POST-GRADUATE DEGREE PROGRAMME (CBCS)

M.SC. IN MATHEMATICS

SEMESTER-III

Paper Code: COR 3.2
(Pure and Applied Streams)

Calculus of $\mathbb{R}^n$

Fuzzy Set Theory

Calculus of Variations

Self-Learning Material

DIRECTORATE OF OPEN AND DISTANCE LEARNING
UNIVERSITY OF KALYANI
Kalyani, Nadia
West Bengal

Course Content Writers	
Course Name	**Writer**
Block I: Calculus of $\mathbb{R}^n$	Ms. Audrija Choudhury Assistant Professor Department of Mathematics Directorate of Open and Distance Learning University of Kalyani
Block II: Fuzzy Set Theory	Ms. Audrija Choudhury Assistant Professor Department of Mathematics Directorate of Open and Distance Learning University of Kalyani
Block III: Calculus of Variations	Dr. Tanchar Molla Assistant Professor Department of Mathematics Dumkal College

August, 2023

Directorate of Open and Distance Learning, University of Kalyani

Published by the Directorate of Open and Distance Learning

University of Kalyani, 741235, West Bengal

Printed by East India Photo Composing Centre, 209A, Bidhan Sarani, Kolkata-700006

Director's Message

Satisfying the varied needs of distance learners, overcoming the obstacle of Distance and reaching the unreached students are the three fold functions catered by Open and Distance Learning (ODL) systems. The onus lies on writers, editors, production professionals and other personnel involved in the process to overcome the challenges inherent to curriculum design and production of relevant Self Learning Materials (SLMs). At the University of Kalyani a dedicated team under the able guidance of the Hon'ble Vice-Chancellorhas invested its best efforts, professionally and in keeping with the demands of Post Graduate CBCS Programmes in Distance Mode to devise a self-sufficient curriculum for each course offered by the Directorate of Open and Distance Learning (DODL), University of Kalyani.

Development of printed SLMs for students admitted to the DODL within a limited time to cater to the academic requirements of the Course as per standards set by Distance Education Bureau of the University Grants Commission, New Delhi, India under Open and Distance Mode UGC Regulations, 2020 had been our endeavor. We are happy to have achieved our goal.

Utmost care and precision have been ensured in the development of the SLMs, making them useful to the learners, besides avoiding errors as far as practicable. Further suggestions from the stakeholders in this would be welcome.

During the production-process of the SLMs, the team continuously received positive stimulations and feedback from Professor **(Dr.) Amalendu Bhunia, Hon'ble Vice-Chancellor, University of Kalyani**, who kindly accorded directions, encouragements and suggestions, offered constructive criticism to develop it with in proper requirements. We gracefully, acknowledge his inspiration and guidance.

Sincere gratitude is due to the respective chairpersons as well as each and every member of PGBOS (DODL), University of Kalyani. Heartfelt thanks are also due to the Course Writers-faculty members at the DODL, subject-experts serving at University Post Graduate departments and also to the authors and academicians whose academic contributions have enriched the SLMs. We humbly acknowledge their valuable academic contributions. I wouldespecially like to convey gratitude to all other University dignitaries and personnel involved either at the conceptual or operational level of the DODL of University of Kalyani.

Their persistent and coordinated efforts have resulted in the compilation of comprehensive, learner-friendly, flexible texts that meet the curriculum requirements of the Post Graduate Programme through Distance Mode.

All the Self Learning Materials are self writing and collected from e-book, journals and websites.

Professor (Dr.) Sanjib Kumar Datta
Director
Directorate of Open and Distance Learning
University of Kalyani

Post Graduate Board of Studies (PGBOS) Members of Department of Mathematics, Directorate of Open and Distance Learning (DODL), University of Kalyani.

Sl. No.	Name & Designation	Role
1	Prof. (Dr.) Samares Pal, Professor & Head, Department of Mathematics, University of Kalyani.	Chairperson
2	Prof. (Dr.) Pulak Sahoo, Professor, Department of Mathematics, University of Kalyani.	Member
3	Dr. Sahidul Islam, Associate Professor, Department of Mathematics, University of Kalyani.	Member
4	Prof. (Dr.) Sushanta Kumar Mohanta, Professor, Department of Mathematics, West Bengal State University.	External Nominated Member
5	Ms. Audrija Choudhury, Assistant Professor, Department of Mathematics, DODL, University of Kalyani.	Member
6	Prof. (Dr.) Sanjib Kumar Datta, Director, DODL, University of Kalyani.	Convener

Core Paper

PURE & APPLIED STREAMS

COR 3.2
Marks : 100 (SEE : 80; IA : 20); Credit : 6

Calculus of $\mathbb{R}^n$ (Marks : 50 (SEE: 40; IA: 10))
Fuzzy Set Theory (Marks : 25 (SEE: 20; IA: 05))
Calculus of Variations (Marks : 25 (SEE: 20; IA: 05))

