ (addition) and $\cdot$ (scalar multiplication) satisfying:

Axiom	Name	Formula
V1	Associativity of addition	$(x+y)+z = x+(y+z)$
V2	Commutativity of addition	$x+y = y+x$
V3	Additive identity	$\exists 0 \in V : x+0 = x$
V4	Additive inverse	$\forall x, \exists(-x) : x+(-x) = 0$
V5	Multiplicative identity	$1 \cdot x = x$
V6	Associativity of scalar multiplication	$(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$
V7	Distributivity (scalar over vector)	$(\alpha+\beta) \cdot x = \alpha \cdot x + \beta \cdot x$
V8	Distributivity (vector over scalar)	$\alpha \cdot (x+y) = \alpha \cdot x + \alpha \cdot y$

5.1.3 Key Theorems

Theorem 1 (Basic Properties): In any vector space V over a field K:

1. $0 \cdot x = 0$ for all $x \in V$.
2. $\alpha \cdot 0 = 0$ for all $\alpha \in K$.
3. $(-1) \cdot x = -x$ for all $x \in V$.
4. If $\alpha \cdot x = 0$, then $\alpha = 0$ or $x = 0$.

Proof:

1. $0 \cdot x = (0+0) \cdot x = 0 \cdot x + 0 \cdot x$. Subtracting $0 \cdot x$ from both sides gives $0 \cdot x = 0$.
2. $\alpha \cdot 0 = \alpha \cdot (0+0) = \alpha \cdot 0 + \alpha \cdot 0$. Subtracting $\alpha \cdot 0$ gives $\alpha \cdot 0 = 0$.
3. $x + (-1) \cdot x = 1 \cdot x + (-1) \cdot x = (1+(-1)) \cdot x = 0 \cdot x = 0$. Thus, $(-1) \cdot x = -x$.
4. If $\alpha \neq 0$, then $\alpha^{-1}(\alpha \cdot x) = (\alpha^{-1}\alpha) \cdot x = 1 \cdot x = x$. But $\alpha^{-1}(\alpha \cdot x) = \alpha^{-1} \cdot 0 = 0$. Hence $x = 0$.

□

5.2 Linear Independence and Span

5.2.1 Motivation

The concepts of linear independence and span solve the problem of **describing a vector space efficiently**. Instead of listing every vector in a space, we can find a small set of "building blocks" that generate the entire space. Linear independence ensures that these building blocks are minimal—no block is redundant.

Geometric Interpretation in $\mathbb{R}^2$: Two vectors are linearly independent if they point in different directions (not multiples of each other). Their span is the entire plane. Three vectors in $\mathbb{R}^2$ are always linearly dependent: one can be written as a combination of the other two. A basis is a set of independent vectors that "covers" the entire space.

5.2.2 Formal Definitions

Definition 2 (Linear Combination): A **linear combination** of vectors $u_1, \dots, u_n \in V$ is an expression:

$$\alpha_1 u_1 + \cdots + \alpha_n u_n \quad (1)$$

for some scalars $\alpha_1, \dots, \alpha_n \in K$.

Definition 3 (Linear Independence): A subset $U \subset V$ is **linearly independent** if, for any finite collection $\{u_1, \dots, u_n\} \subset U$:

$$\alpha_1 u_1 + \cdots + \alpha_n u_n = 0 \implies \alpha_1 = \cdots = \alpha_n = 0 \quad (2)$$

Definition 4 (Linear Dependence): A subset $U \subset V$ is **linearly dependent** if it is not linearly independent. Equivalently, some vector $u_n \in U$ can be written as a linear combination of other vectors in U:

$$u_n = \alpha_1 u_1 + \cdots + \alpha_{n-1} u_{n-1} \quad (3)$$

Definition 5 (Span): The **span** of a subset $U \subset V$ is:

$$\text{span}(U) = \{\alpha_1 u_1 + \cdots + \alpha_n u_n : n \in \mathbb{N}, u_i \in U, \alpha_i \in K\} \quad (4)$$

Definition 6 (Basis): A **basis** of V is a linearly independent subset of V that spans V.

Definition 7 (Finite-Dimensional): A vector space is **finite-dimensional** if it has a finite basis.

5.2.3 Key Theorems

Theorem 2 (Homogeneous System Lemma): A homogeneous linear system with more unknowns than equations ($m < n$) admits a non-trivial solution.

Proof:

1. **Case $m = 1$:** The equation $a_{11}x_1 + \cdots + a_{1n}x_n = 0$ with $n > 1$ has a non-trivial solution. If $a_{1n} \neq 0$, then:
$$x_n = -\frac{a_{11}}{a_{1n}}x_1 - \cdots - \frac{a_{1n-1}}{a_{1n}}x_{n-1}$$
Choosing non-trivial values for $x_1, \dots, x_{n-1}$ gives a non-trivial solution.

2. **Case $m > 1$:** Suppose the result holds for $m-1$ equations. Assume $a_{mn} \neq 0$. From the m-th equation:
$$x_n = -\frac{a_{m1}}{a_{mn}}x_1 - \cdots - \frac{a_{mn-1}}{a_{mn}}x_{n-1}$$
Substituting into the first $m-1$ equations gives a system with $m-1$ equations and $n-1$ unknowns. Since $n-1 > m-1$, by induction, it has a non-trivial solution $(\alpha_1, \dots, \alpha_{n-1})$. Then define:
$$\alpha_n = -\frac{a_{m1}}{a_{mn}}\alpha_1 - \cdots - \frac{a_{mn-1}}{a_{mn}}\alpha_{n-1}$$
This gives a non-trivial solution for the original system.

□

Theorem 3 (Spanning Set Property): If vectors $v_1, \dots, v_m$ span V, then any set of $n > m$ vectors in V is linearly dependent.

Proof:

1. Write each w_j as a linear combination of the v_i's:

$$w_j = \alpha_{1j} v_1 + \cdots + \alpha_{mj} v_m, \quad j = 1, \dots, n$$

2. We seek coefficients $x_1, \dots, x_n$, not all zero, such that $x_1 w_1 + \cdots + x_n w_n = 0$.
3. Substituting:

$$\left(\sum_{j=1}^{n} x_j \alpha_{1j} \right) v_1 + \cdots + \left(\sum_{j=1}^{n} x_j \alpha_{mj} \right) v_m = 0$$

4. It suffices that:

$$\alpha_{11} x_1 + \cdots + \alpha_{1n} x_n = 0, \quad \dots, \quad \alpha_{m1} x_1 + \cdots + \alpha_{mn} x_n = 0$$

5. Since $n > m$, by Theorem 2, this system has a non-trivial solution.

□

Corollary 1: If vectors $v_1, \dots, v_m$ span V, and vectors $u_1, \dots, u_n$ are linearly independent, then $n \leq m$.

Corollary 2: Every basis of a finite-dimensional vector space has the same number of elements. This number is called the **dimension** of V, denoted $\dim V$.

Proof:

1. Let $B = \{u_1, \dots, u_n\}$ and $B' = \{v_1, \dots, v_m\}$ be two bases.
2. Since B' spans V and B is linearly independent, $n \leq m$ (by Corollary 1).
3. Since B spans V and B' is linearly independent, $m \leq n$.
4. Therefore, $n = m$.

□

Theorem 4 (Equivalent Characterizations of a Basis): The following statements are equivalent for $B \subset V$:

1. B is a basis of V.
2. B is **maximally linearly independent**: no proper superset of B is linearly independent.
3. B is **minimally spanning**: no proper subset of B spans V.

5.3 Inner Products

5.3.1 Motivation

The problem that inner products solve is the **introduction of geometric structure** into vector spaces. While vector spaces allow addition and scalar multiplication, they lack notions of length, angle, and orthogonality. An inner product equips a vector space with these geometric concepts, enabling us to measure distances and define projections.

Geometric Interpretation in $\mathbb{R}^2$: The standard inner product $\langle x, y\rangle = x_1y_1 + x_2y_2$ measures the projection of one vector onto another. The norm $\|x\| = \sqrt{\langle x, x\rangle}$ is the length of the vector. Two vectors are orthogonal if their inner product is zero, meaning they are perpendicular.

5.3.2 Formal Definitions

Definition 8 (Inner Product): An **inner product** on a real vector space V is a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ satisfying:

1. **Symmetry:** $\langle x, y\rangle = \langle y, x\rangle$ for all $x, y \in V$.
2. **Bilinearity:** For all $x, y, z \in V$ and $\alpha \in \mathbb{R}$:
   - $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
   - $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$
   - $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$
   - $\langle x, \alpha y\rangle = \alpha\langle x, y\rangle$
3. **Positive Definiteness:** $\langle x, x\rangle \geq 0$ for all $x \in V$, and $\langle x, x\rangle = 0 \iff x = 0$.

Example: The Euclidean inner product on $\mathbb{R}^n$:

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i \quad (5)$$

5.3.3 Key Theorems

Theorem 5 (Cauchy-Schwarz Inequality): For any inner product space:

$$\boxed{|\langle x, y\rangle| \leq \|x\|\|y\|}$$

where $\|x\| = \sqrt{\langle x, x\rangle}$.

Proof (for $\|x\| = \|y\| = 1$):

1. Let $z = x - \alpha y$, where $\alpha = \langle x, y\rangle$.
2. Then $\langle z, y\rangle = \langle x - \alpha y, y\rangle = \langle x, y\rangle - \alpha\langle y, y\rangle = \alpha - \alpha = 0$.
3. Thus z is orthogonal to y.
4. Now:
$$\|x\|^2 = \langle z + \alpha y, z + \alpha y\rangle = \|z\|^2 + \alpha^2\|y\|^2 = \|z\|^2 + \alpha^2$$
5. Since $\|z\|^2 \geq 0$ and $\|x\|^2 = 1$, we have $1 \geq \alpha^2$.
6. Therefore, $|\langle x, y\rangle| \leq 1$.
7. For the general case, set $u = x/\|x\|$ and $v = y/\|y\|$. Then:
$$|\langle u, v\rangle| \leq 1 \implies \frac{|\langle x, y\rangle|}{\|x\|\|y\|} \leq 1$$
   Hence, $|\langle x, y\rangle| \leq \|x\|\|y\|$.

□

Theorem 6 (Norm from Inner Product): If $\langle\cdot,\cdot\rangle$ is an inner product, then $\|x\| = \sqrt{\langle x, x\rangle}$ is a norm.

Proof:

1. **Triangle Inequality:** $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle \leq \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = (\|x\| + \|y\|)^2$.

2. **Homogeneity:** $\|\alpha x\| = \sqrt{\langle \alpha x, \alpha x \rangle} = \sqrt{\alpha^2 \langle x, x \rangle} = |\alpha| \|x\|$.
3. **Positive Definiteness:** Since $\langle x, x \rangle \geq 0$, $\|x\| \geq 0$, and $\|x\| = 0 \iff x = 0$.

□

5.4 Norms and Distances

5.4.1 Motivation

Norms generalize the concept of length to arbitrary vector spaces. They allow us to measure the "size" of vectors and the "distance" between them. In $\mathbb{R}^n$, different norms (Euclidean, maximum, sum) provide different ways of measuring distance, but they are all equivalent in the sense that they generate the same topology.

Geometric Interpretation: The unit ball of a norm—the set of vectors with norm less than or equal to 1—visualizes the norm. For the Euclidean norm, it's a circle (in $\mathbb{R}^2$); for the maximum norm, it's a square; for the sum norm, it's a diamond.

5.4.2 Formal Definitions

Definition 9 (Norm): A **norm** on a vector space V is a function $\|\cdot\| : V \to \mathbb{R}$ satisfying:

1. **Triangle Inequality:** $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.
2. **Homogeneity:** $\|\alpha x\| = |\alpha| \|x\|$ for all $\alpha \in \mathbb{R}$ and $x \in V$.
3. **Positive Definiteness:** $\|x\| \geq 0$, and $\|x\| = 0 \iff x = 0$.

Definition 10 (Metric/Distance): A **metric** on a set X is a function $d : X \times X \to \mathbb{R}$ satisfying:

1. **Non-negativity:** $d(x, y) \geq 0$, and $d(x, y) = 0 \iff x = y$.
2. **Symmetry:** $d(x, y) = d(y, x)$.
3. **Triangle Inequality:** $d(x, z) \leq d(x, y) + d(y, z)$.

Proposition: If $\|\cdot\|$ is a norm on V, then $d(x, y) = \|x - y\|$ is a metric on V.

Definition 11 (Equivalent Norms): Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on $\mathbb{R}^n$ are **equivalent** if there exist constants $c > 0$ and $C > 0$ such that:

$$c\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1 \quad \forall x \in \mathbb{R}^n \quad (6)$$

5.4.3 Key Theorems

Theorem 7 (Examples of Norms on $\mathbb{R}^n$): The following are norms on $\mathbb{R}^n$:

Norm	Formula
Maximum Norm	$\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$
Sum Norm	$\|x\|_1 = |x_1| + \cdots + |x_n|$
***p*-Norm**	$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$ for $p \geq 1$
Euclidean Norm	$\|x\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$

Theorem 8 (Equivalence of Norms in $\mathbb{R}^n$): All norms on $\mathbb{R}^n$ are equivalent.

Proof Sketch:

1. It suffices to show that any norm $\|\cdot\|$ is equivalent to the sum norm $\|\cdot\|_1$.

2. **Upper bound:** Let e_i be the canonical basis. For any $x = \sum x_i e_i$:

$$\|x\| \leq \sum |x_i| \|e_i\| \leq M \|x\|_1$$

where $M = \max\{\|e_1\|, \ldots, \|e_n\|\}$.

3. **Lower bound:** Suppose no such lower bound exists. Then $\forall k \in \mathbb{N}$, $\exists x_k$ with $\|x_k\|_1 = 1$ and $\|x_k\| < 1/k$.
4. By Bolzano-Weierstrass in $\mathbb{R}^n$ (with $\|\cdot\|_1$), (x_k) has a convergent subsequence $x_{k_j} \to x$ in $\|\cdot\|_1$.
5. Then $\|x\|_1 = \lim \|x_{k_j}\|_1 = 1$, so $x \neq 0$.
6. But $\|x\| = \lim \|x_{k_j}\| = 0$, so $x = 0$, a contradiction.
7. Therefore, there exists $c > 0$ such that $c\|x\|_1 \leq \|x\|$.

□

5.5 Summary of Key Results

Concept	Definition	Key Formula/Result
Vector Space	Set with addition and scalar multiplication	Satisfies 8 axioms
Linear Independence	No non-trivial linear combination equals zero	$\sum \alpha_i u_i = 0 \implies \alpha_i = 0$
Span	Set of all linear combinations	$\mathrm{span}(U) = \{\sum \alpha_i u_i : u_i \in U, \alpha_i \in K\}$
Basis	Linearly independent spanning set	All bases have the same cardinality
Dimension	Number of elements in a basis	$\dim V$
Inner Product	Bilinear, symmetric, positive definite form	$\langle x, y \rangle$
Cauchy-Schwarz	$\lvert\langle x, y \rangle\rvert \leq \|x\|\|y\|$	Fundamental inequality
Norm	Length function	Triangle inequality, homogeneity, positive definite
Equivalent Norms	$c\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1$	All norms on $\mathbb{R}^n$ are equivalent
Metric	Distance function	From norm: $d(x, y) = \|x - y\|$

Chapter 6: Topology

6.1 Convergence in $\mathbb{R}^n$

6.1.1 Motivation

The problem that topology solves is the **formalization of "closeness"** and "limits." In $\mathbb{R}$, convergence is defined using absolute value. In $\mathbb{R}^n$, we need a notion of distance that works for vectors. Topology provides the language to discuss open sets, closed sets, and continuity in a way that is independent of any particular norm, relying on the concept of a metric instead.

Geometric Interpretation in $\mathbb{R}^n$: Visualize convergence in $\mathbb{R}^n$ as a sequence of points approaching a limit point. The "distance" between points is measured by the chosen norm. In $\mathbb{R}^2$, convergence means that the points eventually lie within any small disk centered at the limit.

6.1.2 Formal Definitions

Definition 1 (Convergence in $\mathbb{R}^n$): A sequence (x_k) in $\mathbb{R}^n$ **converges** to $x \in \mathbb{R}^n$ if:

$$\forall \epsilon > 0, \exists k_0 \in \mathbb{N} \text{ such that } k \geq k_0 \implies \|x_k - x\| < \epsilon \quad (1)$$

6.1.3 Key Theorems

Theorem 1 (Coordinate-wise Convergence): A sequence (x_k) in $\mathbb{R}^n$ converges to $x \in \mathbb{R}^n$ iff each coordinate converges:

$$x_k \to x \iff x_{ki} \to x_i \quad \forall i = 1, \ldots, n \quad (2)$$

Proof:

1. ($\Rightarrow$) If $x_k \to x$, then $\|x_k - x\|_s = \sum_{i=1}^n |x_{ki} - x_i| \to 0$. Hence, each $|x_{ki} - x_i| \leq \|x_k - x\|_s \to 0$, so $x_{ki} \to x_i$.
2. ($\Leftarrow$) If $x_{ki} \to x_i$ for each i, then:

$$\|x_k - x\|_s = \sum_{i=1}^{n} |x_{ki} - x_i| \to \sum_{i=1}^{n} 0 = 0$$

   Hence, $x_k \to x$ in any norm (since all norms are equivalent).

□

Theorem 2 (Bolzano-Weierstrass in $\mathbb{R}^n$): Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Proof:

1. Let (x_k) be bounded in $\mathbb{R}^n$. Then each coordinate sequence (x_{ki}) is bounded in $\mathbb{R}$.
2. By the Bolzano-Weierstrass theorem in $\mathbb{R}$, (x_{k1}) has a convergent subsequence indexed by $\mathbb{N}_1 \subset \mathbb{N}$.
3. The sequence $(x_{k2})_{k \in \mathbb{N}_1}$ has a convergent subsequence indexed by $\mathbb{N}_2 \subset \mathbb{N}_1$.
4. Continuing, we obtain $\mathbb{N} \supset \mathbb{N}_1 \supset \cdots \supset \mathbb{N}_n$ such that each coordinate converges.
5. Thus, the subsequence indexed by $\mathbb{N}_n$ converges coordinate-wise, hence in $\mathbb{R}^n$.

□

Theorem 3 (Cauchy Sequences in $\mathbb{R}^n$): Every Cauchy sequence in $\mathbb{R}^n$ is convergent.