Syllabus

Block I

- **Unit 1:** Differentiation on $\mathbb{R}^n$: Directional derivatives and continuity, the total derivative and continuity.
- **Unit 2:** Total derivative in terms of partial derivatives, the matrix transformation of $T : \mathbb{R}^n \to \mathbb{R}^n$ Jacobian matrix.
- **Unit 3:** The chain rule and its matrix form. Mean value theorem for vector valued function. Mean value inequality.
- **Unit 4:** A sufficient condition for differentiability. A sufficient condition for mixed partial derivatives.
- **Unit 5:** Functions with non-zero Jacobian determinant, the inverse function theorem, the implicit function theorem as an application of Inverse function theorem.
- **Unit 6:** Extremum problems with side conditions – Lagrange's necessary conditions as an application of Inverse function theorem.
- **Unit 7:** Integration on $\mathbb{R}^n$: Integral of $f : A \to \mathbb{R}$ when $A \subset \mathbb{R}^n$ closed rectangle.
- **Unit 8:** Conditions of inerrability. Integrals of $f : C \to \mathbb{R},\ C \subset \mathbb{R}^n$ is not a rectangle, concept of Jordan measurability of a set in $\mathbb{R}$.
- **Unit 9:** Fubini's theorem for integral of $f : A \times B \to \mathbb{R},\ A \subset \mathbb{R}^n,\ B \subset \mathbb{R}^n$ are closed rectangles.
- **Unit 10:** Fubini's theorem for $f : C \to \mathbb{R},\ C \subset A \times B$, Formula for change of variables in an integral in $\mathbb{R}^n$.

Block II

- **Unit 11:** Interval Arithmetic: Interval numbers, arithmetic operations on interval numbers, distance between intervals, two level interval numbers.
- **Unit 12:** Basic concepts of fuzzy sets: Types of fuzzy sets, α-cuts and its properties, representations of fuzzy sets, support, convexity, normality, cardinality, standard set-theoretic operations on fuzzysets
- **Unit 13:** Decomposition theorems, Zadeh's extension principle.
- **Unit 14:** Fuzzy Relations: Crisp versus fuzzy relations, fuzzy matrices and fuzzy graphs, composition of fuzzy relations, relational join, binary fuzzy relations.
- **Unit 15:** Fuzzy Arithmetic: Fuzzy numbers, arithmetic operations on fuzzy numbers (multiplication and division on $\mathbb{R}^+$ only), fuzzy equations.

Block III

- **Unit 16:** Variational Problems with fixed Boundaries: Variation, Linear functional, Euler-Lagrange equation, Functionals dependent on higher order derivatives, Functionals dependent on functions of several variables.
- **Unit 17:** Applications of Calculus of variations on the problems of shortest distance, minimum surface of revolution, Brachistochrone problem, geodesic etc. Isoperimetric problem.
- **Unit 18:** Variational Problems with Moving Boundaries: Transversality conditions, Orthogonality conditions, Functional dependent on two functions, One sided variations.
- **Unit 19:** Sufficient Conditions for an Extremum: Proper field, Central field, Field of extremals, Embedding in a field of extremals and in a central field.
- **Unit 20:** Sufficient condition for extremum-Weirstrass condition, Legendre condition. Weak and strong extremum.

Contents

Unit 1

Course Structure

- Differentiation on $\mathbb{R}^n$: Directional derivatives and continuity, the total derivative and continuity.

1.1 Introduction

You are familiar with the concepts of one-variable calculus, that is, functions of the form $f(x)$ for one real variable x, the codomain also being $\mathbb{R}$. When we extend this from one variable to more than one variables, it is called the multivariable calculus. Basically, the most general multivariable function is of the form $f : A \to B$, where $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$. We shall first recapitulate the basic definitions of the Euclidean space $\mathbb{R}^n$.

Objectives

After reading this unit, you will be able to

- define basic terms related to the Euclidean plane space $\mathbb{R}^n$
- define directional and partial derivatives of multivariable functions
- define the differentiation of multivariable functions and discuss its implications

1.2 Preliminaries

Euclidean n-space $\mathbb{R}^n$ is defined as the set of all n-tuples $(x_1, \ldots, x_n)$ of real numbers x_i. An element of $\mathbb{R}^n$ is often called a point in $\mathbb{R}^n$, and $\mathbb{R}^1$, $\mathbb{R}^2$, $\mathbb{R}^3$ are often called the line, the plane, and space, respectively. If x denotes an element of $\mathbb{R}^n$, then x is an n-tuple of numbers, the ith component of which is denoted by x_i; thus we can write

$$x = (x_1, \ldots, x_n)$$

A point in $\mathbb{R}^n$ is also called a vector in $\mathbb{R}^n$, because $\mathbb{R}^n$, with $x + y = (x_1 + y_1, \ldots, x_n + y_n)$ and $ax = (ax_1, \ldots, ax_n)$, as operations, is a vector space (over the real numbers, of dimension n). In this vector space there is the notion of the length of a vector x, usually called the norm $|x|$ of x and is defined by $|x| = \sqrt{(x_1)^2 + \cdots + (x_n)^2}$. If $n = 1$, then $|x|$ is the usual absolute value of x. The relation between the norm and the vector space structure of $\mathbb{R}^n$ is very important. Let us list a few properties of the norm function.

Theorem 1.2.1. If $x, y \in \mathbb{R}^n$ and $a \in \mathbb{R}$, then

1. $|x| \geq 0$, and $|x| = 0$ if and only if $x = 0$.
2. $\left|\sum_{i=1}^{n} x_i y_i\right| \leq |x| \cdot |y|$, equality holds if and only if x and y are linearly dependent.
3. $|x + y| \leq |x| + |y|$.
4. $|ax| = |a| \cdot |x|$.

The quantity $\left|\sum_{i=1}^{n} x_i y_i\right|$ defines the inner product of x and y in $\mathbb{R}^n$ and is denoted by $\langle x, y \rangle$. The vector $(0, 0, \ldots, 0)$, usually denoted by 0 is the zero vector of $\mathbb{R}^n$ and $\{e_1, e_2, \ldots, e_n\}$ is its usual basis, where e_i is the vector whose all components are zero except for the ith component, which is 1.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, the matrix of T with respect to the usual bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is the $m \times n$ matrix $A = (a_{ij})$, where $T(e_i) = \sum_{j=1}^{m} a_{ji} e_j$, the coefficients of $T(e_i)$ appear in the ith column of the matrix. If $S : \mathbb{R}^m \to \mathbb{R}^p$ has the $p \times m$ matrix B, then $S \circ T$ has the $p \times n$ matrix BA.

Let us now look into some standard subsets in $\mathbb{R}^n$. The closed interval $[a, b]$ has a natural analogue in $\mathbb{R}^2$. This is the closed rectangle $[a, b] \times [c, d]$, defined as the collection of all pairs (x, y) with $x \in [a, b]$ and $y \in [c, d]$. More generally, if $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$, then $A \times B \subset \mathbb{R}^{m+n}$ is defined as the set of all $(x, y) \in \mathbb{R}^{m+n}$ with $x \in A$ and $y \in B$. In particular, $\mathbb{R}^{m+n} = \mathbb{R}^m \times \mathbb{R}^n$. If $A \subset \mathbb{R}^m, B \subset \mathbb{R}^n$, and $C \subset \mathbb{R}^p$, then $(A \times B) \times C = A \times (B \times C)$, and both of these are denoted simply $A \times B \times C$; this convention is extended to the product of any number of sets. The set $[a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ is called a closed rectangle in $\mathbb{R}^n$, while the set $(a_1, b_1) \times \cdots \times (a_n, b_n) \subset \mathbb{R}^n$ is called an open rectangle. More generally a set $U \subset \mathbb{R}^n$ is called open if for each $x \in U$ there is an open rectangle A such that $x \in A \subset U$. A subset C of $\mathbb{R}^n$ is closed if $\mathbb{R}^n \setminus C$ is open. For example, if C contains only finitely many points, then C is closed.

If $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$, then we have one of the following possibilities:

1. There exists an open rectangle B such that $x \in B \subset A$;
2. There exists an open rectangle B such that $x \in B \subset \mathbb{R}^n \setminus A$;
3. If B is any open rectangle with $x \in B$, then B contains points of both A as well as $\mathbb{R}^n \setminus A$.

In case of 1, x is called an *interior point* of A; in case of 2, it is called an *exterior point* of A and in case of 3, it is a *boundary point* of A. The set of all interior points of A constitute the *interior* of A and all exterior points constitute the *exterior* of A and the boundary points the *boundary* of A. It can be easily seen that the interior of any set A is open, and the same is true for the exterior of A, which is, in fact, the interior of $\mathbb{R}^n \setminus A$. Thus their union is open, and what remains, the boundary, must be closed.

A collection $\mathcal{O}$ of open sets is an open cover of A (or, briefly, covers A) if every point $x \in A$ is in some open set in the collection $\mathcal{O}$. For example if $\mathcal{O}$ is the collection of all open intervals $(a, a + 1)$ for $a \in \mathbb{R}$, then $\mathcal{O}$ is a cover of $\mathbb{R}$. Clearly no finite number of the open sets in $\mathcal{O}$ will cover $\mathbb{R}$ or, for that matter, any unbounded subset of $\mathbb{R}$. A similar situation can also occur for bounded sets. If $\mathcal{O}$ is the collection of all open intervals $(1/n, 1 - 1/n)$ for all integers $n > 1$, then $\mathcal{O}$ is an open cover of $(0, 1)$, but again no finite collection of sets in $\mathcal{O}$ will cover $(0, 1)$. Although this phenomenon may not appear particularly scandalous, sets for which this state of affairs cannot occur are of such importance that they have received a special designation: a set A is called compact if every open cover $\mathcal{O}$ contains a finite subcollection of open sets which also covers A.

A set with only finitely many points is obviously compact and so is the infinite set A which contains 0 and the numbers $1/n$ for all integers n (reason: if $\mathcal{O}$ is a cover, then $0 \in U$ for some open set U in $\mathcal{O}$; there are only finitely many other points of A not in U, each requiring at most one more open set). Recognizing compact sets is greatly simplified by the following results, of which only the first has any depth (i.e., uses any facts about the real numbers).

Theorem 1.2.2. (Heine-Borel). The closed interval $[a, b]$ is compact.

If $B \subset \mathbb{R}^m$ is compact and $x \in \mathbb{R}^n$, it is easy to see that $\{x\} \times B \subset \mathbb{R}^{n+m}$ is compact. The following result however leads to a stronger and more generalised conclusion.