Proof:

1. If (x_k) is Cauchy in $\mathbb{R}^n$, then each coordinate sequence (x_{ki}) is Cauchy in $\mathbb{R}$.
2. Since $\mathbb{R}$ is complete, each coordinate converges: $x_{ki} \to x_i$.
3. By Theorem 1, $x_k \to x = (x_1, \ldots, x_n)$.

□

6.2 Balls, Open and Closed Sets

6.2.1 Motivation

Balls are the fundamental building blocks of topology. An open ball is a set of points within a certain distance of a center. Open sets are unions of open balls; they capture the idea of a set where every point has some "wiggle room." Closed sets contain all their limit points. These concepts are essential for defining continuity and compactness.

Geometric Interpretation in $\mathbb{R}^2$: An open ball is the interior of a disk (excluding the boundary). An open set is a region where you can move a small distance in any direction and stay in the set. A closed set includes its boundary—like a closed disk.

6.2.2 Formal Definitions

Definition 2 (Open Ball): The **open ball** of radius $r > 0$ centered at $a \in \mathbb{R}^n$ is:

$$B_r(a) = \{x \in \mathbb{R}^n : \|x - a\| < r\} \quad (3)$$

Definition 3 (Closed Ball): The **closed ball** of radius $r > 0$ centered at $a \in \mathbb{R}^n$ is:

$$\overline{B}_r(a) = \{x \in \mathbb{R}^n : \|x - a\| \leq r\} \quad (4)$$

Definition 4 (Bounded Set): A set $X \subset \mathbb{R}^n$ is **bounded** if $\exists M > 0$ such that $\|x\| \leq M$ for all $x \in X$. Equivalently, $X \subset B_M(0)$.

6.2.3 Interior, Closure, and Boundary

Definition 5 (Interior Point): A point $a \in X$ is an **interior point** of X if:

$$\exists \epsilon > 0 \text{ such that } B_\epsilon(a) \subset X \quad (5)$$

Definition 6 (Closure Point): A point $a \in \mathbb{R}^n$ is a **closure point** of X if:

$$\forall \epsilon > 0, B_\epsilon(a) \cap X \neq \emptyset \quad (6)$$

Definition 7 (Interior and Closure): The **interior** of X is:

$$\text{int}(X) = \{x \in X : x \text{ is an interior point of } X\} \quad (7)$$

The **closure** of X is:

$$\overline{X} = \{x \in \mathbb{R}^n : x \text{ is a closure point of } X\} \quad (8)$$

Definition 8 (Open and Closed Sets): - A set $X \subset \mathbb{R}^n$ is **open** if $X = \text{int}(X)$. - A set $X \subset \mathbb{R}^n$ is **closed** if $X = \overline{X}$.

Definition 9 (Boundary Point): A point $a \in \mathbb{R}^n$ is a **boundary point** of X if:

$$\forall \epsilon > 0, B_\epsilon(a) \cap X \neq \emptyset \quad \text{and} \quad B_\epsilon(a) \cap X^c \neq \emptyset \quad (9)$$

6.2.4 Key Theorems

Theorem 4 (Closure Characterization): A point $a \in \mathbb{R}^n$ is in the closure of X iff there exists a sequence (x_k) in X with $x_k \to a$.

Proof:

1. ($\Rightarrow$) If $a \in \overline{X}$, then for each $k \in \mathbb{N}$, $B_{1/k}(a) \cap X \neq \emptyset$. Choose $x_k \in B_{1/k}(a) \cap X$. Then $\|x_k - a\| < 1/k \to 0$, so $x_k \to a$.
2. ($\Leftarrow$) If $x_k \to a$ with $x_k \in X$, then for any $\epsilon > 0$, there exists k such that $\|x_k - a\| < \epsilon$, so $x_k \in B_\epsilon(a) \cap X$. Hence, $B_\epsilon(a) \cap X \neq \emptyset$, so $a \in \overline{X}$.

□

Theorem 5 (Complement Characterization): A set $F \subset \mathbb{R}^n$ is closed iff its complement F^c is open.

Proof:

1. ($\Rightarrow$) Suppose F is closed. Let $x \in F^c$. Since $x \notin F = \overline{F}$, x is not a closure point of F. Hence, $\exists \epsilon > 0$ such that $B_\epsilon(x) \cap F = \emptyset$. Thus, $B_\epsilon(x) \subset F^c$, so F^c is open.
2. ($\Leftarrow$) Suppose F^c is open. Let $x \notin F$. Then $x \in F^c$, so $\exists \epsilon > 0$ such that $B_\epsilon(x) \subset F^c$. Thus, $B_\epsilon(x) \cap F = \emptyset$, so $x \notin \overline{F}$. Hence, $\overline{F} \subset F$, so F is closed.

□

Theorem 6 (Finite Unions and Intersections): If $F_1, \ldots, F_m$ are closed (open), then:

1. $\bigcup_{j=1}^m F_j$ is closed (open).
2. $\bigcap_{j=1}^m F_j$ is closed (open).

Proof for open sets:

1. For union: If $x \in \bigcup F_j$, then $x \in F_k$ for some k. Since F_k is open, $\exists \epsilon > 0$ such that $B_\epsilon(x) \subset F_k \subset \bigcup F_j$. Hence, the union is open.
2. For intersection: If $x \in \bigcap F_j$, then for each j, $\exists \epsilon_j > 0$ such that $B_{\epsilon_j}(x) \subset F_j$. Let $\epsilon = \min\{\epsilon_1, \ldots, \epsilon_m\}$. Then $B_\epsilon(x) \subset \bigcap F_j$. Hence, the intersection is open.
3. Closedness follows by taking complements and using De Morgan's laws.

□

6.3 Compactness

6.3.1 Motivation

Compactness is one of the most important concepts in analysis. It captures the idea of a set being "small" in a topological sense—every sequence has a convergent subsequence, and every open cover has a finite subcover. In $\mathbb{R}^n$, compact sets are exactly the closed and bounded sets (Heine-Borel Theorem).

Geometric Interpretation: A compact set in $\mathbb{R}^n$ is a set that is both closed (contains its boundary) and bounded (fits inside a large ball). It cannot have "holes" to infinity and cannot be open.

6.3.2 Formal Definitions

Definition 10 (Compactness in $\mathbb{R}^n$): A set $K \subset \mathbb{R}^n$ is **compact** if it is closed and bounded.

Definition 11 (Open Cover): A collection $\{G_\lambda\}_{\lambda \in L}$ of open sets is an **open cover** of K if:

$$K \subset \bigcup_{\lambda \in L} G_\lambda \quad (10)$$

Definition 12 (Finite Subcover): A **finite subcover** of an open cover is a finite subcollection $\{G_{\lambda_1}, \ldots, G_{\lambda_m}\}$ that still covers K.

6.3.3 Key Theorems

Theorem 7 (Heine-Borel Theorem): A set $K \subset \mathbb{R}^n$ is compact iff every open cover of K has a finite subcover.

Theorem 8 (Characterization of Compactness): For $K \subset \mathbb{R}^n$, the following are equivalent:

1. K is compact (closed and bounded).
2. Every sequence in K has a cluster point in K (i.e., a convergent subsequence).
3. Every sequence of nonempty closed subsets $F_1 \supset F_2 \supset \cdots$ of K has nonempty intersection:

$$\bigcap_{k=1}^{\infty} F_k \neq \emptyset$$

4. Every open cover of K has a finite subcover.

Proof of (i) $\Rightarrow$ (ii):

1. Let (x_k) be a sequence in K. Since K is bounded, (x_k) is bounded.
2. By Bolzano-Weierstrass in $\mathbb{R}^n$, (x_k) has a convergent subsequence $x_{k_j} \to x \in \mathbb{R}^n$.
3. Since K is closed, it contains all its limit points, so $x \in K$.

□

Proof of (ii) $\Rightarrow$ (iii):

1. Let (F_k) be a sequence of nonempty closed subsets with $F_1 \supset F_2 \supset \cdots$.
2. Choose $x_k \in F_k$ for each k. Then (x_k) is a sequence in K.
3. By (ii), there exists a subsequence $x_{k_j} \to x \in K$.
4. Fix $m \in \mathbb{N}$. For sufficiently large j, $k_j \geq m$, so $x_{k_j} \in F_{k_j} \subset F_m$.
5. Thus, $x \in \overline{F_m} = F_m$ since F_m is closed.
6. Therefore, $x \in \bigcap_{m=1}^{\infty} F_m \neq \emptyset$.

□

Proof of (iii) $\Rightarrow$ (iv):

1. Let (G_j) be a sequence of open sets covering K.
2. Define $F_k = K \setminus \bigcup_{j=1}^{k} G_j$. Each F_k is closed in K, and $F_1 \supset F_2 \supset \cdots$.
3. If all F_k were nonempty, then by (iii), $\bigcap_{k=1}^{\infty} F_k \neq \emptyset$.
4. But $\bigcap_{k=1}^{\infty} F_k = K \setminus \bigcup_{j=1}^{\infty} G_j = \emptyset$, since the G_j's cover K.
5. Contradiction. Hence, some $F_k = \emptyset$, so $K \subset \bigcup_{j=1}^{k} G_j$.

□

Proof of (iv) $\Rightarrow$ (i):

1. **Bounded:** $K \subset \mathbb{R}^n = \bigcup_{j=1}^{\infty} B_j(0)$. By (iv), there is a finite subcover: $K \subset \bigcup_{j=1}^{m} B_j(0) = B_m(0)$. Hence, K is bounded.
2. **Closed:** Suppose $a \in \overline{K} \setminus K$. Then $K \subset \bigcup_{j=1}^{\infty} \left(\overline{B_{1/j}(a)}\right)^c$. By (iv), there is a finite subcover: $K \subset \bigcup_{j=1}^{m} \left(\overline{B_{1/j}(a)}\right)^c$. Then $B_{1/m}(a) \cap K = \emptyset$, contradicting that $a \in \overline{K}$. Hence, K is closed.

□

6.4 Lindelöf's Theorem

6.4.1 Motivation

Lindelöf's Theorem states that every open cover of a subset of $\mathbb{R}^n$ has a countable subcover. This is a crucial technical result: it allows us to reduce arbitrary open covers (which may be uncountable) to countable ones, which are easier to handle. It follows from the fact that $\mathbb{R}^n$ has a countable basis of open sets.

6.4.2 Key Theorems

Theorem 9 (Lindelöf's Theorem): There exists a countable family $\mathcal{B}$ of open sets in $\mathbb{R}^n$ such that every open set G is a union of sets from $\mathcal{B}$:

$$G = \bigcup_{B \in \mathcal{B}} B \quad (11)$$

Proof:

1. Let $\mathcal{B} = \left\{B_{1/j}(q) : j \in \mathbb{N}, q \in \mathbb{Q}^n\right\}$. This is countable since $\mathbb{N} \times \mathbb{Q}^n$ is countable.
2. Let $G \subset \mathbb{R}^n$ be open. Pick $a \in G$. Since G is open, $\exists \epsilon > 0$ such that $B_\epsilon(a) \subset G$.
3. Choose $j \in \mathbb{N}$ such that $1/j < \epsilon/2$. Since $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$, $\exists q \in \mathbb{Q}^n \cap B_{1/j}(a)$.
4. Then $a \in B_{1/j}(q) \in \mathcal{B}$. For any $x \in B_{1/j}(q)$:
$$\|x - a\| \leq \|x - q\| + \|q - a\| < \frac{1}{j} + \frac{1}{j} < \epsilon$$
5. Hence, $x \in B_\epsilon(a) \subset G$, so $B_{1/j}(q) \subset G$.
6. Therefore, $G = \bigcup_{B \in \mathcal{B}} B$.

□

Corollary: Every set in $\mathbb{R}^n$ has a countable dense subset.

Proof:

1. Let $X \subset \mathbb{R}^n$. Let $\mathcal{B}$ be the countable family from Lindelöf's Theorem.
2. Define $\mathcal{D} = \{B \in \mathcal{B} : B \cap X \neq \emptyset\}$, which is countable.
3. For each $B_j \in \mathcal{D}$, choose $y_j \in B_j \cap X$.
4. Let $Y = \{y_j : j \in \mathbb{N}\}$. Then Y is countable and dense in X.

□

6.5 Summary of Key Results

Concept	Definition	Key Formula/Result
Convergence in $\mathbb{R}^n$	$\|x_k - x\| \to 0$	Coordinate-wise convergence
Open Ball	$B_r(a) = \{x : \|x - a\| < r\}$	Basis of topology
Closed Ball	$\overline{B}_r(a) = \{x : \|x - a\| \leq r\}$	Closed set
Interior	$\text{int}(X) = \{x : \exists \epsilon > 0, B_\epsilon(x) \subset X\}$	Largest open subset
Closure	$\overline{X} = \{x : \forall \epsilon > 0, B_\epsilon(x) \cap X \neq \emptyset\}$	Smallest closed superset
Open Set	$X = \text{int}(X)$	Complement is closed

Concept	Definition	Key Formula/Result
Closed Set	$X = \overline{X}$	Complement is open
Bolzano-Weierstrass	Every bounded sequence has convergent subsequence	In $\mathbb{R}^n$
Compactness	Closed and bounded	Equivalent to finite subcover property
Heine-Borel	Compact $\iff$ every open cover has finite subcover	In $\mathbb{R}^n$
Lindelöf's Theorem	Every open cover has countable subcover	$\mathbb{R}^n$ is second-countable

---

Chapter 7: Limits and Continuity

7.1 Limits of Functions

7.1.1 Motivation

The problem that limits solve is the **analysis of function behavior near a point**. Even if a function is not defined at a point, we may still want to know what value it approaches as the input gets arbitrarily close. Limits are the foundation of continuity, differentiability, and integration—the core concepts of calculus and analysis.

Geometric Interpretation in $\mathbb{R}^2$: Visualize the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$ as a surface. The limit $\lim_{x\to a} f(x) = b$ means that as points x approach a from any direction, the heights $f(x)$ approach b. The limit exists only if the height is the same regardless of the path taken to a.

7.1.2 Formal Definitions

Definition 1 (Accumulation Point): A point $a \in \mathbb{R}^n$ is an **accumulation point** of $X \subset \mathbb{R}^n$ if:

$$\forall \epsilon > 0, B_\epsilon(a) \cap (X \setminus \{a\}) \neq \emptyset \quad (1)$$

Definition 2 (Isolated Point): A point $a \in X$ is an **isolated point** of X if:

$$\exists \epsilon > 0 \text{ such that } B_\epsilon(a) \cap X = \{a\} \quad (2)$$

Definition 3 (Limit of a Function): Let $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$, $a \in X'$ (an accumulation point of X), and $b \in \mathbb{R}^m$. We say that:

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

if:

$$\forall \epsilon > 0, \exists \delta > 0 \text{ such that } \forall x \in X, 0 < \|x - a\| < \delta \implies \|f(x) - b\| < \epsilon \quad (3)$$

7.1.3 Key Theorems

Theorem 1 (Sequential Characterization of Limits): $\lim_{x\to a} f(x) = b$ iff for every sequence (x_k) in $X \setminus \{a\}$ with $x_k \to a$, one has $f(x_k) \to b$.

Proof:

1. ($\Rightarrow$) Suppose $\lim_{x\to a} f(x) = b$. Let (x_k) be a sequence with $x_k \to a$ and $x_k \neq a$. Given $\epsilon > 0$, choose $\delta > 0$ such that $0 < \|x - a\| < \delta \implies \|f(x) - b\| < \epsilon$. Since $x_k \to a$, $\exists k_0$ such that $k \geq k_0 \implies \|x_k - a\| < \delta$. Then $\|f(x_k) - b\| < \epsilon$, so $f(x_k) \to b$.

2. ($\Leftarrow$) Suppose the sequential condition holds. If $\lim_{x\to a} f(x) \neq b$, then $\exists\epsilon > 0$ such that for every $\delta > 0$, there exists $x \in X$ with $0 < \|x - a\| < \delta$ but $\|f(x) - b\| \geq \epsilon$. Choose x_k with $\|x_k - a\| < 1/k$ and $\|f(x_k) - b\| \geq \epsilon$. Then $x_k \to a$ but $f(x_k) \not\to b$, contradicting the condition.

□

Theorem 2 (One-Sided Limits in $\mathbb{R}$): For a function $f : X \subset \mathbb{R} \to \mathbb{R}$, the limit $\lim_{x\to a} f(x)$ exists iff both one-sided limits exist and are equal:

$$\lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x) \quad (4)$$

7.2 Continuity

7.2.1 Motivation

Continuity formalizes the intuitive idea that a function has no "jumps" or "breaks." A continuous function maps nearby points to nearby values. The ϵ-δ definition captures this precisely: small changes in the input produce small changes in the output.

Geometric Interpretation: The graph of a continuous function can be drawn without lifting the pen from the paper. In $\mathbb{R}^2$, a continuous surface has no holes or tears.

7.2.2 Formal Definitions

Definition 4 (Continuity at a Point): A function $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is **continuous at** $a \in X$ if:

$$\forall\epsilon > 0, \exists\delta > 0 \text{ such that } \forall x \in X, \|x - a\| < \delta \implies \|f(x) - f(a)\| < \epsilon \quad (5)$$

Definition 5 (Continuity on a Set): A function f is **continuous** on X if it is continuous at every point $a \in X$.

7.2.3 Key Theorems

Theorem 3 (Equivalent Characterizations of Continuity): For $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ and $a \in X$, the following are equivalent:

1. f is continuous at a.
2. $\lim_{x\to a} f(x) = f(a)$ (if a is an accumulation point).
3. For every sequence (x_k) in X with $x_k \to a$, one has $f(x_k) \to f(a)$.

Proof: Similar to the sequential characterization of limits.

□

Theorem 4 (Topological Characterization of Continuity): For $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$, the following are equivalent:

1. f is continuous.
2. For every open set $G \subset \mathbb{R}^m$, the preimage $f^{-1}(G)$ is open in X.
3. For every closed set $F \subset \mathbb{R}^m$, the preimage $f^{-1}(F)$ is closed in X.

Proof: (Standard; follows from the ϵ-δ definition and the definition of open/closed sets.)

□

7.3 Continuous Functions on Compact Sets

7.3.1 Motivation

Continuous functions on compact sets have several powerful properties: they attain their maximum and minimum (Extreme Value Theorem), they are uniformly continuous, and their image is compact. These results are fundamental in optimization and analysis.

Geometric Interpretation: The image of a compact set under a continuous function is also compact—it is closed and bounded. The function must reach its highest and lowest points somewhere on the compact set.

7.3.2 Key Theorems

Theorem 5 (Continuous Image of Compact is Compact): If $K \subset \mathbb{R}^n$ is compact and $f : K \to \mathbb{R}^m$ is continuous, then $f(K)$ is compact.

Proof:

1. Let (y_k) be a sequence in $f(K)$. Then $y_k = f(x_k)$ for some $x_k \in K$.
2. Since K is compact, there exists a subsequence $x_{k_j} \to x \in K$.
3. By continuity, $y_{k_j} = f(x_{k_j}) \to f(x) \in f(K)$.
4. Thus, every sequence in $f(K)$ has a convergent subsequence in $f(K)$, so $f(K)$ is compact.

□

Theorem 6 (Extreme Value Theorem): If $K \subset \mathbb{R}^n$ is compact and $f : K \to \mathbb{R}$ is continuous, then f attains its maximum and minimum on K:

$$\exists x_{\max}, x_{\min} \in K \text{ such that } f(x_{\max}) = \sup f(K), \quad f(x_{\min}) = \inf f(K) \quad (6)$$

Proof:

1. By Theorem 5, $f(K)$ is compact in $\mathbb{R}$, hence closed and bounded.
2. Since $f(K)$ is closed, it contains all its limit points, so $\sup f(K) \in f(K)$ and $\inf f(K) \in f(K)$.
3. Therefore, there exist $x_{\max}, x_{\min} \in K$ such that $f(x_{\max}) = \sup f(K)$ and $f(x_{\min}) = \inf f(K)$.

□

Theorem 7 (Uniform Continuity): If $K \subset \mathbb{R}^n$ is compact and $f : K \to \mathbb{R}^m$ is continuous, then f is uniformly continuous on K.

Proof:

1. Suppose f is not uniformly continuous. Then $\exists \epsilon > 0$ and sequences $(x_k), (y_k)$ in K such that $\|x_k - y_k\| \to 0$ but $\|f(x_k) - f(y_k)\| \geq \epsilon$.
2. By compactness, there exist subsequences $x_{k_j} \to x$ and $y_{k_j} \to y$ with $x, y \in K$.
3. Since $\|x_k - y_k\| \to 0$, we have $x = y$.
4. By continuity, $f(x_{k_j}) \to f(x)$ and $f(y_{k_j}) \to f(x)$, so $\|f(x_{k_j}) - f(y_{k_j})\| \to 0$, contradicting $\|f(x_k) - f(y_k)\| \geq \epsilon$.
5. Therefore, f is uniformly continuous.

□

7.4 Connectedness and the Intermediate Value Theorem

7.4.1 Motivation

Connectedness captures the idea of a set being "in one piece." A continuous image of a connected set is connected. In $\mathbb{R}$, connected sets are exactly intervals. The Intermediate Value Theorem follows: a continuous function on a connected set takes all values between any two of its values.

Geometric Interpretation: A connected set cannot be split into two nonempty disjoint open subsets. Think of it as a set with no gaps—a single continuous piece.

7.4.2 Formal Definitions

Definition 6 (Connected Set): A set $C \subset \mathbb{R}^n$ is **connected** if, whenever $U, V \subset C$ are disjoint open sets in C with $U \cup V = C$, one has $U = \emptyset$ or $V = \emptyset$.

7.4.3 Key Theorems

Theorem 8 (Continuous Image of Connected is Connected): If $C \subset \mathbb{R}^n$ is connected and $f : C \to \mathbb{R}^m$ is continuous, then $f(C)$ is connected.

Proof:

1. Let U, V be disjoint open sets in $f(C)$ with $U \cup V = f(C)$.
2. Then $f^{-1}(U)$ and $f^{-1}(V)$ are open in C, disjoint, and $C = f^{-1}(U) \cup f^{-1}(V)$.
3. Since C is connected, either $f^{-1}(U) = \emptyset$ or $f^{-1}(V) = \emptyset$.
4. Hence, either $U = \emptyset$ or $V = \emptyset$, so $f(C)$ is connected.

□

Theorem 9 (Characterization of Connected Sets in $\mathbb{R}$): A set $C \subset \mathbb{R}$ is connected iff it is an interval.

Theorem 10 (Intermediate Value Theorem - IVT): If $C \subset \mathbb{R}^n$ is connected and $f : C \to \mathbb{R}$ is continuous, then for any $a, b \in C$ and any $\gamma \in \mathbb{R}$ with $f(a) \leq \gamma \leq f(b)$, there exists $x \in C$ such that $f(x) = \gamma$.

Proof:

1. By Theorem 8, $f(C)$ is connected in $\mathbb{R}$.
2. By Theorem 9, $f(C)$ is an interval in $\mathbb{R}$.
3. Since $f(a), f(b) \in f(C)$ and $f(a) \leq \gamma \leq f(b)$, the interval $f(C)$ contains all points between $f(a)$ and $f(b)$.
4. Hence, $\gamma \in f(C)$, so there exists $x \in C$ with $f(x) = \gamma$.

□

Theorem 11 (Boundary Crossing Theorem): Let $X, C \subset \mathbb{R}^n$ with C connected. If $X \cap C \neq \emptyset$ and $X^c \cap C \neq \emptyset$, then $\partial X \cap C \neq \emptyset$.

Proof:

1. Suppose $\partial X \cap C = \emptyset$. Then $C = (C \cap \text{int} X) \cup (C \cap \text{int} X^c)$.
2. These are disjoint open sets in C, and both are nonempty by hypothesis.

3. This contradicts the connectedness of C.
4. Therefore, $\partial X \cap C \neq \emptyset$.

□

7.5 Summary of Key Results

Concept	Definition	Key Formula/Result
Accumulation Point	Every neighborhood intersects $X \setminus \{a\}$	$a \in X'$
Limit	$f(x) \to b$ as $x \to a$	$\forall \epsilon > 0, \exists \delta > 0$
Sequential Limit	$x_k \to a \implies f(x_k) \to b$	Equivalent to ϵ-δ
Continuity at a Point	$x \to a \implies f(x) \to f(a)$	ϵ-δ definition
Topological Continuity	Preimage of open sets is open	Equivalent to ϵ-δ
Continuous Image of Compact	$f(K)$ is compact	Extreme Value Theorem follows
Extreme Value Theorem	Continuous on compact attains max/min	$\exists x_{\max}, x_{\min} \in K$
Uniform Continuity	δ independent of point	Continuous on compact is uniformly continuous
Connected Set	Cannot be split into disjoint open sets	In $\mathbb{R}$, connected $\iff$ interval
Intermediate Value Theorem	Continuous on connected takes all intermediate values	$f(a) \leq \gamma \leq f(b) \implies \exists x : f(x) = \gamma$

Chapter 8: Differentiability - Part I

8.1 Differentiability in $\mathbb{R}$

8.1.1 Motivation

The problem that differentiability solves is the **local linear approximation of functions**. While continuity tells us that a function has no jumps, differentiability tells us that the function can be approximated by a straight line (its tangent) near a point. This linear approximation is the foundation of optimization, numerical methods, and the study of rates of change in economics and physics.

Geometric Interpretation in $\mathbb{R}^2$: Visualize the graph of a function $f : \mathbb{R} \to \mathbb{R}$ as a curve. If f is differentiable at a, the curve has a well-defined tangent line at a. The slope of this tangent line is $f'(a)$. Zooming in near a, the curve becomes indistinguishable from its tangent line.

8.1.2 Formal Definitions

Definition 1 (Derivative in $\mathbb{R}$): A function $f : I \subset \mathbb{R} \to \mathbb{R}$, where I is an interval, is **differentiable** at $a \in I$ if the following limit exists:

$$\boxed{f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}} \quad (1)$$

This limit is called the **derivative** of f at a.

8.1.3 Key Theorems

Theorem 1 (Differentiability Implies Continuity): If f is differentiable at a, then f is continuous at a.

Proof:

1. Since f is differentiable at a:
$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$
2. Let $h(x) = x - a$. Then:
$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$
3. Multiplying both sides by $\lim_{x \to a}(x - a)$:
$$f'(a) \cdot 0 = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \cdot \lim_{x \to a}(x - a)$$
4. By the product rule for limits:
$$0 = \lim_{x \to a}[f(x) - f(a)]$$
5. Therefore: $\lim_{x \to a} f(x) = f(a)$, so f is continuous at a.

□

Theorem 2 (Local Growth): If $f : I \subset \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in I$:

1. If $f'(a) > 0$, then $\exists \epsilon > 0$ such that $f(a+h) > f(a)$ for all $0 < h < \epsilon$.
2. If $f'(a) < 0$, then $\exists \epsilon > 0$ such that $f(a+h) < f(a)$ for all $0 > h > -\epsilon$.

Theorem 3 (Chain Rule in $\mathbb{R}$): If $f : I \subset \mathbb{R} \to \mathbb{R}$ is differentiable at $a \in I$, and $g : J \subset \mathbb{R} \to \mathbb{R}$ is differentiable at $b = f(a)$, with $f(I) \subset J$, then $g \circ f$ is differentiable at a and:

$$\boxed{(g \circ f)'(a) = g'(f(a)) \cdot f'(a)} \quad (2)$$

Proof:

1. Consider the quotient:
$$\frac{g(f(a+h)) - g(f(a))}{h} = \frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)} \cdot \frac{f(a+h) - f(a)}{h}$$
(whenever $f(a+h) - f(a) \neq 0$; when the latter is zero, the quotient is also zero).
2. As $h \to 0$:
   - $\frac{g(f(a+h)) - g(f(a))}{f(a+h) - f(a)} \to g'(f(a))$
   - $\frac{f(a+h) - f(a)}{h} \to f'(a)$
3. Therefore: $(g \circ f)'(a) = g'(f(a)) \cdot f'(a)$.

□

8.2 Rolle's Theorem and the Mean Value Theorem

8.2.1 Motivation

Rolle's Theorem and the Mean Value Theorem (MVT) are among the most important results in calculus. They provide a bridge between the derivative of a function and its values. The MVT states that for a differentiable function on an interval, there exists a point where the instantaneous rate of change equals the average rate of change. This result is the foundation for understanding monotonicity, optimization, and Taylor's theorem.

Geometric Interpretation: The MVT says that if you travel from point A to point B on a smooth curve, there must be some point where the tangent line is parallel to the secant line connecting A and B.

8.2.2 Formal Statements

Theorem 4 (Rolle's Theorem): Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there exists $c \in (a, b)$ such that:

$$\boxed{f'(c) = 0} \quad (3)$$

Proof:

1. If f is constant on $[a, b]$, then $f'(c) = 0$ for all $c \in (a, b)$.
2. Otherwise, by the Extreme Value Theorem, f attains a maximum or a minimum at some $c \in [a, b]$.
3. Since $f(a) = f(b)$, the extremum cannot occur at the endpoints unless the function is constant.
4. Thus, $c \in (a, b)$. By the local growth theorem, $f'(c) = 0$.

□

Theorem 5 (Mean Value Theorem): Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that:

$$\boxed{f'(c) = \frac{f(b) - f(a)}{b - a}} \quad (4)$$

Proof:

1. Define $g(x) = f(x) - \left[f(a) + \frac{f(b)-f(a)}{b-a}(x - a)\right]$.
2. Then $g(a) = 0$ and $g(b) = 0$, so by Rolle's Theorem, $\exists c \in (a, b)$ such that $g'(c) = 0$.
3. But $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$.
4. Therefore: $f'(c) = \frac{f(b)-f(a)}{b-a}$.

□

Theorem 6 (Darboux's Theorem): Let $f : [a, b] \to \mathbb{R}$ be differentiable, and suppose $f'(a) < d < f'(b)$. Then there exists $c \in (a, b)$ such that $f'(c) = d$.

Proof:

1. Apply the previous result to $g(x) = f(x) - dx$. Then $g'(a) < 0 < g'(b)$.
2. Since g is continuous on $[a, b]$, it attains a minimum at some $c \in [a, b]$.
3. Since $g'(a) < 0 < g'(b)$, $c \in (a, b)$.
4. By the local growth theorem, $g'(c) = 0$, so $f'(c) = d$.

□

8.3 Partial and Directional Derivatives

8.3.1 Motivation

In $\mathbb{R}^n$, we need to understand how functions change in different directions. Partial derivatives measure the rate of change along coordinate axes, while directional derivatives measure the rate of change along arbitrary directions. These concepts are essential for multivariable optimization, gradient descent, and the study of economic functions with multiple inputs.

Geometric Interpretation: Visualize a function $f : \mathbb{R}^2 \to \mathbb{R}$ as a surface. The partial derivative $\partial_1 f(a)$ is the slope of the curve obtained by slicing the surface with a vertical plane parallel to the x-axis. The directional derivative $\partial_v f(a)$ is the slope in the direction of the vector v.

8.3.2 Formal Definitions

Definition 2 (Partial Derivative): Let $U \subset \mathbb{R}^n$ be open, $a \in U$, and $f : U \to \mathbb{R}$. For $j = 1, \ldots, n$, the j-th **partial derivative** of f at a is:

$$\boxed{\partial_j f(a) = \lim_{t \to 0} \frac{f(a + te_j) - f(a)}{t}} \tag{5}$$

where e_j is the j-th standard basis vector.

Definition 3 (Directional Derivative): For $v \in \mathbb{R}^n \setminus \{0\}$, the **directional derivative** of f at a in the direction v is:

$$\boxed{\partial_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}} \tag{6}$$

8.3.3 Key Results

Result 1: Existence of partial derivatives does **not** imply continuity.

Example: Consider:

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2}, & (x, y) \neq (0, 0) \\ 0, & (x, y) = (0, 0) \end{cases}$$

At $(0, 0)$: - $\partial_1 f(0, 0) = \lim_{t \to 0} \frac{f(t,0)-f(0,0)}{t} = 0$ - $\partial_2 f(0, 0) = \lim_{t \to 0} \frac{f(0,t)-f(0,0)}{t} = 0$

But f is not continuous at $(0, 0)$, since $f(x, x) = 1/2$ for $x \neq 0$.

Result 2: Existence of partial derivatives does **not** imply existence of all directional derivatives.

Using the same function as above, for $v = (\alpha, \beta)$:

$$\frac{f(t\alpha, t\beta)}{t} = \frac{1}{t} \cdot \frac{\alpha\beta}{\alpha^2 + \beta^2}$$

This limit does not exist at the origin if $\alpha\beta \neq 0$.

Result 3: Existence of directional derivatives does **not** imply linearity of $v \mapsto \partial_v f(a)$.

The mapping $v \mapsto \partial_v f(a)$ may not satisfy additivity or scalar multiplication. That is, the directional derivative is not necessarily a linear function of the direction.

8.3.4 Key Theorems

Theorem 7 (Mean Value Theorem in $\mathbb{R}^n$): Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ be differentiable on an open set U, and suppose the line segment from a to $a + v$ is contained in U. Then there exists $\theta \in (0, 1)$ such that:

$$\boxed{f(a + v) - f(a) = \partial_v f(a + \theta v)} \quad (7)$$

Proof:

1. Define $\xi(t) = f(a + tv)$ for $t \in [0, 1]$.
2. Then ξ is a single-variable differentiable function.
3. By the Mean Value Theorem in $\mathbb{R}$, there exists $\theta \in (0, 1)$ such that:
$$\xi(1) - \xi(0) = \xi'(\theta)$$
4. But $\xi'(t) = \partial_v f(a + tv)$, so:
$$f(a + v) - f(a) = \partial_v f(a + \theta v)$$

□

Corollary: Let $U \subset \mathbb{R}^n$ be open and connected. If f has directional derivatives at every point in U and every direction, and $\partial_v f(x) = 0$ for all $x \in U$ and all $v \in \mathbb{R}^n$, then f is constant on U.

8.4 Summary of Key Results

Concept	Definition	Key Formula
Derivative in $\mathbb{R}$	Limit of difference quotient	$f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}$
Differentiability $\implies$ Continuity	Differentiable at a implies continuous at a	$\lim_{x\to a} f(x) = f(a)$
Chain Rule	Derivative of composition	$(g \circ f)'(a) = g'(f(a)) \cdot f'(a)$
Rolle's Theorem	Equal endpoints $\implies$ zero derivative	$f(a) = f(b) \implies \exists c : f'(c) = 0$
Mean Value Theorem	Average rate equals instantaneous rate	$f'(c) = \frac{f(b)-f(a)}{b-a}$
Darboux's Theorem	Derivatives have intermediate value property	$f'(a) < d < f'(b) \implies \exists c : f'(c) = d$
Partial Derivative	Derivative along coordinate axis	$\partial_j f(a) = \lim_{t\to 0} \frac{f(a+te_j)-f(a)}{t}$
Directional Derivative	Derivative along arbitrary direction	$\partial_v f(a) = \lim_{t\to 0} \frac{f(a+tv)-f(a)}{t}$

Chapter 9: Linear Maps

9.1 Definition and Basic Properties

9.1.1 Motivation

The problem that linear maps solve is the **study of structure-preserving transformations** between vector spaces. Linear maps are the "homomorphisms" of vector spaces: they preserve addition and scalar multiplication. They are the mathematical framework for representing systems of linear equations, transformations of coordinates, and the derivative in multivariable calculus.

Geometric Interpretation: Visualize a linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ as a transformation of the plane. Lines are mapped to lines, and the origin is fixed. Linear maps can rotate, scale, shear, or project vectors.

9.1.2 Formal Definitions

Definition 1 (Linear Map): A function $T : V \to W$ between vector spaces over $\mathbb{R}$ is **linear** if: 1. $T(v_1 + v_2) = T(v_1) + T(v_2)$ for all $v_1, v_2 \in V$. 2. $T(\alpha v) = \alpha T(v)$ for all $\alpha \in \mathbb{R}$ and $v \in V$.

Definition 2 (Matrix of a Linear Map): Given a basis $\{e_1, \dots, e_m\}$ of V and $\{f_1, \dots, f_n\}$ of W, the matrix of a linear map $T : V \to W$ is the $n \times m$ matrix (α_{ij}) where:

$$T(e_j) = \sum_{i=1}^{n} \alpha_{ij} f_i \quad (1)$$

9.1.3 Key Theorems

Theorem 1 (Properties of Linear Maps): If $T : V \to W$ is linear, then: 1. $T(0) = 0$. 2. $T\left(\sum_{i=1}^{m} \alpha_i v_i\right) = \sum_{i=1}^{m} \alpha_i T(v_i)$.