Theorem 1.2.3. If B is compact and $\mathcal{O}$ is an open cover of $\{x\} \times B$, then there is an open set $U \subset \mathbb{R}^n$ containing x such that $U \times B$ is covered by a finite number of sets in $\mathcal{O}$.

Corollary 1.2.4. If $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ are compact, then $A \times B \subset \mathbb{R}^{n+m}$ is compact.

And finally we have the following.

Corollary 1.2.5. If A_i is compact for each $i = 1, \ldots, n$, then $A_1 \times A_2 \times \ldots \times A_n$ is compact.

Corollary 1.2.6. Every closed and bounded subset of $\mathbb{R}^n$ is compact.

The converse is also true.

1.2.1 Functions in $\mathbb{R}^n$

A function from $\mathbb{R}^n$ to $\mathbb{R}^m$ (sometimes called a (vectorvalued) function of n variables) is a rule which associates to each point in $\mathbb{R}^n$ some point in $\mathbb{R}^m$; the point a function f associates to x is denoted $f(x)$. We write $f : \mathbb{R}^n \to \mathbb{R}^m$ to indicate that $f(x) \in \mathbb{R}^m$ is defined for $x \in \mathbb{R}^n$. The notation $f : A \to \mathbb{R}^m$ indicates that $f(x)$ is defined only for x in the set A, which is called the domain of f. If $B \subset A$, we define $f(B)$ as the set of all $f(x)$ for $x \in B$, and if $C \subset \mathbb{R}^m$ we define $f^{-1}(C) = \{x \in A : f(x) \in C\}$. The notation $f : A \to B$ indicates that $f(A) \subset B$.

If $f, g : \mathbb{R}^n \to \mathbb{R}$, the functions $f + g$, $f - g$, $f \cdot g$, and f/g are defined precisely as in the one-variable case. If $f : A \to \mathbb{R}^m$ and $g : B \to \mathbb{R}^p$, where $B \subset \mathbb{R}^m$, then the composition $g \circ f$ is defined by $g \circ f(x) = g(f(x))$; the domain of $g \circ f$ is $A \cap f^{-1}(B)$. If $f : A \to \mathbb{R}^m$ is $1 - 1$, that is, if $f(x) \neq f(y)$ when $x \neq y$, we define $f^{-1} : f(A) \to \mathbb{R}^n$ by the requirement that $f^{-1}(z)$ is the unique $x \in A$ with $f(x) = z$.

A function $f : A \to \mathbb{R}^m$ determines m component functions $f^1, \ldots f^m : A \to \mathbb{R}$ by $f(x) = \big(f^1(x), \ldots f^m(x)\big)$. If conversely, m functions $g_1, \ldots g_m : A \to \mathbb{R}$ are given, there is a unique function $f : A \to \mathbb{R}^m$ such that $f^i = g_i$, namely $f(x) = (g_1(x), \ldots, g_m(x))$. This function f will be denoted $(g_1, \ldots, g_m)$, so that we always have $f = \big(f^1, \ldots, f^m\big)$. Thus, the characteristics of these component functions f^i greatly determine the characteristics of f. If $\boldsymbol{\pi} : \mathbb{R}^n \to \mathbb{R}^n$ is the identity function, $\pi(x) = x$, then $\pi^i(x) = x^i$; the function π^i is called the i th projection function.

We are familiar with the notation $\lim\limits_{x \to a} f(x) = l$. This means that the value of the function f goes arbitrarily close to l as x goes arbitrarily close to a. In technical terms, we can say that for any $\epsilon > 0$, we can find a $\delta > 0$ such that $|f(x) - l| < \epsilon$ whenever $0 < |x - a| < \delta$. We can define the limit of a vectorvalued function similarly.

Definition 1.2.7. Let $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$ be a function. Let $a \in \mathbb{R}^n$ and $l \in \mathbb{R}^m$. Then we say that f has limit l as x tends to a, written as

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

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

$$|f(x) - l| < \epsilon \text{ whenever } 0 < |x - a| < \delta.$$

Example 1.2.8. Prove that

$$\lim_{(x,y)\to(0,0)} \frac{2x^2y}{x^2+y^2} = 0.$$

Solution. For any x and y, we know that $(x^2+y^2) \geq x^2$(or y^2). Thus,

$$\left|\frac{2x^2y}{x^2+y^2}\right| \leq \left|\frac{2x^2y}{x^2}\right| = |2y|.$$

Let $\epsilon > 0$. Choose $\delta = \frac{\epsilon}{2}$. Then

$$0 < |(x,y)-(0,0)| = \sqrt{x^2+y^2} < \delta$$

implies $|y| < \delta$. Thus, with this choice of δ, we have

$$\left|\frac{2x^2y}{x^2+y^2} - 0\right| < \epsilon.$$

Hence proved. ■

The sum, multiplication and quotient properties of limits are valid in multivariable case too. Let us quickly define the notion of continuity in the multivariable case.

Definition 1.2.9. A function $f : A \to \mathbb{R}^m$ is called continuous at $a \in A$ if $\lim_{x\to a} f(x) = f(a)$. If f is continuous at each point of A, then it is said to be continuous on A, or simply continuous.

From the definition of continuity, it is clear that continuity implies the existence of limit. However, the converse may not be true always. The following is a necessary and sufficient condition for continuity of f.