Proof:

1. $T(0) = T(0 + 0) = T(0) + T(0)$. Thus, $T(0) = 0$.
2. Follows by induction from the definition.

□

Theorem 2 (Matrix Representation): For $v = \sum_{j=1}^{m} v_j e_j$, the coordinate vector of $T(v)$ is:

$$\begin{pmatrix} \sum_{j=1}^{m} \alpha_{1j} v_j \\ \vdots \\ \sum_{j=1}^{m} \alpha_{nj} v_j \end{pmatrix} \quad (2)$$

Proof:

1. $T(v) = \sum_{j=1}^{m} v_j T(e_j) = \sum_{j=1}^{m} v_j \sum_{i=1}^{n} \alpha_{ij} f_i$.
2. Thus: $T(v) = \sum_{i=1}^{n} \left(\sum_{j=1}^{m} \alpha_{ij} v_j \right) f_i$.
3. Therefore, the coordinate vector of $T(v)$ is as claimed.

□

9.2 Kernel and Image

9.2.1 Motivation

The kernel and image of a linear map capture two fundamental aspects of the transformation: what gets "crushed" to zero (the kernel) and what is "hit" by the transformation (the image). The dimension theorem (Rank-Nullity Theorem) relates these two quantities to the dimension of the domain.

Geometric Interpretation: For a linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$, the kernel is the set of vectors that map to the origin—these are vectors that are "lost" by the transformation. The image is the set of all possible outputs—the "reach" of the transformation.

9.2.2 Formal Definitions

Definition 3 (Kernel): The **kernel** (or null space) of a linear map $T : V \to W$ is:

$$\ker T = T^{-1}(\{0\}) = \{v \in V : T(v) = 0\} \quad (3)$$

Definition 4 (Image): The **image** (or range) of a linear map $T : V \to W$ is:

$$\operatorname{Im} T = T(V) = \{T(v) : v \in V\} \quad (4)$$

9.2.3 Key Theorems

Theorem 3 (Kernel and Image are Subspaces):

1. $\ker T$ is a subspace of V.
2. $\operatorname{Im} T$ is a subspace of W.

Proof: (Exercise)

Theorem 4 (Injectivity Characterization): For a linear map $T : V \to W$, the following are equivalent: 1. T is injective. 2. $\ker T = \{0\}$. 3. For every linearly independent set $L \subset V$, $T(L)$ is linearly independent.

Proof:

1. (1) $\Rightarrow$ (2): If $T(v) = 0 = T(0)$, then injectivity gives $v = 0$. Thus, $\ker T = \{0\}$.
2. (2) $\Rightarrow$ (3): Let $\{e_1, \dots, e_k\}$ be linearly independent. Suppose $\sum \alpha_i T(e_i) = 0$. Then $T(\sum \alpha_i e_i) = 0$, so $\sum \alpha_i e_i \in \ker T = \{0\}$. Hence, $\sum \alpha_i e_i = 0$. By linear independence, all $\alpha_i = 0$.
3. (3) $\Rightarrow$ (1): If $T(u) = T(v)$, then $T(u - v) = 0$. If $u - v \neq 0$, then $\{T(u - v)\}$ is not linearly independent, contradicting (3). Hence, $u = v$.

□

Theorem 5 (Surjectivity Characterization): $T : V \to W$ is surjective iff for every spanning set $S \subset V$, $T(S)$ spans W.

Corollary: $T : V \to W$ is bijective iff every basis of V is mapped to a basis of W.

9.3 Direct Sums and Projections

9.3.1 Motivation

Direct sums allow us to decompose a vector space into complementary subspaces. This decomposition is fundamental for understanding the structure of linear maps, projections, and the relationship between the kernel and image.

Geometric Interpretation: In $\mathbb{R}^2$, any two non-parallel lines through the origin form a direct sum decomposition of the plane: every vector can be uniquely written as the sum of a vector on one line and a vector on the other.

9.3.2 Formal Definitions

Definition 5 (Direct Sum): Let $U, V \subset E$ be subspaces with $U \cap V = \{0\}$. The **direct sum** of U and V is:

$$U \oplus V = \{u + v : u \in U, v \in V\} \quad (5)$$

Definition 6 (Projection): Given a direct sum decomposition $E = U \oplus V$, the **projection** $P_U : E \to U$ maps each $w \in E$ to the unique $u \in U$ such that $w - u \in V$.

9.3.3 Key Theorems

Theorem 6 (Unique Decomposition): If $E = U \oplus V$, then for every $w \in E$, there exist unique $u \in U$ and $v \in V$ such that $w = u + v$.

Proof: (Exercise)

Theorem 7 (Properties of Projections):

1. P_U is linear and idempotent: $P_U^2 = P_U$.
2. If $P : E \to U$ is onto and idempotent, then P is the projection onto U relative to $E = U \oplus \ker P$.

Theorem 8 (Dimension of Direct Sum): If $\{e_1, \ldots, e_n\}$ is a basis of E, then:

$$E = \operatorname{span}\{e_1, \ldots, e_k\} \oplus \operatorname{span}\{e_{k+1}, \ldots, e_n\} \quad (6)$$

Conversely, if $E = U \oplus V$ and $\{u_1, \ldots, u_n\}$, $\{v_1, \ldots, v_m\}$ are bases of U and V, then $\{u_1, \ldots, u_n, v_1, \ldots, v_m\}$ is a basis of E.

In particular:

$$\boxed{\dim E = \dim U + \dim V} \quad (7)$$

9.4 Rank-Nullity Theorem

9.4.1 Motivation

The Rank-Nullity Theorem is one of the most important results in linear algebra. It states that the dimension of the domain equals the sum of the dimensions of the kernel and the image. This theorem is essential for understanding the structure of linear systems and the behavior of linear transformations.

Geometric Interpretation: The kernel measures the "loss" of dimension caused by the transformation. The image measures the "reach" of the transformation. The theorem says that the dimension of the domain is exactly the sum of what is lost and what is reached.

9.4.2 Formal Statement

Theorem 9 (Rank-Nullity Theorem): Let $T : V \to W$ be a linear map. Then:

$$\boxed{\dim V = \dim(\ker T) + \dim(\operatorname{Im} T)} \quad (8)$$

Proof:

1. Since $\ker T \subset V$, by the direct sum theorem, there exists a subspace $F \subset V$ such that $V = \ker T \oplus F$.
2. Claim: $T|_F : F \to \operatorname{Im} T$ is a bijection.
3. **Injectivity:** If $v \in F$ and $T(v) = 0$, then $v \in \ker T \cap F = \{0\}$, so $v = 0$.
4. **Surjectivity:** For any $w \in \operatorname{Im} T$, $w = T(v)$ for some $v \in V$. Write $v = e + f$ with $e \in \ker T$, $f \in F$. Then $w = T(v) = T(e) + T(f) = T(f)$. Thus, $w \in \operatorname{Im}(T|_F)$.
5. Therefore, $T|_F : F \to \operatorname{Im} T$ is a bijection, so $\dim F = \dim(\operatorname{Im} T)$.
6. Since $V = \ker T \oplus F$, $\dim V = \dim(\ker T) + \dim F = \dim(\ker T) + \dim(\operatorname{Im} T)$.

□

9.5 Adjoint Linear Maps

9.5.1 Motivation

The adjoint of a linear map is the generalization of the transpose of a matrix to inner product spaces. It allows us to "reverse" a linear map in the sense of inner products. The adjoint is fundamental for understanding orthogonal projections, least squares, and the Fundamental Theorem of Linear Algebra.

Geometric Interpretation: For a linear map $T : V \to W$, the adjoint $T^* : W \to V$ is the unique map such that the inner product of w with $T(v)$ equals the inner product of $T^*(w)$ with v.

9.5.2 Formal Definitions

Definition 7 (Adjoint): Let V, W be vector spaces with inner products. The **adjoint** of a linear map $T : V \to W$ is the unique linear map $T^* : W \to V$ such that:

$$\boxed{\langle T^*(w), v\rangle_V = \langle w, T(v)\rangle_W} \quad (9)$$

for all $v \in V$ and $w \in W$.

Definition 8 (Orthogonal Complement): For a set $A \subset V$, the **orthogonal complement** is:

$$A^\perp = \{v \in V : \langle v, a\rangle = 0 \text{ for all } a \in A\} \quad (10)$$

9.5.3 Key Theorems

Theorem 10 (Properties of Orthogonal Complements):

1. If $U \subset V$ is a subspace, then $V = U \oplus U^\perp$.
2. If $A \subset B$, then $B^\perp \subset A^\perp$.

Theorem 11 (Fundamental Theorem of Linear Algebra): For a linear map $T : V \to W$:

$$\boxed{\operatorname{Im} T^* = (\ker T)^\perp} \quad (11)$$

Equivalently: $\operatorname{Im} T = (\ker T^*)^\perp$.

Proof:

1. Show $\operatorname{Im} T^* \subset (\ker T)^\perp$: Let $v = T^*(w) \in \operatorname{Im} T^*$. For any $u \in \ker T$:

$$\langle v, u\rangle = \langle T^*(w), u\rangle = \langle w, T(u)\rangle = \langle w, 0\rangle = 0$$

   Thus, $v \perp u$ for all $u \in \ker T$.
2. Show $(\ker T)^\perp \subset \operatorname{Im} T^*$: Let $v \in (\ker T)^\perp$. Then $\langle v, u\rangle = 0$ for all $u \in \ker T$. This implies $v \in \operatorname{Im} T^*$.
3. Therefore, $\operatorname{Im} T^* = (\ker T)^\perp$.

□

Theorem 12 (Theorem of the Alternative): Exactly one of the following statements holds: 1. There exists $x \in \mathbb{R}^m$ such that $Ax = b$. 2. There exists $y \in \mathbb{R}^n$ such that $y^T A = 0$ and $b \cdot y \neq 0$.

This is the basic duality result of linear programming.

9.6 Summary of Key Results

Concept	Definition	Key Formula
Linear Map	Preserves addition and scalar multiplication	$T(u+v) = T(u) + T(v)$, $T(\alpha v) = \alpha T(v)$
Matrix of Linear Map	Columns are images of basis vectors	$T(e_j) = \sum_i \alpha_{ij} f_i$
Kernel	Vectors mapping to zero	$\ker T = \{v : T(v) = 0\}$
Image	Set of all outputs	$\operatorname{Im} T = \{T(v) : v \in V\}$
Injectivity	Kernel is trivial	$\ker T = \{0\} \iff T$ injective
Rank-Nullity	Dimension of domain equals sum	$\dim V = \dim(\ker T) + \dim(\operatorname{Im} T)$
Direct Sum	Unique decomposition	$V = U \oplus W$ with $U \cap W = \{0\}$
Projection	Idempotent linear map	$P^2 = P$
Adjoint	Reverse map w.r.t. inner product	$\langle T^* w, v\rangle = \langle w, Tv\rangle$
Fundamental Theorem	Image of adjoint = orthogonal complement of kernel	$\operatorname{Im} T^* = (\ker T)^\perp$

Chapter 10: Differentiability - Part II

10.1 Differentiability in $\mathbb{R}^n$

10.1.1 Motivation

The problem that differentiability in $\mathbb{R}^n$ solves is the **local linear approximation of multivariable functions**. While in $\mathbb{R}$ differentiability means the existence of a tangent line, in $\mathbb{R}^n$ it means the existence of a tangent plane (or hyperplane). The derivative is no longer a single number but a linear map—the Jacobian—that captures how the function changes in all directions simultaneously.

Geometric Interpretation: Visualize a function $f : \mathbb{R}^2 \to \mathbb{R}$ as a surface. Differentiability at a point means the surface has a well-defined tangent plane. The derivative is the linear map that sends a small change in the input to the corresponding change in the output, to first order.

10.1.2 Formal Definitions

Definition 1 (Differentiability in $\mathbb{R}^n$): Let $U \subset \mathbb{R}^n$ be open, $a \in U$, and $f : U \to \mathbb{R}^m$. We say that f is **differentiable** at a if there exists a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ such that:

$$\boxed{f(a+h) = f(a) + T(h) + r(h), \quad \lim_{h\to 0} \frac{r(h)}{\|h\|} = 0} \tag{1}$$

We write $Df(a) = T$ and call $Df(a)$ the **derivative** of f at a.

Definition 2 (Jacobian Matrix): In the canonical basis, the matrix of $Df(a) : \mathbb{R}^n \to \mathbb{R}^m$ is the **Jacobian**:

$$\boxed{Df(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}} \tag{2}$$

10.1.3 Key Theorems

Theorem 1 (Differentiability Implies Continuity): If $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $a \in U$, then f is continuous at a.

Proof:

1. Let (x_k) be a sequence converging to a.
2. By differentiability:

$$\|f(x_k) - f(a)\| \leq \|Df(a) \cdot (x_k - a)\| + \|r(x_k - a)\|$$

3. Since $Df(a)$ is linear, $\|Df(a) \cdot (x_k - a)\| \to 0$.
4. Since $\lim_{h \to 0} \frac{r(h)}{\|h\|} = 0$, we have $r(x_k - a) \to 0$.
5. Therefore, $f(x_k) \to f(a)$, so f is continuous at a.

□

Theorem 2 (Coordinate-wise Differentiability): A function $f = (f_1, \dots, f_m) : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a iff each coordinate function f_i is differentiable at a.

Proof: Follows directly from the definition: the remainder term $r(h)$ goes to zero iff each coordinate remainder $r_i(h)$ goes to zero.

□

Theorem 3 (Directional Derivatives from Derivative): If f is differentiable at a, then the directional derivative $\partial_v f(a)$ exists and:

$$\boxed{\partial_v f(a) = Df(a) \cdot v} \tag{3}$$

Proof:

1. By differentiability:

$$f(a + tv) = f(a) + Df(a) \cdot (tv) + r(tv)$$

2. Then:

$$\frac{f(a + tv) - f(a)}{t} = Df(a) \cdot v + \frac{r(tv)}{t}$$

3. Since $\frac{r(tv)}{t} = \frac{r(tv)}{\|tv\|} \cdot \|v\| \to 0$, the limit exists:

$$\partial_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t} = Df(a) \cdot v$$

□

10.2 The Chain Rule in $\mathbb{R}^n$

10.2.1 Motivation

The Chain Rule in $\mathbb{R}^n$ generalizes the one-dimensional chain rule to multivariable functions. It states that the derivative of a composition is the composition of the derivatives. This is one of the most important results in multivariable calculus, allowing us to compute derivatives of complex functions by breaking them down into simpler pieces.

Geometric Interpretation: The derivative $Df(a)$ is a linear map that approximates f near a. The derivative $Dg(b)$ approximates g near $b = f(a)$. The composition $Dg(b) \circ Df(a)$ approximates $g \circ f$ near a.

10.2.2 Formal Statement

Theorem 4 (Chain Rule): Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open sets, $a \in U$. Let $f : U \to \mathbb{R}^m$ be differentiable at a, and $g : V \to \mathbb{R}^p$ be differentiable at $b = f(a)$, with $f(U) \subset V$. Then $g \circ f : U \to \mathbb{R}^p$ is differentiable at a and:

$$\boxed{D(g \circ f)(a) = Dg(b) \cdot Df(a)} \quad (4)$$

10.2.3 Proof

Proof:

1. Since g is differentiable at b:

$$g(f(a+h)) - g(f(a)) = Dg(b) \cdot (f(a+h) - f(a)) + R(f(a+h) - f(a))$$

where $\lim_{v \to 0} \frac{R(v)}{\|v\|} = 0$.

2. Since f is differentiable at a:

$$f(a+h) - f(a) = Df(a) \cdot h + r(h)$$

where $\lim_{h \to 0} \frac{r(h)}{\|h\|} = 0$.

3. Substituting:

$$g(f(a+h)) - g(f(a)) = Dg(b) \cdot Df(a) \cdot h + Dg(b) \cdot r(h) + R(f(a+h) - f(a))$$

4. We must show:

$$\lim_{h \to 0} \frac{Dg(b) \cdot r(h) + R(f(a+h) - f(a))}{\|h\|} = 0$$

5. For the first term:

$$\lim_{h \to 0} \frac{Dg(b) \cdot r(h)}{\|h\|} = Dg(b) \cdot \lim_{h \to 0} \frac{r(h)}{\|h\|} = 0$$

6. For the second term:

$$\frac{R(f(a+h) - f(a))}{\|h\|} = \frac{R(f(a+h) - f(a))}{\|f(a+h) - f(a)\|} \cdot \frac{\|f(a+h) - f(a)\|}{\|h\|}$$

7. The first factor tends to 0 by differentiability of g. The second factor is bounded (since $\frac{\|Df(a) \cdot h\|}{\|h\|}$ is bounded).

8. Therefore, the composition is differentiable and the formula holds.

□

10.3 Sufficient Conditions for Differentiability

10.3.1 Motivation

Not every function with partial derivatives is differentiable. We need conditions that guarantee differentiability. A sufficient condition is that the partial derivatives exist and are continuous in a neighborhood of the point—the function is then said to be of class C^1.

Geometric Interpretation: If the partial derivatives are continuous, the tangent plane varies smoothly with the point. The function has no "kinks" or abrupt changes in slope.

10.3.2 Formal Statement

Theorem 5 (Sufficient Condition for Differentiability): Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$, $a \in U$. Suppose there exists $\delta > 0$ such that for all $j = 1, \dots, n$, the partial derivative $\partial_j f(x)$ exists at all $x \in B_\delta(a) \subset U$, and $\partial_j f : B_\delta(a) \to \mathbb{R}$ is continuous at a. Then f is differentiable at a.