Theorem 1.2.10. If $A \subset \mathbb{R}^n$, a function $f : A \to \mathbb{R}^m$ is continuous if and only if for every open set $U \subset \mathbb{R}^m$, there is some open set $V \subset \mathbb{R}^n$ such that $f^{-1}(U) = V \cap A$.

Proof. Left as an exercise. □

Theorem 1.2.11. If $f : A \to \mathbb{R}^m$ is called continuous, where $A \subset \mathbb{R}^n$ is compact, then $f(A) \subset \mathbb{R}^m$ is compact.

Proof. Let $\mathcal{O}$ be an open cover of $f(A)$. For each open set U in $\mathcal{O}$, there is an open set V_U such that $f^{-1}(U) = V_U \cap A$. The collection of all V_U is an open cover of A. Since A is compact, there exists a finite natural number n such that $V_{U_1}, V_{U_2}, \ldots, V_{U_n}$ cover A. Thus, $U_1, U_2, \ldots, U_n$ cover $f(A)$. Hence the result. □

However, if f is not continuous at some point, say a, then there is a measure to which f fails to be so. This is called the oscillation of f at a. Let $\delta > 0$ and

$$\begin{aligned} M(a,f,\delta) &= \sup\{f(x) : x \in A \text{ and } |x-a| < \delta\}, \\ m(a,f,\delta) &= \inf\{f(x) : x \in A \text{ and } |x-a| < \delta\}. \end{aligned}$$

Then, the oscillation of f at a, denoted by $o(f,a)$ is defined by

$$o(f,a) = \lim_{\delta\to 0}[M(a,f,\delta) - m(a,f,\delta)].$$

The limit always exists since the value of $M(a,f,\delta)-m(a,f,\delta)$ decreases with δ. Clearly, from the definition of $o(f,a)$, one may say that the continuity of f will be true as long as the value of $o(f,a)$ is the minimum, that is, 0. The next theorem says exactly that.

Theorem 1.2.12. The bounded function f is continuous at a if and only if $o(f,a) = 0$.

Proof. Let f be continuous at a. For every number $\epsilon > 0$, we can choose a real number $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon \text{ for all } \; x \in A \text{ whenever } |x - a| < \delta.$$

Thus, $M(a,f,\delta) - m(a,f,\delta) \le 2\epsilon$. Since $\epsilon > 0$ is arbitrary, this is true for all ϵ and hence $o(f,a) = 0$. The converse can also be shown similarly. □

Exercise 1.2.13.

1. Let $f : A \to \mathbb{R}^m$ and $a \in A$. Show that $\lim\limits_{x\to a} f(x) = l$ if and only if $\lim\limits_{x\to a} f^i(x) = l_i$, for all i, where $l = (l_1, l_2, \ldots, l_m) \in \mathbb{R}^m$.

2. Find the following limits, if they exist.

   (a) $\lim\limits_{(x,y)\to(0,0)} \dfrac{4x^2 + 3y^2 + x^3y^3}{x^2 + y^2 + x^4y^4}$

   (b) $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^2y + y^3}{x^2 + y^2}$

   (c) $\lim\limits_{(x,y)\to(0,1)} \mathrm{e}^x\, y$

   (d) $\lim\limits_{(x,y)\to(0,0)} \dfrac{\mathrm{e}^{xy}}{x + 1}$

3. Prove that $f : A \to \mathbb{R}^m$ is continuous at a if and only if every f^i is so.

4. Prove that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is continuous.

5. Show that the function $f(x,y) = y\,\mathrm{e}^x + \sin x + (xy)^4$ is continuous.

6. Show that $f(x,y) = \dfrac{x+y}{x-y}$ is continuous at $(1,2)$.

7. Can $\dfrac{xy}{x^2+y^2}$ be made continuous by suitably defining it at $(0,0)$?

1.3 Differentiation

We saw that the characteristics of a multivariable function depends upon its component functions f^i. So notion of one-variable functions is necessary to move forward with the idea of differentiability of a multivariable function. Hence, let's start with the definition of differentiability of a single variable function $f : A \to \mathbb{R}$. Let $a \in A$ and A contains a neighbourhood containing a. We define the derivative of f at a as the limit

$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h},$$

if the limit exists, in which case, it is denoted by $f'(a)$. If the limit exists, then f is said to be differentiable at a. The following facts immediately follow as a result of differentiability.

1. Differentiable functions are continuous;

2. Composites of differentiable functions are differentiable.

We seek such a definition of derivative of the multivariable function $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$ so that the above two conditions are satisfied.

1.3.1 Partial Derivatives

First let us begin with the definition.

Definition 1.3.1. Let $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$ be a function and A contains a neighbourhood of $a = (a_1, a_2, \ldots, a_n)$. Then the ith partial derivative of f, denoted by $\dfrac{\partial f}{\partial x_i}$ (or $D_i f$, $_{x_i}$), is defined by

$$\frac{\partial f}{\partial x_i} = \lim_{t \to 0} \frac{f(a_1, a_2, \ldots, a_{i-1}, a_i + t, a_{i+1}, \ldots, a_n) - f(a_1, a_2, \ldots, a_n)}{t},$$

provided the limit exists. From the definition, we can say that f can have n partial derivatives $\dfrac{\partial f}{\partial x_i}$, $i = 1, \ldots, n$, if all the limits exist.

The partial derivatives at some point a estimate the rate of change of f along the direction of the axes. Further, if all the partial derivative of $f : \mathbb{R}^n \to \mathbb{R}$ exist at $a \in \mathbb{R}^n$, then the vector

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

is called the gradient of f at a.

Example 1.3.2. Let $f : \mathbb{R}^2 \to \mathbb{R}$ such that

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

Then

$$\begin{aligned} \frac{\partial f}{\partial x}(0, 0) &= 0, \\ \text{and} \quad \frac{\partial f}{\partial y}(0, 0) &= 0. \end{aligned}$$

However, f is not continuous at $(0, 0)$.

Thus, the mere existence of partial derivatives may not guarantee continuity at some point. Hence the partial derivatives do not satisfy the first necessary condition of differentiability for one-variable function. Thus, this is not a suitable candidate for deriavtive of f. However, it is important as we shall see later. Let us also check whether continuity imply the existence of partial derivatives or not.

Example 1.3.3. Let $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\begin{aligned} f(x) &= x \sin \frac{1}{y} + y \sin \frac{1}{x}, \text{ when } x \neq 0, y \neq 0 \\ &= x \sin \frac{1}{x}, \text{ when } x \neq 0, y = 0 \\ &= y \sin \frac{1}{y}, \text{ when } x = 0, y \neq 0 \\ &= 0, \text{ when } x = 0, y = 0. \end{aligned}$$

Then f is continuous at $(0, 0)$ but neither $\dfrac{\partial f}{\partial x}(0, 0)$ nor $\dfrac{\partial f}{\partial y}(0, 0)$ exist.

Before concluding this section, let us list down some of the properties of the partial derivatives.

Theorem 1.3.4. Let $f, g : \mathbb{R}^n \to \mathbb{R}$ and $a \in \mathbb{R}^n$. Suppose $\dfrac{\partial f}{\partial x_i}(a)$ and $\dfrac{\partial g}{\partial x_i}(a)$ exist. Then

1. $\dfrac{\partial \alpha f}{\partial x_i}(a) = \alpha \dfrac{\partial f}{\partial x_i}(a)$ for $\alpha \in \mathbb{R}$;

2. $\dfrac{\partial \alpha(f+g)}{\partial x_i}(a) = \dfrac{\partial f}{\partial x_i}(a) + \dfrac{\partial g}{\partial x_i}(a)$;

3. $\dfrac{\partial \alpha(fg)}{\partial x_i}(a) = \dfrac{\partial \alpha f}{\partial x_i}(a)g(a) + f(a)\dfrac{\partial g}{\partial x_i}(a)$.

The proof of the above are easy and left as exercise.

1.3.2 Directional derivative

Since the partial derivatives fail to fulfill the necessary criteria, so we try to generalise the definition further.

Definition 1.3.5. Let $f : A \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$ be a function and A contains a neighbourhood of a. Then, for $0 \neq u \in \mathbb{R}^n$, the directional derivative of f at a with respect to the vector u is denoted by $f'(a; u)$ and is define by

$$f'(a; u) = \lim_{t \to 0} \frac{f(a + tu) - f(a)}{t}$$

provided the limit exists.

Let us see an example.

Example 1.3.6. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(x, y) = xy.$$

Then the directional derivative of f at $a = (a_1, a_2)$ with respect to the vector $(1, 0)$ is

$$f'(a; u) = \lim_{t \to 0} \frac{(a_1 + t)a_2 - a_1 a_2}{t} = a_2.$$

With respect to the vector $v = (1, 2)$, the directional derivative is

$$f'(a; u) = \lim_{t \to 0} \frac{(a_1 + t)(a_2 + 2t) - a_1 a_2}{t} = a_2 + 2a_1.$$

If $u = e_i$, then $f'(a; u) = \dfrac{\partial f}{\partial x_i}$. So, directional derivatives are more generic in nature. It estimates the rate of change of f at a in the direction of u. It should be stated that the sum, product and chain rules for directional derivatives are similar as those in the partial derivatives case. Also, if $f'(a; u)$ exists for all non-zero $u \in \mathbb{R}^n$, then f is said to have directional derivatives in all directions. Obviously, if $f'(a; u)$ exists for all u, then this implies that all the partial derivatives of f exist. However, the converse may not be true.

Example 1.3.7. Let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as follows.

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

Then $\dfrac{\partial f}{\partial x}(0,0) = 0$ and $\dfrac{\partial f}{\partial y}(0,0) = 0$. But, for $u = \left(\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}\right)$,

$$\begin{aligned} f'((0,0);u) &= \lim_{t\to 0} \frac{f\left(\frac{t}{\sqrt{2}}, \frac{t}{\sqrt{2}}\right) - f(0,0)}{t} \\ &= \lim_{t\to 0} \frac{1}{2t} \end{aligned}$$

which does not exist finitely. Thus, despite the existence of all partial derivatives, directional derivative may not exist at some for along all directions.

One might get tempted to believe that perhaps the directional derivative is the best candidate to play the role of the derivatives in the multivariable case. But that is not the case as seen by the example below.