10.3.3 Proof (for $n = 2$)

Proof:

1. For $h = (h_1, h_2)$, by the Mean Value Theorem:

$$f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2 + h_2) = \partial_1 f(a_1 + \theta_1(h_2)h_2, a_2 + h_2)h_1$$

   and

$$f(a_1, a_2 + h_2) - f(a) = \partial_2 f(a_1, a_2 + \theta_2(h_2)h_2)h_2$$

2. Define the remainder:

$$r(h) = f(a + h) - f(a) - \partial_1 f(a)h_1 - \partial_2 f(a)h_2$$

3. Then:

$$r(h) = (\partial_1 f(a_1 + \theta_1(h)h_1, a_2 + h_2) - \partial_1 f(a))h_1 + (\partial_2 f(a_1, a_2 + \theta_2(h_2)h_2) - \partial_2 f(a))h_2$$

4. Therefore:

$$\frac{|r(h)|}{\sqrt{h_1^2 + h_2^2}} \le \frac{|\partial_1 f(\cdot) - \partial_1 f(a)|}{\to 0} \cdot \frac{|h_1|}{\sqrt{h_1^2 + h_2^2}} + \frac{|\partial_2 f(\cdot) - \partial_2 f(a)|}{\to 0} \cdot \frac{|h_2|}{\sqrt{h_1^2 + h_2^2}}$$

5. Since $\frac{|h_i|}{\sqrt{h_1^2+h_2^2}} \le 1$, the whole expression tends to 0. Thus, f is differentiable at a.

□

10.4 Summary of Key Results

Concept	Definition	Key Formula
Differentiability in $\mathbb{R}^n$	Local linear approximation	$f(a+h) = f(a) + Df(a)h + r(h)$, $\frac{r(h)}{\Vert h\Vert} \to 0$
Jacobian Matrix	Matrix of partial derivatives	$Df(a) = \big(\partial_j f_i(a)\big)$
Derivative $\implies$ Continuity	Differentiable implies continuous	$x_k \to a \implies f(x_k) \to f(a)$
Directional Derivative	From the derivative	$\partial_v f(a) = Df(a) \cdot v$
Chain Rule	Derivative of composition	$D(g \circ f)(a) = Dg(f(a)) \cdot Df(a)$
Sufficient Condition	Continuous partial derivatives	$C^1 \implies$ differentiable
Coordinate-wise Differentiability	f diff iff each f_i diff	Follows from definition

---

Chapter 11: Implicit Functions Derivatives

11.1 Motivation and Intuition

11.1.1 Motivation

The problem that the Implicit Function Theorem (IFT) solves is the **local representation of implicitly defined relationships**. Many economic and physical relationships are not given explicitly as $y = f(x)$, but rather as equations $F(x, y) = 0$. The IFT provides conditions under which we can solve for y as a function of x locally, and compute its derivatives without actually finding the explicit formula.

Geometric Interpretation: Visualize the equation $f(x, y) = 0$ as a curve in $\mathbb{R}^2$. The Implicit Function Theorem says that if the curve is "smooth" and not vertical at a point (a, b), then near that point, the curve can be represented as the graph of a function $y = \xi(x)$. The condition $\partial_y f(a, b) \neq 0$ ensures the curve is not vertical—it has a well-defined slope.

11.1.2 Formal Definitions

Definition 1 (Continuously Differentiable): A function $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$ is **continuously differentiable** (or C^1) if f is differentiable and $Df : U \to L(\mathbb{R}^m, \mathbb{R}^n)$ is continuous. Equivalently, each partial derivative $\partial_j f : U \to \mathbb{R}^n$ is continuous.

11.2 The Implicit Function Theorem

11.2.1 Formal Statement

Theorem 1 (Implicit Function Theorem for $f : \mathbb{R}^{n+1} \to \mathbb{R}$): Let $f : U \subset \mathbb{R}^{n+1} \to \mathbb{R}$ be continuously differentiable. Let $(\bar{x}, \bar{y}) \in U$ be such that $f(\bar{x}, \bar{y}) = 0$ and $\partial_{n+1} f(\bar{x}, \bar{y}) \neq 0$. Then there exist $\delta > 0$ and $\epsilon > 0$ such that:

1. For every $x \in B_\epsilon(\bar{x}) \subset \mathbb{R}^n$, there exists a unique $\xi(x) \in \mathbb{R}$ such that:

$$f(x, \xi(x)) = 0 \quad \text{and} \quad (x, \xi(x)) \in B_\delta(\bar{x}, \bar{y})$$

2. The function $\xi : B_\epsilon(\bar{x}) \to \mathbb{R}$ is continuously differentiable, and:

$$\boxed{\frac{\partial \xi(x)}{\partial x_j} = -\frac{\frac{\partial f}{\partial x_j}(x, \xi(x))}{\frac{\partial f}{\partial y}(x, \xi(x))}} \qquad (1)$$

11.2.2 Proof (for $n = 1$)

Proof:

1. **Existence and Uniqueness:** Suppose $\frac{\partial f}{\partial y}(\bar{x}, \bar{y}) > 0$. Then $\exists \delta > 0$ such that $\frac{\partial f}{\partial y}(x, y) > 0$ for all $(x, y) \in B_\delta(\bar{x}, \bar{y})$. Thus, there exists $\mu > 0$ such that:

$$f(\bar{x}, \bar{y} - \mu) < 0 = f(\bar{x}, \bar{y}) < f(\bar{x}, \bar{y} + \mu)$$

2. By continuity of f, there exists $\epsilon > 0$ such that for all $x \in B_\epsilon(\bar{x})$:

$$f(x, \bar{y} - \mu) < 0 < f(x, \bar{y} + \mu)$$

3. By the Intermediate Value Theorem, for each $x \in B_\epsilon(\bar{x})$, there exists $\xi(x) \in (\bar{y} - \mu, \bar{y} + \mu)$ such that $f(x, \xi(x)) = 0$.
4. Uniqueness follows because $f(x, \cdot)$ is strictly increasing on $B_\delta(\bar{x}, \bar{y})$.
5. **Differentiability:** For $h \in (-\epsilon, \epsilon)$, by the Mean Value Theorem, $\exists \theta(h) \in (0, 1)$ such that:

$$0 = f(x + h, \xi(x + h)) - f(x, \xi(x))$$

$$= \frac{\partial f}{\partial x}(x + \theta h, \xi(x) + \theta(\xi(x + h) - \xi(x)))h + \frac{\partial f}{\partial y}(\cdot)(\xi(x + h) - \xi(x))$$

6. Therefore, for $h \neq 0$:

$$\frac{\xi(x+h)-\xi(x)}{h} = -\frac{\frac{\partial f}{\partial x}(x+\theta h, \xi(x)+\theta(\xi(x+h)-\xi(x)))}{\frac{\partial f}{\partial y}(x+\theta h, \xi(x)+\theta(\xi(x+h)-\xi(x)))}$$

7. Taking the limit as $h \to 0$, we obtain:

$$\xi'(x) = -\frac{\frac{\partial f}{\partial x}(x,\xi(x))}{\frac{\partial f}{\partial y}(x,\xi(x))}$$

$\square$

11.2.3 Derivative Formula via Chain Rule

Alternatively, if we know ξ is well-defined and differentiable, since:

$$g(x_1,\dots,x_n) = f(x_1,\dots,x_n,\xi(x_1,\dots,x_n)) = 0$$

for all $x \in B_\epsilon(\bar{x})$, then for each $j = 1,\dots,n$:

$$0 = \frac{\partial g}{\partial x_j}(x) = \frac{\partial f}{\partial x_j}(x,\xi(x)) + \frac{\partial f}{\partial y}(x,\xi(x)) \cdot \frac{\partial \xi}{\partial x_j}(x)$$

Solving for $\frac{\partial \xi}{\partial x_j}$:

$$\boxed{\frac{\partial \xi(x)}{\partial x_j} = -\frac{\frac{\partial f}{\partial x_j}(x,\xi(x))}{\frac{\partial f}{\partial y}(x,\xi(x))}}$$

11.3 Implicit Function Theorem in $\mathbb{R}^n$

11.3.1 Motivation

The general version of the Implicit Function Theorem handles systems of equations. Instead of one equation in $n+1$ variables, we have n equations in $m+n$ variables. The condition for solvability is that the Jacobian of the system with respect to the y-variables is invertible.

Geometric Interpretation: Visualize a system of n equations in $m+n$ variables. The solution set is an m-dimensional surface. The IFT says that if the Jacobian with respect to y is invertible at a point, the surface locally looks like the graph of a function $y = \xi(x)$.

11.3.2 Formal Statement

Theorem 2 (Implicit Function Theorem for Systems): Let $f : U \subset \mathbb{R}^m \oplus \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. Let $(a_1, a_2) \in U$ be such that $f(a_1, a_2) = 0$. Suppose:

$$\det(D_2 f(a_1, a_2)) \neq 0$$

where $D_2 f$ is the $n \times n$ matrix of partial derivatives with respect to the y-variables.

Then there exist $\epsilon > 0$ and $\delta > 0$ such that:

1. For every $x \in B_\epsilon(a_1)$, there exists a unique $\xi(x) \in \mathbb{R}^n$ such that:

$$f(x,\xi(x)) = 0 \quad \text{and} \quad (x,\xi(x)) \in B_\delta(a_1, a_2)$$

2. The function $\xi : B_\epsilon(a_1) \to \mathbb{R}^n$ is continuously differentiable, and:

$$\boxed{D\xi(x) = -(D_2 f(x,\xi(x)))^{-1} D_1 f(x,\xi(x))} \quad (2)$$

11.3.3 Matrix Form

Write $f = (f_1, \ldots, f_n)$. The derivative $Df(x, y)$ has the block form:

$$Df(x,y) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} & \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} & \frac{\partial f_n}{\partial y_1} & \cdots & \frac{\partial f_n}{\partial y_n} \end{pmatrix}$$

Then: - $D_1 f(a) \in \mathbb{R}^{n\times m}$ (derivatives with respect to x) - $D_2 f(a) \in \mathbb{R}^{n\times n}$ (derivatives with respect to y)

The condition $\det(D_2 f(a)) \neq 0$ ensures that the system can be solved for y in terms of x near a.

11.3.4 Motivation for the Formula

The formula $D\xi(x) = -(D_2 f)^{-1} D_1 f$ comes from differentiating the identity:

$$f(x, \xi(x)) = 0$$

By the Chain Rule:

$$D_1 f(x, \xi(x)) + D_2 f(x, \xi(x)) \cdot D\xi(x) = 0$$

Solving for $D\xi(x)$ gives the formula.

11.4 Summary of Key Results

Concept	Definition	Key Formula
Continuously Differentiable	Differentiable with continuous derivative	$f \in C^1$
Implicit Function Theorem (1D)	Solve $f(x, y) = 0$ for $y = \xi(x)$	$\xi'(x) = -\frac{f_x}{f_y}$
Implicit Function Theorem (Systems)	Solve $f(x, y) = 0$ for $y = \xi(x)$	$D\xi(x) = -(D_2 f)^{-1} D_1 f$
Condition for Solvability	Jacobian w.r.t. y is invertible	$\det(D_2 f(a)) \neq 0$
Derivative via Chain Rule	Differentiate $f(x, \xi(x)) = 0$	$D_1 f + D_2 f \cdot D\xi = 0$

Chapter 12: Higher Order Derivatives

12.1 Higher Order Derivatives in $\mathbb{R}$

12.1.1 Motivation

Higher order derivatives measure the rate of change of the derivative itself. The second derivative measures concavity—how the slope changes. Higher order derivatives are essential for Taylor approximation, optimization (second derivative test), and the study of curvature.

Geometric Interpretation: The first derivative gives the slope of the tangent line. The second derivative gives the curvature—whether the curve bends upward (convex) or downward (concave). Higher order derivatives capture more subtle aspects of the shape of the function.

12.1.2 Formal Definitions

Definition 1 (Second Derivative): A function $f : I \subset \mathbb{R} \to \mathbb{R}$ is **twice differentiable** at $a \in I$ if there exists $\delta > 0$ such that f is differentiable at every $x \in (a-\delta, a+\delta)$ and $f' : (a-\delta, a+\delta) \to \mathbb{R}$ is differentiable at a. The **second derivative** is:

$$f''(a) = \lim_{h \to 0} \frac{f'(a+h) - f'(a)}{h} \qquad (1)$$

Definition 2 (p-th Derivative): A function $f : I \subset \mathbb{R} \to \mathbb{R}$ is **p-times differentiable** at $a \in I$ if there exists $\delta > 0$ such that f is differentiable at every $x \in (a-\delta, a+\delta)$ and $f' : (a-\delta, a+\delta) \to \mathbb{R}$ is $(p-1)$-times differentiable at a. The p-th derivative is:

$$f^{(p)}(a) = \lim_{h \to 0} \frac{f^{(p-1)}(a+h) - f^{(p-1)}(a)}{h} \qquad (2)$$

12.2 Taylor's Theorem

12.2.1 Motivation

Taylor's Theorem is one of the most important results in analysis. It shows that a smooth function can be approximated by a polynomial of degree p, with an error term that goes to zero faster than h^p. This is the foundation of numerical methods, optimization, and the study of local behavior of functions.

Geometric Interpretation: Taylor's Theorem says that near a point a, the graph of a smooth function looks like its Taylor polynomial of degree p. The error is smaller than h^p, so the approximation improves as p increases.

12.2.2 Auxiliary Lemma

Lemma 1: Let $r : I \to \mathbb{R}$ be n-times differentiable at $0 \in I$. Then:

$$\lim_{x \to 0} \frac{r(x)}{x^n} = 0 \iff r(0) = r'(0) = \cdots = r^{(n)}(0) = 0$$

Proof for ($\Leftarrow$, $n = 1$): If $r(0) = r'(0) = 0$, then:

$$\frac{r(x)}{x} = \frac{r(x) - r(0)}{x - 0} \to r'(0) = 0$$

Inductive Step: Suppose the result holds for $n-1$. Assume $r(0) = \cdots = r^{(n)}(0) = 0$. By the Mean Value Theorem, for each x, there exists $\theta_x \in (0,1)$ such that:

$$r(x) = r'(\theta_x)x$$

Then:

$$\frac{r(x)}{x^n} = \frac{r'(\theta_x)}{x^{n-1}} \cdot \frac{\theta_x^{n-1}}{\theta_x^{n-1}} = \frac{r'(\theta_x)}{\theta_x^{n-1}} \cdot \left(\frac{\theta_x}{x}\right)^{n-1}$$

By the induction hypothesis applied to r', $\frac{r'(\theta_x)}{\theta_x^{n-1}} \to 0$. Thus, $\frac{r(x)}{x^n} \to 0$.

$\square$

12.2.3 Taylor's Formula

Theorem 1 (Taylor's Formula with Peano Remainder): Let $f : I \to \mathbb{R}$ be p-times differentiable at $a \in I$. Then:

$$\boxed{f(a+h) = \sum_{i=0}^{p} \frac{f^{(i)}(a)}{i!} h^i + r_p(h)}$$

where $r_p(h) = f(a+h) - \sum_{i=0}^{p} \frac{f^{(i)}(a)}{i!} h^i$ and:

$$\boxed{\lim_{h \to 0} \frac{r_p(h)}{h^p} = 0} \quad (3)$$

Proof:

1. Define $r(h) = f(a+h) - \sum_{i=0}^{p} \frac{f^{(i)}(a)}{i!} h^i$.
2. Then $r(0) = r'(0) = \cdots = r^{(p)}(0) = 0$.
3. By Lemma 1, $\lim_{h \to 0} \frac{r(h)}{h^p} = 0$.

□

Theorem 2 (Taylor's Formula with Lagrange Remainder): Let $f : I \to \mathbb{R}$ be p-times differentiable on I. Then for all $h \in \mathbb{R}$ with $a + h \in I$, there exists $\theta \in (0, 1)$ such that:

$$\boxed{f(a+h) = \sum_{i=0}^{p-1} \frac{f^{(i)}(a)}{i!} h^i + \frac{f^{(p)}(a + \theta h)}{p!} h^p}$$

12.3 Higher Order Derivatives in $\mathbb{R}^n$

12.3.1 Motivation

In $\mathbb{R}^n$, higher order derivatives become tensors. The second derivative is a bilinear map (the Hessian matrix). The p-th derivative is a p-linear map. These higher derivatives are essential for Taylor approximations in multiple dimensions and for optimization.

Geometric Interpretation: The Hessian matrix captures the curvature of a function in all directions. Its eigenvalues determine whether a critical point is a minimum, maximum, or saddle point.

12.3.2 Formal Definitions

Definition 3 (Mixed Partial Derivatives): For $f : U \subset \mathbb{R}^n \to \mathbb{R}$, the mixed partial derivative is:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right)(x) \quad (4)$$

12.3.3 Key Theorems

Theorem 3 (Schwarz's Theorem): If $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at $a \in U$, then for all $i, j = 1, \ldots, n$:

$$\boxed{\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)} \quad (5)$$

This theorem states that mixed partial derivatives commute when the function is sufficiently smooth.

Theorem 4 (Higher Order Schwarz Theorem): If $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is p-times differentiable at $a \in U$, then $D^p f(a) \in L^p(\mathbb{R}^n \times \cdots \times \mathbb{R}^n, \mathbb{R}^m)$ is symmetric. That is, for any permutation π of $\{1, \ldots, p\}$:

$$D^p f(a)(v_1, \ldots, v_p) = D^p f(a)(v_{\pi(1)}, \ldots, v_{\pi(p)}) \quad (6)$$

12.4 The Hessian Matrix

12.4.1 Formal Definition

Definition 4 (Hessian Matrix): For $f : U \subset \mathbb{R}^n \to \mathbb{R}$, the **Hessian matrix** at a is the $n \times n$ matrix:

$$\boxed{\nabla^2 f(a) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{pmatrix}} \quad (7)$$

The Hessian is symmetric by Schwarz's Theorem.