Example 1.3.8. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as

$$\begin{aligned} f(x,y) &= \frac{x^2 y}{x^4 + y^2}, \text{ if } (x,y) \neq (0,0) \\ &= 0, \text{ if } (x,y) = (0,0). \end{aligned}$$

Then f is not continuous at $(0,0)$. However, for $u = (u_1, u_2) \neq (0,0)$, we have

$$\begin{aligned} f'((0,0);u) &= \lim_{t\to 0} \frac{f(tu_1, tu_2) - f(0,0)}{t} \\ &= \lim_{t\to 0} \frac{u_1^2 u_2}{t^2 u_1^4 + u_2^2} \\ &= \frac{u_1^2}{u_2}, \text{ when } u_2 \neq 0. \end{aligned}$$

Also, $f'((0,0);u) = 0$ when exactly one of u_1 and u_2 is zero. Thus, the directional derivative of f exists at $(0,0)$ along every direction but it is not continuous there.

Exercise 1.3.9.

1. Let $A \in \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Show that if $f'(a;u)$ exists, then $f'(a;cu)$ exists and equals $cf'(a;u)$.

2. Find the partial derivatives of the following three variable functions (if they exist).

   (a) $f(x,y,z) = \mathrm{e}^{xyz}(xy + xz + yz)$
   (b) $f(x,y,z) = \sin(xy^2 z^3)$
   (c) $f(x,y,z) = \mathrm{e}^x \cos(yz^2)$
   (d) $f(x,y,z) = \dfrac{xy^3 + \mathrm{e}^z}{x^3 y - e^z}$

3. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined as follows.

$$\begin{aligned} f(x,y) &= \frac{x^2 y^2}{x^2 y^2 + (y-x)^2} \quad \text{if } (x,y) \neq (0,0) \\ &= 0, \quad \text{if } (x,y) = (0,0). \end{aligned}$$

   (a) For which vectors $u \neq 0$ does $f'(0;u)$ exist? Evaluate it when it exists.
   (b) Do $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ exist at $(0,0)$?
   (c) Is f continuous at $(0,0)$?

4. Repeat the same exercise as 3 for $f(x,y) = |x| + |y|$.

1.3.3 Derivative of f

Before starting with the definition of multivariable function f, let us rewrite the definition of derivative of a one-variable function f.

A function $f : A \to \mathbb{R}$ of a single variable x is said to be differentiable at some point a such that A contains a neighbourhood of a if the following limit exists.

$$\lim_{t\to 0} \frac{f(a+t) - f(a)}{t}.$$

If the above limit exists, let it be some real number λ. Then the above limit can be rewritten as

$$\frac{f(a+t) - f(a) - \lambda t}{t} \to 0 \text{ as } t \to 0,$$

where λt is a unique linear transformation (or a linear approximation of f at a) on $\mathbb{R}$. We need to generalise this idea for multivariable function.

Definition 1.3.10. Let $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Suppose A contains a neighbourhood of a. We say that f is differentiable at a of there exists an $m \times n$ matrix B such that

$$\frac{f(a+h) - f(a) - Bh}{|h|} \to 0 \text{ as } h \to 0.$$

The matrix B, if it exists, is called the derivative or total derivative of f at a. It is denoted by $Df(a)$.

The matrix is known as the Jacobian matrix of f. We will discuss more about it in the next unit. It is easy to check that B is unique. Suppose C is another matrix satisfying the same condition. Then subtracting the two conditions yield

$$\frac{(C - B)h}{|h|} \to 0 \text{ as } h \to 0.$$

Let u be a fixed vector and set $h = tu$. Let $t \to 0$. It follows that $(C - B)u = 0$. Since u is arbitrary, so $C = B$.

The above definition seems promising. However, we need to check for continuity and composition results. Continuity is proved by the following result.

Theorem 1.3.11. Let $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. If f is differentiable at a, then f is continuous at a.

Proof. Let $B = Df(a)$. For h near 0 but different from 0, we write

$$f(a+h) - f(a) = |h| \left[\frac{f(a+h) - f(a) - Bh}{|h|}\right] + Bh.$$

By hypothesis, the expression inside the brackets on the right hand side of the above equation tends to 0 as h tends to 0. Thus, by the basic theorems on limits,

$$\lim_{h\to 0}[f(a+h) - f(a)] = 0.$$

Thus, f is continuous at a. □

If f is differentiable at every point of A, then f is called a differentiable function.

Example 1.3.12. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function defined as follows.

$$f(x) = Bx + b$$

where B is an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then f is differentiable and $Df(x) = B$.

Exercise 1.3.13. 1. Check whether each of the following functions are differentiable at the indicated point.

   (a) $f(x, y) = (\mathrm{e}^x, \sin xy)$ at $(1, 3)$;

   (b) $f(x, y, z) = (x - y, y + z)$ at $(1, 0, 1)$;

   (c) $f(x, y, z) = (x + \mathrm{e}^z + y, yx^2)$ at $(1, 1, 0)$;

   (d) $f(x, y) = (x\,\mathrm{e}^y + \cos y, x, x + \mathrm{e}^y)$ at $(1, 0)$.