12.4.2 The Gradient

Definition 5 (Gradient): The **gradient** of f at a is:

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a) \right)^T \quad (8)$$

12.4.3 Taylor's Formula in $\mathbb{R}^n$

Theorem 5 (Taylor's Formula in $\mathbb{R}^n$): Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ be p-times differentiable at $a \in U$. Then:

$$\boxed{f(a+h) = f(a) + Df(a)h + \frac{1}{2} D^2 f(a)(h,h) + \cdots + \frac{1}{p!} D^p f(a)(h, \ldots, h) + r_p(h)}$$

with:

$$\lim_{h \to 0} \frac{r_p(h)}{\|h\|^p} = 0 \quad (9)$$

In particular, for $p = 2$:

$$\boxed{f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} h^T \nabla^2 f(a) h + r_2(h)}$$

with:

$$\lim_{h \to 0} \frac{r_2(h)}{\|h\|^2} = 0 \quad (10)$$

12.4.4 Relationship Between $D^2 f$ and the Hessian

For a twice differentiable function $f : U \subset \mathbb{R}^n \to \mathbb{R}$:

$$D^2 f(a)(v, w) = v^T \nabla^2 f(a) w \quad (11)$$

The Hessian matrix represents the bilinear form $D^2 f(a)$ in the canonical basis:

$$\nabla^2 f(a)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \quad (12)$$

12.5 Summary of Key Results

Concept	Definition	Key Formula
Second Derivative	Derivative of derivative	$f''(a) = \lim_{h\to 0} \frac{f'(a+h)-f'(a)}{h}$
p-th Derivative	p-times differentiable	$f^{(p)}(a) = \lim_{h\to 0} \frac{f^{(p-1)}(a+h)-f^{(p-1)}(a)}{h}$
Taylor's Formula (Peano)	Polynomial approximation	$f(a+h) = \sum_{i=0}^{p} \frac{f^{(i)}(a)}{i!}h^i + r_p(h)$, $\frac{r_p(h)}{h^p} \to 0$
Taylor's Formula (Lagrange)	Exact remainder	$f(a+h) = \sum_{i=0}^{p-1} \frac{f^{(i)}(a)}{i!}h^i + \frac{f^{(p)}(a+\theta h)}{p!}h^p$
Schwarz's Theorem	Mixed partials commute	$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$
Gradient	Vector of first partials	$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\right)^T$
Hessian Matrix	Matrix of second partials	$\nabla^2 f(a)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(a)$
Taylor in $\mathbb{R}^n$ ($p=2$)	Quadratic approximation	$f(a+h) = f(a) + \langle \nabla f(a), h\rangle + \frac{1}{2}h^T\nabla^2 f(a)h + r_2(h)$

Chapter 13: Optimization

13.1 Basic Concepts and Definitions

13.1.1 Motivation

The problem that optimization solves is the **selection of the best alternative** among a set of feasible options. In economics, this translates to maximizing utility, profit, or welfare, or minimizing cost, risk, or error. Optimization theory provides the mathematical tools to identify optimal solutions and characterize their properties.

Geometric Interpretation: Visualize the objective function $f : \mathbb{R}^n \to \mathbb{R}$ as a surface over the feasible set X. Maximization is the search for the highest point on this surface, subject to staying within the feasible region.

13.1.2 Formal Definitions

Definition 1 (Optimization Problem): A general optimization problem is:

$$(P) : \max_{x \in U \subset \mathbb{R}^m} f(x) \quad \text{subject to} \quad g_i(x) \geq 0, \quad i = 1, \ldots, n$$

where $f, g_1, \ldots, g_n : U \subset \mathbb{R}^m \to \mathbb{R}$ are twice differentiable (C^2).

Definition 2 (Feasible Set): The **feasible set** is:

$$X = \{x \in U : g_i(x) \geq 0,\ i = 1, \ldots, n\} \quad (1)$$

13.1.3 Types of Optimizers

Definition 3: Let $a \in X$. We say that:

Type	Condition
Global maximizer	$f(a) \geq f(x)$ for all $x \in X$
Strict global maximizer	$f(a) > f(x)$ for all $x \in X \setminus \{a\}$

Type	Condition
Local maximizer	$\exists\delta > 0$ such that $f(a) \geq f(x)$ for all $x \in X \cap B_\delta(a)$
Strict local maximizer	$\exists\delta > 0$ such that $f(a) > f(x)$ for all $x \in X \cap B_\delta(a) \setminus \{a\}$

13.2 Definiteness of Matrices

13.2.1 Formal Definitions

Definition 4 (Positive Definite/Semi-definite): Let A be a symmetric $n \times n$ matrix.

Type	Condition
Positive semi-definite ($A \succeq 0$)	$v^T A v \geq 0$ for all $v \in \mathbb{R}^n$
Positive definite ($A \succ 0$)	$v^T A v > 0$ for all $v \in \mathbb{R}^n \setminus \{0\}$
Negative semi-definite ($A \preceq 0$)	$-A \succeq 0$
Negative definite ($A \prec 0$)	$-A \succ 0$

For a 2×2 matrix $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$: - $A \succ 0 \iff a > 0$ and $ac - b^2 > 0$ - $A \prec 0 \iff a < 0$ and $ac - b^2 > 0$

13.3 Unconstrained Optimization

13.3.1 Motivation

Unconstrained optimization seeks the maximum or minimum of a function over the entire domain, without any restrictions. The first-order condition ($\nabla f(a) = 0$) identifies critical points. The second-order condition (definiteness of the Hessian) determines whether a critical point is a maximum, minimum, or saddle point.

13.3.2 Key Theorems

Theorem 1 (Necessary Conditions for Local Maximum): If f is twice differentiable at $a \in U$ and a is a local maximizer of f, then: 1. $\nabla f(a) = 0$ (First-order condition) 2. $\nabla^2 f(a) \preceq 0$ (Second-order condition)

Proof of Second-Order Condition:

1. Since a is a local maximizer, for any $v \in \mathbb{R}^n$ and small $t > 0$:

$$f(a + tv) - f(a) \leq 0$$

2. By Taylor's formula:

$$f(a + tv) - f(a) = \nabla f(a) \cdot (tv) + \frac{1}{2}(tv)^T \nabla^2 f(a)(tv) + r(tv)$$

3. Since $\nabla f(a) = 0$, divide by $t^2 > 0$:

$$0 \geq v^T \nabla^2 f(a) v + \frac{r(tv)}{t^2}$$

4. Taking the limit as $t \to 0$:

$$0 \geq v^T \nabla^2 f(a) v$$

Thus, $\nabla^2 f(a) \preceq 0$.

□

Theorem 2 (Sufficient Conditions for Strict Local Maximum): If f is twice differentiable at $a \in U$, $\nabla f(a) = 0$, and $\nabla^2 f(a) \prec 0$, then a is a strict local maximizer.

Proof:

1. Suppose, for contradiction, that a is not a strict local maximizer.
2. Then there exists a sequence $x_k \in U$, $x_k \neq a$, $x_k \to a$, such that $f(x_k) \geq f(a)$.
3. By Taylor's formula:

$$f(x_k) - f(a) = \frac{1}{2}(x_k - a)^T \nabla^2 f(a)(x_k - a) + r(x_k - a)$$

   where $\frac{r(h)}{\|h\|^2} \to 0$.

4. Dividing by $\|x_k - a\|^2$:

$$0 \leq \frac{1}{2}\left(\frac{x_k - a}{\|x_k - a\|}\right)^T \nabla^2 f(a)\left(\frac{x_k - a}{\|x_k - a\|}\right) + \frac{r(x_k - a)}{\|x_k - a\|^2}$$

5. By Bolzano-Weierstrass, there is a subsequence with $\frac{x_k - a}{\|x_k - a\|} \to v$ with $\|v\| = 1$.
6. Taking the limit gives $0 \leq \frac{1}{2} v^T \nabla^2 f(a) v$, contradicting $\nabla^2 f(a) \prec 0$.

□

13.4 Equality Constraints

13.4.1 Motivation

Equality constraints arise when resources must be exactly allocated or when specific relationships must be satisfied. The Lagrange multiplier theorem provides necessary conditions for optimality under equality constraints.

13.4.2 Formal Statement

Theorem 3 (Lagrange Multiplier Theorem): Let a be a local maximizer of f subject to $g_i(x) = 0$ for $i = 1, \ldots, n$. Suppose that f is differentiable at a, each g_i is C^1, and $\nabla g_1(a), \ldots, \nabla g_n(a)$ are linearly independent. Then there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that:

$$\boxed{\nabla f(a) = \sum_{i=1}^{n} \lambda_i \nabla g_i(a)} \quad (2)$$

13.4.3 Proof Sketch

1. Define $g = (g_1, \ldots, g_n) : U \subset \mathbb{R}^m \to \mathbb{R}^n$.
2. Since the rows of $Dg(a)$ are linearly independent, there exist n linearly independent columns.
3. Relabel variables so that $a = (a_1, a_2)$ with $a_2 \in \mathbb{R}^n$, and $D_2 g(a)$ is invertible.
4. By the Implicit Function Theorem, there exists a C^1 function $\xi : B_\delta(a_1) \to \mathbb{R}^n$ such that:

$$g(x, \xi(x)) = 0 \quad \text{and} \quad (x, \xi(x)) \in B_\epsilon(a)$$

5. Define $\phi(x) = (x, \xi(x))$. Then a_1 is an unconstrained local maximum of $f \circ \phi$.
6. Therefore, $D(f \circ \phi)(a_1) = 0$, so $\nabla f(a)^T D\phi(a_1) = 0$.
7. This implies $\nabla f(a) \in (\operatorname{Im} D\phi(a_1))^\perp$.
8. Since $g(\phi(x)) = 0$, we have $Dg(a) \cdot D\phi(a_1) = 0$.
9. The LICQ condition implies equality, so $\nabla f(a) \in \operatorname{span}\{\nabla g_1(a), \ldots, \nabla g_n(a)\}$.

□

13.5 Inequality Constraints

13.5.1 Motivation

Inequality constraints are the most common in economics: resources are typically limited by upper or lower bounds, not exact equalities. The Kuhn-Tucker conditions generalize the Lagrange multiplier method to handle inequality constraints.

13.5.2 Formal Definitions

Definition 5 (Active Set): For $a \in X$, the **active set** of inequality constraints is:

$$I(a) = \{i : g_i(a) = 0\} \quad (3)$$

Definition 6 (Tangent Cone): The **tangent cone** to X at a is:

$$T_X(a) = \{\lambda_k(x_k - a) : x_k \to a,\ x_k \in X,\ \lambda_k \geq 0\} \quad (4)$$

Definition 7 (Polar Cone): The **polar** of a set C is:

$$C^0 = \{y : y^T x \leq 0 \text{ for all } x \in C\} \quad (5)$$

13.5.3 Key Theorems

Theorem 4 (Necessary Condition for Local Maximum): If f, g, h are differentiable at $a \in X$ and a is a local maximizer of (NP), then:

$$\boxed{\nabla f(a) \in (T_X(a))^0} \quad (6)$$

Proof:

1. Since a is a local maximizer, there exists $\delta > 0$ such that $f(x) \leq f(a)$ for all $x \in X \cap B_\delta(a)$.
2. Let $z \in T_X(a)$. Then $z = \lim_{k\to\infty} \lambda_k(x_k - a)$ for some $x_k \in X$, $x_k \to a$, $\lambda_k \geq 0$.
3. By differentiability:
$$f(x_k) - f(a) = \nabla f(a)^T (x_k - a) + r(x_k - a)$$
4. Since $f(x_k) \leq f(a)$:
$$0 \geq \nabla f(a)^T (x_k - a) + r(x_k - a)$$
5. Multiplying by $\lambda_k \geq 0$ and taking the limit:
$$0 \geq \nabla f(a)^T z$$

Thus, $\nabla f(a) \in (T_X(a))^0$.

□

13.6 Summary of Key Results

Concept	Definition	Key Formula
Feasible Set	Set satisfying constraints	$X = \{x : g_i(x) \geq 0, h_j(x) = 0\}$
Global Maximizer	Best over entire feasible set	$f(a) \geq f(x)\ \forall x \in X$
Local Maximizer	Best in a neighborhood	$f(a) \geq f(x)\ \forall x \in X \cap B_\delta(a)$
Positive Definite	$v^T Av > 0$ for $v \neq 0$	All eigenvalues positive
Negative Definite	$v^T Av < 0$ for $v \neq 0$	All eigenvalues negative
First-Order Condition	Necessary for interior optimum	$\nabla f(a) = 0$
Second-Order Condition	Hessian negative definite	$\nabla^2 f(a) \prec 0$
Tangent Cone	Directions of feasible movement	$T_X(a)$
Polar Cone	Directions with non-positive product	$C^0 = \{y : y^T x \leq 0\ \forall x \in C\}$

Chapter 14: Kuhn-Tucker Conditions

14.1 Motivation and Intuition

14.1.1 Motivation

The Kuhn-Tucker (KT) conditions generalize the Lagrange multiplier method to problems with inequality constraints. They provide necessary conditions for a feasible point to be optimal. The key innovation is the concept of **complementary slackness**: either a constraint is binding (active) or its associated multiplier is zero.

Geometric Interpretation: Visualize the feasible set X as a region bounded by constraint surfaces. At an optimum, the gradient of the objective must lie in the cone generated by the gradients of the active constraints. The Lagrange multipliers are the non-negative coefficients in this linear combination.

14.1.2 Formal Statement

Kuhn-Tucker Conditions: A feasible point a satisfies the **Kuhn-Tucker conditions** if there exist $\lambda_1, \ldots, \lambda_n \geq 0$ and $\mu_1, \ldots, \mu_p \in \mathbb{R}$ such that:

1. **First-Order Condition:**

$$\boxed{\nabla f(a) + \sum_{i=1}^{n} \lambda_i \nabla g_i(a) + \sum_{j=1}^{p} \mu_j \nabla h_j(a) = 0} \quad (1)$$

2. **Complementary Slackness:**

$$\boxed{\lambda_i g_i(a) = 0 \quad \forall i = 1, \ldots, n} \quad (2)$$

$$\lambda_i \geq 0 \quad \forall i = 1, \ldots, n$$

14.2 Farkas' Lemma

14.2.1 Formal Statement

Lemma 1 (Farkas' Lemma): Let A be an $n \times m$ matrix and $b \in \mathbb{R}^n$. Exactly one of the following two systems is feasible:

1. $Ax = b$, $x \geq 0$
2. $y^T A \geq 0$ and $b^T y < 0$

Proof:

1. **If (i) is feasible, (ii) is not:** If $Ax = b$ with $x \geq 0$, then for any y with $y^T A \geq 0$:
$$y^T b = y^T Ax \geq 0$$
So $b^T y \geq 0$, contradicting $b^T y < 0$.
2. **If (i) is not feasible:** Let $C = \{Ax : x \geq 0\}$, which is a closed convex cone. Since $b \notin C$, by the separation theorem, there exist $y \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that:
$$b^T y < c < y^T Ax \quad \forall x \geq 0$$
3. Since $0 \in C$, $c < y^T A \cdot 0 = 0$, so $c < 0$.
4. If $y^T Ae_i < 0$ for some i, then $y^T A(\alpha e_i) \to -\infty$ as $\alpha \to \infty$, contradicting $c < y^T Ax$ for all $x \geq 0$.
5. Thus, $y^T A \geq 0$ and $b^T y < c < 0$, so system (ii) is feasible.

□

14.3 Equivalence of KT and Polar Cone Conditions

14.3.1 Key Theorems

Theorem 1 (KT Equivalence): Let $a \in X$. Then the following are equivalent:

1. a satisfies the Kuhn-Tucker conditions.
2. $\nabla f(a) \in (L(a))^0$, where:
$$L(a) = \{z : Dg^*(a)z \geq 0,\ Dh(a)z = 0\} \quad (3)$$

Proof:

1. $(i) \Rightarrow (ii)$: If KT holds, then for any $z \in L(a)$:
$$\nabla f(a)^T z = -\sum_{i \in I(a)} \lambda_i \nabla g_i(a)^T z - \sum_{j=1}^{p} \mu_j \nabla h_j(a)^T z \leq 0$$
since $\lambda_i \geq 0$, $\nabla g_i(a)^T z \geq 0$, and $\nabla h_j(a)^T z = 0$.
2. $(ii) \Rightarrow (i)$: If $\nabla f(a) \in (L(a))^0$, then there do not exist z such that $Dg^*(a)z \geq 0$, $Dh(a)z = 0$, and $-\nabla f(a)^T z > 0$.
3. By Farkas' Lemma, this implies there exist $\lambda \in \mathbb{R}_+^{|I(a)|}$ and $\mu \in \mathbb{R}^p$ such that:
$$-\nabla f(a) = \sum_{i \in I(a)} \lambda_i \nabla g_i(a) + \sum_{j=1}^{p} \mu_j \nabla h_j(a)$$
4. Setting $\lambda_i = 0$ for $i \notin I(a)$ gives the KT conditions.

□

14.4 Constraint Qualifications

14.4.1 Motivation

Constraint qualifications (CQs) are conditions that ensure the KT conditions are necessary for optimality. Without a CQ, a local maximizer may not satisfy the KT conditions. The most common CQs are LICQ and MFCQ.

14.4.2 Formal Definitions

Definition 1 (LICQ - Linear Independence Constraint Qualification): The gradients of all active constraints are linearly independent:

$$\{\nabla g_i(a) : i \in I(a),\ \nabla h_j(a) : j = 1, \ldots, p\} \text{ are linearly independent} \quad (4)$$

Definition 2 (MFCQ - Mangasarian-Fromovitz Constraint Qualification): The gradients of equality constraints are linearly independent, and there exists $z \in \mathbb{R}^m$ such that:

$$Dg^*(a)z \gg 0 \quad \text{and} \quad Dh(a)z = 0 \quad (5)$$

Definition 3 (General CQ): A feasible point satisfies the general CQ if:

$$(T_X(a))^0 = (L(a))^0 \quad (6)$$

14.4.3 Key Theorems

Theorem 2 (Kuhn-Tucker Theorem): If f, g, h are differentiable at $a \in X$, a is a local maximizer, and a satisfies a constraint qualification, then there exist $\lambda \in \mathbb{R}^n_+$ and $\mu \in \mathbb{R}^p$ such that the KT conditions hold.

Theorem 3 (Implications between CQs): For $a \in X$:

$$\boxed{\text{LICQ at } a \implies \text{MFCQ at } a \implies \text{CQ at } a} \quad (7)$$

Proof of LICQ $\Rightarrow$ MFCQ:

1. Under LICQ, the matrix $C = [Dg^*(a)\ \ Dh(a)]$ has linearly independent rows and is therefore surjective.
2. Let $b = [1, \ldots, 1, 0, \ldots, 0]^T$ (ones for active inequalities, zeros for equalities).
3. Since C is surjective, there exists $z \in \mathbb{R}^m$ such that:

$$Dg^*(a)z = (1, \ldots, 1)^T \gg 0, \quad Dh(a)z = 0$$

4. Thus, MFCQ holds.

□

Proof of MFCQ $\Rightarrow$ CQ:

1. We need to show $T_X(a) = L(a)$.
2. Since $T_X(a) \subset L(a)$ always holds, it suffices to show $L(a) \subset T_X(a)$.
3. By MFCQ, there exists $\bar{z}$ with $Dg^*(a)\bar{z} \gg 0$ and $Dh(a)\bar{z} = 0$.
4. For any $z \in L(a)$, define $z_\epsilon = (1 - \epsilon)z + \epsilon\bar{z}$. Then $Dg^*(a)z_\epsilon \gg 0$ and $Dh(a)z_\epsilon = 0$.
5. By the Implicit Function Theorem, there exists a C^1 path $\bar{x}_\epsilon(t)$ such that $\bar{x}_\epsilon(0) = a$, $\bar{x}'_\epsilon(0) = z_\epsilon$, and $\bar{x}_\epsilon(t) \in X$.
6. Taking $\epsilon \to 0$ and $t \to 0$ appropriately shows $z \in T_X(a)$.

□

14.5 The Lagrangian and Sufficient Conditions

14.5.1 Formal Definitions

Definition 4 (Lagrangian): The **Lagrangian** for the optimization problem is:

$$\boxed{L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)} \quad (8)$$

Fact 1: If $x \in X$ and $\lambda \geq 0$, then $L(x, \lambda, \mu) \geq f(x)$.

Fact 2: If $x \in X$ and satisfies KT with respect to (λ, μ), then $f(x) = L(x, \lambda, \mu)$.

14.5.2 Key Theorems

Theorem 4 (Sufficient Second-Order Condition): Let f, g, h be twice differentiable at $a \in X$. Suppose (a, λ, μ) satisfies KT and:

$$z^T \nabla^2_{xx} L(a, \lambda, \mu) z < 0 \quad \forall z \in T_X(a) \setminus \{0\} \quad (9)$$

Then a is a strict local maximizer.

Proof:

1. For any $x \in X$, by Taylor's formula:

$$f(x) \leq L(x, \lambda, \mu) = L(a, \lambda, \mu) + \frac{1}{2}(x-a)^T \nabla^2_{xx} L(a)(x-a) + r(x-a)$$

2. Since $L(a, \lambda, \mu) = f(a)$ and $\nabla_x L(a, \lambda, \mu) = 0$:

$$f(x) \leq f(a) + \frac{1}{2}(x-a)^T \nabla^2_{xx} L(a)(x-a) + r(x-a)$$

3. If a is not a strict local maximizer, there exists $x_k \in X \setminus \{a\}$, $x_k \to a$, with $f(x_k) \geq f(a)$.
4. Then:

$$\frac{1}{2}(x_k - a)^T \nabla^2_{xx} L(a)(x_k - a) + r(x_k - a) \geq 0$$

5. Dividing by $\|x_k - a\|^2$ and taking a subsequence gives $z^T \nabla^2_{xx} L(a) z \geq 0$ for some $z \in T_X(a)$, contradicting (9).

□

14.6 Summary of Key Results

Concept	Definition	Key Formula
Kuhn-Tucker Conditions	Necessary for constrained optimum	$\nabla f + \sum \lambda_i \nabla g_i + \sum \mu_j \nabla h_j = 0$, $\lambda_i g_i = 0$
Complementary Slackness	Either constraint binds or multiplier is zero	$\lambda_i g_i(a) = 0$
Farkas' Lemma	Exactly one of two linear systems is feasible	$Ax = b, x \geq 0$ or $y^T A \geq 0, b^T y < 0$
LICQ	Active constraint gradients are independent	Linear independence of $\{\nabla g_i, \nabla h_j\}$
MFCQ	Exists direction with strict improvement in constraints	$Dg^* z \gg 0, Dhz = 0$

Concept	Definition	Key Formula
Lagrangian	Augmented objective with multipliers	$L = f + \sum \lambda_i g_i + \sum \mu_j h_j$
Second-Order Sufficient	Hessian negative on tangent cone	$z^T \nabla^2_{xx} L z < 0$ for $z \in T_X(a) \setminus \{0\}$

Chapter 15: Convex Analysis

15.1 Convex Sets

15.1.1 Motivation

Convexity is one of the most important concepts in optimization. In convex problems, local optima are global optima, and the KT conditions become sufficient. Convex analysis provides the geometric intuition and mathematical tools for studying convex sets and functions.

Geometric Interpretation: A set is convex if the line segment between any two points in the set lies entirely within the set. Convex sets have no "dents" or "holes."

15.1.2 Formal Definitions

Definition 1 (Convex Set): A set $C \subset \mathbb{R}^n$ is **convex** if for all $x, y \in C$ and all $t \in [0, 1]$:

$$tx + (1-t)y \in C \quad (1)$$

15.1.3 Key Theorems

Theorem 1 (Separation Theorem): Let C and K be closed convex sets with K compact. If $C \cap K = \emptyset$, then there exist $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that:

$$b^T x < c < b^T y \quad \forall x \in C,\ y \in K \quad (2)$$

Corollary: If C is a closed convex set and $a \notin C$, then there exist $b \in \mathbb{R}^n$ and $c \in \mathbb{R}$ such that:

$$b^T x < c < b^T a \quad \forall x \in C \quad (3)$$

Theorem 2 (Best Approximation Property): Let $C \subset \mathbb{R}^n$ be a closed convex set. Then for every $x \in \mathbb{R}^n$, there exists a unique $P(x) \in C$ such that:

$$\|x - P(x)\|^2 = \min_{y \in C} \|x - y\|^2 \quad (4)$$

Moreover, for all $y \in C$:

$$\langle y - P(x), x - P(x) \rangle \leq 0 \quad (5)$$

Proof:

1. **Existence:** Let $\gamma = \inf_{y \in C} \|x - y\|$. Choose $y_k \in C$ such that $\|x - y_k\| \to \gamma$.
2. Since (y_k) is bounded, it has a convergent subsequence $y_{k_j} \to y \in C$ (since C is closed).
3. Then $\|x - y\| = \lim \|x - y_{k_j}\| = \gamma$.

4. **Uniqueness:** If y, y' both minimize, then:

$$\left\| x - \frac{y+y'}{2} \right\|^2 < \frac{1}{2}\|x-y\|^2 + \frac{1}{2}\|x-y'\|^2 = \gamma^2$$

   Contradiction, so $y = y'$.

5. **Optimality condition:** For any $y \in C$ and $t \in [0,1]$:

$$\|x - P(x)\|^2 \leq \|x - P(x) - t(y - P(x))\|^2$$

6. Expanding:

$$0 \leq t^2\|y - P(x)\|^2 - 2t\langle x - P(x), y - P(x)\rangle$$

7. Dividing by $t > 0$ and letting $t \to 0$:

$$\langle x - P(x), y - P(x)\rangle \leq 0$$

□

15.2 Convex and Concave Functions

15.2.1 Motivation

Convex functions have the property that the line segment between any two points on the graph lies above the graph. This makes them easy to optimize: any local minimum is global. Concave functions are the natural functions for maximization in economics.

Geometric Interpretation: The epigraph (set of points above the graph) of a convex function is a convex set. The hypograph (set of points below the graph) of a concave function is a convex set.

15.2.2 Formal Definitions

Definition 2 (Epigraph and Hypograph): For a function $f : \mathbb{R}^n \to \mathbb{R}$:

- The **epigraph** is: $\text{epi}(f) = \{(x, h) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq h\}$.
- The **hypograph** is: $\text{hyp}(f) = \{(x, h) \in \mathbb{R}^n \times \mathbb{R} : f(x) \geq h\}$.

Definition 3 (Convex Function): A function f is **convex** if $\text{epi}(f)$ is a convex set.

Definition 4 (Concave Function): A function f is **concave** if $\text{hyp}(f)$ is a convex set.

15.2.3 Key Characterizations

Theorem 3 (Jensen's Inequality):

- f is convex iff:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \quad \forall x, y, \forall t \in [0,1] \quad (6)$$

- f is concave iff:

$$f(tx + (1-t)y) \geq tf(x) + (1-t)f(y) \quad \forall x, y, \forall t \in [0,1] \quad (7)$$

Theorem 4 (First-Order Characterization):

- If f is convex and differentiable at a, then:

$$\boxed{f(x) \geq f(a) + \nabla f(a)(x-a) \quad \forall x} \tag{8}$$

- If f is concave and differentiable at a, then:

$$\boxed{f(x) \leq f(a) + \nabla f(a)(x-a) \quad \forall x} \tag{9}$$

15.3 Quasiconcavity

15.3.1 Motivation

Quasiconcavity is a weaker condition than concavity that is often sufficient for optimization in economics. A function is quasiconcave if its upper contour sets are convex. This means that the set of points with value at least some threshold is convex.

15.3.2 Formal Definitions

Definition 5 (Quasiconcave Function): A function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ is **quasiconcave** if for all $c \in \mathbb{R}$, the upper contour set:

$$f^{-1}([c,\infty]) = \{x \in D : f(x) \geq c\} \tag{10}$$

is convex.

15.3.3 Key Theorems

Theorem 5 (Quasiconcavity and First-Order Conditions):

1. If f is quasiconcave and differentiable at $a \in D$, then:

$$f(x) \geq f(a) \implies \nabla f(a)(x-a) \geq 0 \tag{11}$$

2. If f is quasiconcave and differentiable at $a \in D$, then:

$$f(x) > f(a) \implies \nabla f(a)(x-a) > 0 \tag{12}$$

15.4 Convexity and Optimization

15.4.1 Key Theorems

Theorem 6 (Sufficiency of KT under Convexity): Consider the problem:

$$\max f(x) \quad \text{s.t.} \quad g_i(x) \geq 0,\ i = 1, \ldots, n$$

Suppose f is concave or quasiconcave, and g_i is quasiconcave for all $i \in I(a^*)$. If (a^*, λ) satisfies the KT conditions, then a^* is a global maximizer.

Proof:

1. Let $x \in X$, so $g_i(x) \geq 0 = g_i(a^*)$ for $i \in I(a^*)$.
2. By quasiconcavity of g_i:
$$\nabla g_i(a^*)(x-a^*) \geq 0 \quad \forall i \in I(a^*)$$
3. From KT: $\nabla f(a^*) = -\sum_{i \in I(a^*)} \lambda_i \nabla g_i(a^*)$.

4. Therefore:
$$\nabla f(a^*)(x-a^*) = -\sum_{i\in I(a^*)} \lambda_i \nabla g_i(a^*)(x-a^*) \le 0$$

5. If f is concave:
$$f(x) \le f(a^*) + \nabla f(a^*)(x-a^*) \le f(a^*)$$

6. If f is quasiconcave with $\nabla f(a^*) \neq 0$, then $f(x) \le f(a^*)$.

Thus, a^* is a global maximizer.

□

15.5 Summary of Key Results

Concept	Definition	Key Formula
Convex Set	Contains all line segments	$tx + (1-t)y \in C$
Separation Theorem	Separates disjoint convex sets	$b^T x < c < b^T y$
Best Approximation	Closest point in convex set	$\|x - P(x)\| = \min_{y\in C} \|x-y\|$
Convex Function	Epigraph is convex	$f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$
Concave Function	Hypograph is convex	$f(tx + (1-t)y) \ge tf(x) + (1-t)f(y)$
First-Order Convexity	Tangent is below convex function	$f(x) \ge f(a) + \nabla f(a)(x-a)$
First-Order Concavity	Tangent is above concave function	$f(x) \le f(a) + \nabla f(a)(x-a)$
Quasiconcavity	Upper contour sets are convex	$f^{-1}([c,\infty])$ convex
Quasiconcavity Condition	Implication of higher value	$f(x) \ge f(a) \implies \nabla f(a)(x-a) \ge 0$

Chapter 16: Envelope Theorem

16.1 Motivation and Intuition

16.1.1 Motivation

The Envelope Theorem is one of the most important results in economic theory. It answers the question: how does the optimal value of an optimization problem change when a parameter changes? The theorem states that the derivative of the value function with respect to a parameter equals the derivative of the Lagrangian with respect to that parameter, evaluated at the optimal solution. This simplifies the calculation of comparative statics enormously.

Geometric Interpretation: Visualize the value function $v(\theta)$ as the "envelope" of a family of objective functions $f(x,\theta)$ evaluated at the optimal x for each θ. The Envelope Theorem says that the slope of this envelope is determined by the direct effect of the parameter on the objective, ignoring the indirect effect through changes in x (since those effects vanish at the optimum).

16.1.2 Formal Setup

Consider the optimization problem with parameters:

$$(P_\theta): \quad \max_{x\in\mathbb{R}^m} f(x,\theta) \quad \text{s.t.} \quad g_i(x,\theta) \ge 0,\ i = 1,\dots,n$$

where $\theta \in D \subset \mathbb{R}^k$ is a parameter vector.

Definition 1 (Value Function): The **value function** is:

$$v(\theta) = f(\tilde{x}(\theta), \theta) \quad (1)$$

where $\tilde{x}(\theta)$ is the optimal solution for parameter θ.

16.2 The Implicit Function Theorem Setup

16.2.1 The KT System

Let $x_0 \in X(\theta_0)$ and $\lambda_0 \in \mathbb{R}^n_+$ satisfy the KT conditions at θ_0:

1. $\nabla_x f(x_0, \theta_0) + \sum_{i \in I(x_0,\theta_0)} \lambda_i \nabla_x g_i(x_0, \theta_0) = 0$
2. $g_i(x_0, \theta_0) = 0$ for all $i \in I(x_0, \theta_0)$

Define the function:

$$F(\theta, x, \lambda) = \begin{pmatrix} \nabla_x f(x, \theta) + \sum_{i=1}^n \lambda_i \nabla_x g_i(x, \theta) \\ g_i(x, \theta) \quad (i \in I(x_0, \theta_0)) \end{pmatrix} = 0 \quad (2)$$

16.2.2 The Implicit Function Theorem

Theorem 1 (Existence of Differentiable Solutions): Suppose $f, g_1, \ldots, g_n$ are C^2. If the Jacobian with respect to (x, λ) is invertible at $(\theta_0, x_0, \lambda_0)$:

$$\det \begin{pmatrix} \nabla^2_{xx} f(x_0, \theta_0) + \sum \lambda_i \nabla^2_{xx} g_i(x_0, \theta_0) & \nabla_x g^*(x_0, \theta_0)^T \\ \nabla_x g^*(x_0, \theta_0) & 0 \end{pmatrix} \neq 0 \quad (3)$$

then there exist differentiable functions $\tilde{x} : \Theta_0 \to \mathbb{R}^m$ and $\tilde{\lambda} : \Theta_0 \to \mathbb{R}^n$ such that for all $\theta \in \Theta_0$:

$$F(\theta, \tilde{x}(\theta), \tilde{\lambda}(\theta)) = 0 \quad (4)$$

16.3 The Envelope Theorem

16.3.1 Formal Statement

Theorem 2 (Envelope Theorem): Suppose there exist differentiable functions $\tilde{x} : \Theta_0 \to \mathbb{R}^m$ and $\tilde{\lambda} : \Theta_0 \to \mathbb{R}^n$ such that for all $\theta \in \Theta_0$, $\tilde{x}(\theta)$ solves the maximization problem and $(\tilde{x}(\theta), \tilde{\lambda}(\theta))$ satisfy the KT conditions. Then v is differentiable and:

$$\boxed{\nabla v(\theta) = \nabla_\theta L(\tilde{x}(\theta), \tilde{\lambda}(\theta), \theta)} \quad (5)$$

where the Lagrangian is:

$$L(x, \lambda, \theta) = f(x, \theta) + \sum_{i=1}^n \lambda_i g_i(x, \theta) \quad (6)$$

16.3.2 Proof

Proof:

1. Since $g_i(\tilde{x}(\theta), \theta) = 0$ for active constraints and $\tilde{x}(\theta)$ is optimal, we have:

$$v(\theta) = f(\tilde{x}(\theta), \theta) = f(\tilde{x}(\theta), \theta) + \sum_{i=1}^n \tilde{\lambda}_i(\theta) g_i(\tilde{x}(\theta), \theta)$$

2. Differentiate with respect to θ:

$$\nabla v(\theta) = \nabla_x f(\tilde{x}, \theta) \cdot D\tilde{x} + \nabla_\theta f(\tilde{x}, \theta) + \sum_i \tilde{\lambda}_i \left[\nabla_x g_i(\tilde{x}, \theta) \cdot D\tilde{x} + \nabla_\theta g_i(\tilde{x}, \theta)\right] + \sum_i g_i(\tilde{x}, \theta) \cdot D\tilde{\lambda}_i$$

3. The last term is zero because $g_i(\tilde{x}, \theta) = 0$ for active constraints (and $\lambda_i = 0$ for inactive constraints).
4. Rearranging:

$$\nabla v(\theta) = \left[\nabla_x f(\tilde{x}, \theta) + \sum_i \tilde{\lambda}_i \nabla_x g_i(\tilde{x}, \theta)\right] D\tilde{x} + \nabla_\theta f(\tilde{x}, \theta) + \sum_i \tilde{\lambda}_i \nabla_\theta g_i(\tilde{x}, \theta)$$

5. The term in brackets is zero by the first-order KT condition.
6. Therefore:

$$\nabla v(\theta) = \nabla_\theta f(\tilde{x}, \theta) + \sum_{i=1}^{n} \tilde{\lambda}_i(\theta) \nabla_\theta g_i(\tilde{x}, \theta)$$

7. But this is exactly $\nabla_\theta L(\tilde{x}(\theta), \tilde{\lambda}(\theta), \theta)$.

□

16.4 Interpretation and Applications

16.4.1 Economic Interpretation

The Envelope Theorem has a powerful economic interpretation:

- $\nabla_\theta f(\tilde{x}, \theta)$ is the direct effect of the parameter on the objective, holding the optimal choice fixed.
- $\sum \tilde{\lambda}_i \nabla_\theta g_i(\tilde{x}, \theta)$ is the indirect effect through changes in the constraints.
- The theorem says that the total effect (through changes in x) is zero at the optimum because the first-order conditions eliminate it.
- In economic terms: the derivative of the value function equals the derivative of the Lagrangian with respect to the parameter, evaluated at the optimal point.

16.4.2 Shadow Prices

The multipliers $\tilde{\lambda}_i(\theta)$ can be interpreted as **shadow prices**: they measure the marginal value of relaxing a constraint. By the Envelope Theorem:

$$\frac{\partial v}{\partial b_i} = \lambda_i \quad (7)$$

where b_i is the right-hand side of constraint i.

16.5 Summary of Key Results

Concept	Definition	Key Formula
Value Function	Optimal value as function of parameters	$v(\theta) = \max_x f(x, \theta)$ s.t. $g(x, \theta) \geq 0$
Lagrangian	Objective plus constraints	$L(x, \lambda, \theta) = f(x, \theta) + \sum \lambda_i g_i(x, \theta)$
Envelope Theorem	Derivative of value function	$\nabla v(\theta) = \nabla_\theta L(\tilde{x}(\theta), \tilde{\lambda}(\theta), \theta)$
Direct Effect	Holding x fixed	$\nabla_\theta f(\tilde{x}, \theta)$

Concept	Definition	Key Formula
Indirect Effect	Through constraints	$\sum \tilde{\lambda}_i \nabla_\theta g_i(\tilde{x}, \theta)$
Shadow Price	Marginal value of relaxing constraint	$\lambda_i = \frac{\partial v}{\partial b_i}$

Chapter 17: Riemann Integrals

17.1 Definition and Basic Properties

17.1.1 Motivation

The problem that the Riemann integral solves is the **calculation of the area under a curve**. Given a function $f : [a, b] \to \mathbb{R}$, we want to define the "total accumulated value" of f over the interval. The Riemann integral formalizes this by approximating the area using sums of rectangles and taking a limit as the partitions become finer.

Geometric Interpretation: Visualize the graph of f on $[a, b]$. The integral $\int_a^b f(x)dx$ represents the signed area between the curve and the x-axis. The lower sum approximates the area from below using rectangles whose heights are the infima on each subinterval; the upper sum approximates from above using suprema.

17.1.2 Formal Definitions

Definition 1 (Partition): A **partition** of $[a, b]$ is a finite set:

$$P = \{a = t_0 < t_1 < \cdots < t_n = b\} \quad (1)$$

Definition 2 (Upper and Lower Sums): For a bounded function $f : [a, b] \to \mathbb{R}$ and a partition P, define:

$$m_i = \inf_{x \in [t_i, t_{i+1}]} f(x), \qquad M_i = \sup_{x \in [t_i, t_{i+1}]} f(x) \quad (2)$$

The **lower sum** and **upper sum** are:

$$\boxed{s(f, P) = \sum_{i=0}^{n-1} m_i (t_{i+1} - t_i)} \quad (3)$$

$$\boxed{S(f, P) = \sum_{i=0}^{n-1} M_i (t_{i+1} - t_i)} \quad (4)$$

Definition 3 (Lower and Upper Integrals): The **lower integral** and **upper integral** are:

$$\boxed{\underline{\int_a^b} f(x)dx = \sup_P s(f, P)} \quad (5)$$

$$\boxed{\overline{\int_a^b} f(x)dx = \inf_P S(f, P)} \quad (6)$$

Definition 4 (Riemann Integrable): A bounded function f is **Riemann integrable** on $[a, b]$ if:

$$\boxed{\underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx} \quad (7)$$

The common value is denoted $\int_a^b f(x)dx$.

17.1.3 Key Theorems

Theorem 1 (Linearity of the Integral): If f and g are integrable on $[a, b]$, then for all $\alpha, \beta \in \mathbb{R}$:

$$\boxed{\int_a^b (\alpha f + \beta g)(x)dx = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx} \quad (8)$$

Theorem 2 (Monotonicity of the Integral): If f and g are integrable and $f(x) \leq g(x)$ for all $x \in [a, b]$, then:

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx \quad (9)$$

Theorem 3 (Additivity over Intervals): If f is integrable on $[a, b]$ and $c \in (a, b)$, then:

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \quad (10)$$

17.2 The Fundamental Theorem of Calculus

17.2.1 Motivation

The Fundamental Theorem of Calculus (FTC) establishes the deep connection between differentiation and integration. It shows that integration and differentiation are inverse operations. This theorem is the foundation for evaluating definite integrals using antiderivatives.

17.2.2 Key Theorems

Theorem 4 (FTC - Part 1): Let $f : [a, b] \to \mathbb{R}$ be integrable. Define:

$$F(x) = \int_a^x f(t)dt, \qquad x \in [a, b] \quad (11)$$

If f is continuous at $x_0 \in [a, b]$, then F is differentiable at x_0 and:

$$\boxed{F'(x_0) = f(x_0)} \quad (12)$$

Proof:

1. For $h > 0$:
$$\frac{F(x_0 + h) - F(x_0)}{h} = \frac{1}{h} \int_{x_0}^{x_0+h} f(t)dt$$
2. Since f is continuous at x_0, for any $\epsilon > 0$, there exists $\delta > 0$ such that $|f(t) - f(x_0)| < \epsilon$ whenever $|t - x_0| < \delta$.
3. For $|h| < \delta$:
$$\left| \frac{1}{h} \int_{x_0}^{x_0+h} f(t)dt - f(x_0) \right| = \left| \frac{1}{h} \int_{x_0}^{x_0+h} (f(t) - f(x_0))dt \right| \leq \epsilon$$
4. Therefore, $\lim_{h \to 0} \frac{F(x_0+h)-F(x_0)}{h} = f(x_0)$.

□

Theorem 5 (FTC - Part 2): If F is differentiable on $[a, b]$ and F' is integrable on $[a, b]$, then:

$$\boxed{\int_a^b F'(x)dx = F(b) - F(a)} \quad (13)$$

Theorem 6 (Integration by Substitution): If $g : [a, b] \to \mathbb{R}$ is a continuously differentiable bijection, then:

$$\boxed{\int_{g(a)}^{g(b)} f(u)du = \int_a^b f(g(x))g'(x)dx} \quad (14)$$

17.3 Summary of Key Results

Concept	Definition	Key Formula
Partition	Division of interval into subintervals	$P = \{a = t_0 < t_1 < \cdots < t_n = b\}$
Lower Sum	Sum of infima times interval lengths	$s(f, P) = \sum m_i(t_{i+1} - t_i)$
Upper Sum	Sum of suprema times interval lengths	$S(f, P) = \sum M_i(t_{i+1} - t_i)$
Lower Integral	Supremum of lower sums	$\underline{\int_a^b} f(x)dx = \sup_P s(f, P)$
Upper Integral	Infimum of upper sums	$\overline{\int_a^b} f(x)dx = \inf_P S(f, P)$
Riemann Integrable	Lower integral equals upper integral	$\underline{\int} f = \overline{\int} f$
Linearity	Integral of linear combination	$\int(\alpha f + \beta g) = \alpha \int f + \beta \int g$
Monotonicity	$f \leq g \implies \int f \leq \int g$	Order preserved by integration
FTC Part 1	Derivative of indefinite integral	$F'(x) = f(x)$ where $F(x) = \int_a^x f(t)dt$
FTC Part 2	Definite integral from antiderivative	$\int_a^b F'(x)dx = F(b) - F(a)$

Chapter 18: Topics in Mathematical Economics

18.1 Correspondence Theory

18.1.1 Motivation

Correspondences (also called set-valued maps) are fundamental in mathematical economics. They arise naturally in optimization problems where the optimal choice may not be unique, in game theory where best responses are sets, and in general equilibrium theory where demand functions are set-valued. Understanding correspondences is essential for proving existence results in economics.

Geometric Interpretation: A correspondence $\Gamma : X \rightrightarrows Y$ maps each point $x \in X$ to a subset $\Gamma(x) \subset Y$. Think of it as a "fuzzy" function where each input can map to multiple outputs. The graph of a correspondence is a set in $X \times Y$ that may have vertical slices that are not single points.

18.1.2 Formal Definitions

Definition 1 (Correspondence): A **correspondence** $\Gamma : X \rightrightarrows Y$ associates each $x \in X$ with a subset $\Gamma(x) \subset Y$.

Definition 2 (Optimization Problem with Correspondences): Consider the problem:

$$(P): \quad V(x) = \max_{y \in \Gamma(x)} f(x, y) \quad (1)$$

The **solution correspondence** is:

$$\Gamma^*(x) = \{y \in \Gamma(x) : f(x, y) = V(x)\} \quad (2)$$

18.2 Hemicontinuity

18.2.1 Motivation

Hemicontinuity is the generalization of continuity to correspondences. Upper hemicontinuity (u.h.c.) ensures that the correspondence does not "explode" or have sudden jumps. Lower hemicontinuity (l.h.c.) ensures that the correspondence does not have sudden gaps. Together, they are the set-valued analogue of continuity.

Geometric Interpretation: A correspondence is u.h.c. if its graph is closed in the appropriate sense: if $x_k \to x$ and $y_k \in \Gamma(x_k)$ with $y_k \to y$, then $y \in \Gamma(x)$. It is l.h.c. if every point in $\Gamma(x)$ can be approached from nearby values: for every $y \in \Gamma(x)$ and $x_k \to x$, there exist $y_k \in \Gamma(x_k)$ with $y_k \to y$.

18.2.2 Formal Definitions

Definition 3 (Upper Hemicontinuity): A correspondence $\Gamma : X \rightrightarrows Y$ is **upper hemicontinuous (u.h.c.)** at $x \in X$ if:

$$\forall (x_k) \in X,\ x_k \to x,\ \forall (y_k) \in Y,\ y_k \in \Gamma(x_k),\ \exists \text{ subsequence } y_{k_j} \to y \in \Gamma(x) \quad (3)$$

Definition 4 (Lower Hemicontinuity): A correspondence $\Gamma : X \rightrightarrows Y$ is **lower hemicontinuous (l.h.c.)** at $x \in X$ if:

$$\forall y \in \Gamma(x),\ \forall (x_k) \in X,\ x_k \to x,\ \exists (y_k) \in Y \text{ with } y_k \to y \text{ and } y_k \in \Gamma(x_k) \quad (4)$$

Definition 5 (Continuous Correspondence): A correspondence is **continuous** if it is both u.h.c. and l.h.c.

18.3 The Maximum Theorem (Berge's Theorem)

18.3.1 Motivation

The Maximum Theorem is one of the most important results in mathematical economics. It states that the value function of a parametrized optimization problem is continuous and the solution correspondence is u.h.c. when the objective function is continuous and the constraint correspondence is continuous with compact values. This theorem is essential for proving the existence of equilibria in economic models.

18.3.2 Key Theorems

Theorem 1 (Upper Semicontinuity of the Value Function): If f is upper semicontinuous (u.s.c.) and Γ is u.h.c. with nonempty compact values, then the value function $V(x)$ is u.s.c.

Proof:

1. Let $x \in X$ and $x_k \to x$. We must show that $\limsup V(x_k) \le V(x)$.
2. Since $\Gamma(x_k)$ is compact and nonempty, and $f(x_k, \cdot)$ is u.s.c., there exists $y_k \in \Gamma(x_k)$ such that $V(x_k) = f(x_k, y_k)$.

3. By u.h.c., there exists a subsequence $y_{k_j} \to y \in \Gamma(x)$.
4. Then:
$$V(x) \geq f(x, y) \geq \limsup f(x_{k_j}, y_{k_j}) = \limsup V(x_{k_j})$$
(by u.s.c. of f).
5. Therefore, V is u.s.c.

□

Theorem 2 (Lower Semicontinuity of the Value Function): If f is continuous and Γ is l.h.c. with nonempty compact values, then the value function $V(x)$ is l.s.c.

Proof:

1. Let $x \in X$ and $x_k \to x$. We must show that $\liminf V(x_k) \geq V(x)$.
2. Choose $y \in \Gamma(x)$ such that $V(x) = f(x, y)$ (exists since f is continuous on compact $\Gamma(x)$).
3. By l.h.c., there exists $y_k \in \Gamma(x_k)$ with $y_k \to y$.
4. Thus, $V(x_k) \geq f(x_k, y_k)$ for each k.
5. By continuity of f:
$$\liminf V(x_k) \geq f(x, y) = V(x)$$
6. Therefore, V is l.s.c.

□

Theorem 3 (Berge's Maximum Theorem): If f is continuous and Γ is continuous (u.h.c. + l.h.c.) with nonempty compact values, then: 1. The value function V is continuous. 2. The solution correspondence Γ^* is u.h.c.

Proof of Upper Hemicontinuity of Γ^*:

1. Let $y_k \in \Gamma^*(x_k)$ (i.e., $y_k \in \Gamma(x_k)$ and $V(x_k) = f(x_k, y_k)$) and $x_k \to x$.
2. Since $y_k \in \Gamma(x_k)$ and Γ is u.h.c., there exists a subsequence $y_{k_j} \to y \in \Gamma(x)$.
3. Choose any $y' \in \Gamma(x)$. We must show that $f(x, y) \geq f(x, y')$.
4. Since Γ is l.h.c., there exists $y'_k \to y'$ with $y'_k \in \Gamma(x_k)$.
5. By definition of solution: $f(x_k, y_k) \geq f(x_k, y'_k)$ for all k.
6. By continuity of f, taking the limit gives $f(x, y) \geq f(x, y')$.
7. Therefore, $y \in \Gamma^*(x)$, so Γ^* is u.h.c.

□

18.4 Uniqueness of Solutions

18.4.1 Key Theorems

Definition 6 (Strict Quasiconcavity): A function f is **strictly quasiconcave** if for all $x, y \in C$, $x \neq y$, and all $\alpha \in (0, 1)$:
$$f(\alpha x + (1 - \alpha)y) > \min\{f(x), f(y)\} \quad (5)$$

Theorem 4 (Uniqueness under Strict Quasiconcavity): If $\Gamma(x)$ is convex and $y \mapsto f(x, y)$ is strictly quasiconcave, then $\Gamma^*(x)$ is either empty or contains exactly one point.

Proof:

1. Suppose there exist $y_1, y_2 \in \Gamma^*(x)$ with $y_1 \neq y_2$. Then:

$$f(x, y_1) = f(x, y_2) = V(x)$$

2. Let $y_\alpha = \alpha y_1 + (1-\alpha)y_2$. Since $\Gamma(x)$ is convex, $y_\alpha \in \Gamma(x)$.
3. By strict quasiconcavity:

$$f(x, y_\alpha) > \min\{f(x, y_1), f(x, y_2)\} = V(x)$$

4. This contradicts the definition of $V(x)$ as the maximum value.
5. Therefore, $\Gamma^*(x)$ is single-valued.

□

18.5 Fixed Point Theorems

18.5.1 Motivation

Fixed point theorems establish conditions under which a function (or correspondence) has a point that maps to itself. These theorems are fundamental in economics for proving the existence of equilibria. The three most important fixed point theorems are Banach's (contractions), Brouwer's (continuous functions on compact convex sets), and Kakutani's (correspondences).

18.5.2 Key Theorems

Theorem 5 (Banach Fixed Point Theorem): Let (X, d) be a complete metric space and $f : X \to X$ be a **contraction**, i.e., there exists $\lambda \in [0, 1)$ such that:

$$d(f(x), f(y)) \leq \lambda d(x, y) \quad \forall x, y \in X \quad (6)$$

Then there exists a unique fixed point $x^* \in X$ such that $f(x^*) = x^*$. Moreover, for any $x_0 \in X$, the sequence $x_{k+1} = f(x_k)$ converges to x^*.

Proof Sketch:

1. Choose $x_0 \in X$. Define $x_{k+1} = f(x_k)$.
2. Show (x_k) is Cauchy:

$$d(x_{k+1}, x_k) = d(f(x_k), f(x_{k-1})) \leq \lambda d(x_k, x_{k-1})$$

3. By induction: $d(x_{k+1}, x_k) \leq \lambda^k d(x_1, x_0)$.
4. For $m > k$:

$$d(x_m, x_k) \leq \sum_{j=k}^{m-1} d(x_{j+1}, x_j) \leq \sum_{j=k}^{\infty} \lambda^j d(x_1, x_0) = \frac{\lambda^k}{1-\lambda} d(x_1, x_0) \to 0$$

5. Thus, (x_k) converges to some $x^* \in X$.
6. By continuity of f: $f(x^*) = \lim f(x_k) = \lim x_{k+1} = x^*$.
7. Uniqueness: if x^* and y^* are fixed points, then:

$$d(x^*, y^*) = d(f(x^*), f(y^*)) \leq \lambda d(x^*, y^*)$$

   Since $\lambda < 1$, this implies $d(x^*, y^*) = 0$, so $x^* = y^*$.

□

Theorem 6 (Brouwer Fixed Point Theorem): Let $K \subset \mathbb{R}^n$ be a nonempty, compact, convex set. If $f : K \to K$ is continuous, then there exists $x^* \in K$ such that:

$$f(x^*) = x^* \quad (7)$$

Theorem 7 (Kakutani Fixed Point Theorem): Let $K \subset \mathbb{R}^n$ be a nonempty, compact, convex set. If $\Gamma : K \rightrightarrows K$ is a correspondence such that: 1. Γ is u.h.c. 2. $\Gamma(x)$ is nonempty, compact, and convex for all $x \in K$.

Then there exists $x^* \in K$ such that:

$$x^* \in \Gamma(x^*) \quad (8)$$

18.6 Summary of Key Results

Concept	Definition	Key Formula
Correspondence	Set-valued map	$\Gamma : X \rightrightarrows Y$
Upper Hemicontinuity	Limits stay in graph	$x_k \to x, y_k \in \Gamma(x_k), y_k \to y \implies y \in \Gamma(x)$
Lower Hemicontinuity	Every point can be approached	$\forall y \in \Gamma(x), \exists y_k \in \Gamma(x_k)$ with $y_k \to y$
Value Function	Maximum value of optimization	$V(x) = \max_{y \in \Gamma(x)} f(x, y)$
Solution Correspondence	Set of maximizers	$\Gamma^*(x) = \{y \in \Gamma(x) : f(x, y) = V(x)\}$
Berge's Maximum Theorem	Continuity of value and solution	V continuous, Γ^* u.h.c.
Strict Quasiconcavity	Strictly increasing along chords	$f(\alpha x + (1-\alpha)y) > \min\{f(x), f(y)\}$
Banach Fixed Point	Unique fixed point of contraction	$d(f(x), f(y)) \leq \lambda d(x, y)$, $\lambda < 1$
Brouwer Fixed Point	Fixed point of continuous map on compact convex set	$f : K \to K$ continuous $\implies \exists x^* : f(x^*) = x^*$
Kakutani Fixed Point	Fixed point of u.h.c. convex-valued correspondence	$\Gamma : K \rightrightarrows K$ u.h.c. $\implies \exists x^* \in \Gamma(x^*)$