2. Try to find the derivative matrix of f at the points (if it exists).

Few Probable Questions

1. Define the directional derivative of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ at some point $a \in \mathbb{R}^n$. Show that the if the directional derivatives of f with respect to every $v \neq 0$ in $\mathbb{R}^n$ at a exist, then the partial derivatives exist at a. Is the converse true? Justify your answer.

2. Prove that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $ain\mathbb{R}^n$, then f is continuous at a.

3. Show that $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = \sqrt{|xy|}$ is not differentiable at $(0, 0)$.

4. Prove or disprove: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function and $(a, b) \in \mathbb{R}^2$. Then f is differentiable at (a, b) if and only if there exist real numbers A, B and a function $\psi(h, k)$ such that

$$f(a + h, b + k) - f(a, b) = Ah + Bk + \psi(h, k),$$

   where $\dfrac{\psi(h, k)}{\sqrt{h^2 + k^2}} \to 0$ as $(h, k) \to (0, 0)$.

Unit 2

Course Structure

- Total derivative in terms of partial derivatives, the matrix transformation of $T : \mathbb{R}^n \to \mathbb{R}^n$ Jacobian matrix.

2.1 Introduction

In the previous unit, we learnt about the derivative of a multivariable function. It was defined keeping in line with the definition of derivative for the one-variable case. We also saw that the differentiability in fact implies continuity. However, we did not get much idea about the derivative or the Jacobian matrix. We have also seen in the previous unit that a function $f = (f^1, f^2, \ldots, f^m)$ behaves in close connection with the behaviour of the component functions f^i, be it in the case of limits or continuity. So it only natural to think that it should be the same in case of differentiability as well. In fact, more than that. The component functions are the building blocks of the Jacobian matrix as we shall see shortly. So here, in this unit, we will come across two things.

1. The differentiability of f and its component functions;

2. The structure of the Jacobian matrix.

We will start with the properties of differentiable functions, gradually developing the Jacobian matrix.

Objectives

After reading this unit, you will be able to

- learn some basic properties of a differentiable function f and talk about its partial and directional derivatives;

- learn the structure of the Jacobian matrix of f.

2.2 Total Derivative of f

We came to see that the differentiability of f implies continuity. Also, as we tried to find a good definition of derivative, we cam across two concepts, viz., the partial derivatives and the directional derivatives. It is quite natural question to pose whether the existence of derivative of f has any connection with the existence of partial or directional derivatives of f. The following results answer these questions.

Theorem 2.2.1. Let $A \subset \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. If f is differentiable at a, then all the directional derivatives of f at a exist and

$$f'(a; u) = Df(a) \cdot u.$$

Proof. Let $B = Df(a)$. Since f is differentiable at a, so there exists an $m \times n$ matrix B such that

$$\frac{f(a+h) - f(a) - B \cdot h}{|h|} \to 0 \ \text{ as } \ h \to 0.$$

Let $0 \neq u \in \mathbb{R}^n$. Set $h = tu$ in the above equation, where $t \neq 0$. Then the above expression becomes

$$\frac{f(a+tu) - f(a) - B \cdot tu}{|tu|} \to 0 \ \text{ as } \ t \to 0. \tag{2.2.1}$$

If t approaches 0 through positive values, we multiply 2.2.1 by $|u|$ to conclude that

$$\frac{f(a+tu) - f(a)}{t} - B \cdot u \to 0 \ \text{ as } \ t \to 0.$$

If t approaches 0 through negative values, we multiply 2.2.1 by $-|u|$ to conclude the same. Thus, $f'(a; u) = B \cdot u$. □

Also, since the existence of directional derivatives imply existence of partial derivatives, so we can say that if the derivative exist at some point a, then all its partial derivatives exist at that point. However, the converse may not be true always.

Example 2.2.2. The function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\begin{aligned} f(x,y) &= \frac{x^2 y}{x^4 + y^2}, \ \text{if } (x,y) \neq (0,0) \\ &= 0, \ \text{if } (x,y) = (0,0). \end{aligned}$$

Then the directional derivatives of f exist for all $(u_1, u_2) \neq (0,0)$ at $(0,0)$ but it is not even continuous at $(0,0)$.

We can similarly take a function whose partial derivatives exist but it is not differentiable. For example, the function $f(x,y) = \dfrac{xy}{x^2+y^2}$, $(x,y) \neq (0,0)$ and $f(0,0) = 0$.

Let us now see how the partial derivatives make up the structure of the derivative of f.

Theorem 2.2.3. Let $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}$. If f is differentiable at a, then

$$Df(a) = \left[\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \ldots, \frac{\partial f}{\partial x_n}(a) \right]$$

This means that if $Df(a)$ exists, then it is the row matrix whose entries are the partial derivatives of f at a.

Proof. By hypothesis, $Df(a)$ exists and is a $1 \times n$ matrix of the form

$$Df(a) = [c_1, c_2, \ldots, c_n].$$

From the preceding theorem, we get

$$\frac{\partial f}{\partial x_j}(a) = f'(a; e_j) = Df(a) \cdot e_j = c_j.$$

□

The mere existence of partial derivatives do not guarantee the differentiability of f. However, if f is differentiable, then the derivative matrix is given by the above theorem. Also, it should be noted that the above matrix may exist even if f is not differentiable at a. The above theorem can be further generalised when f is a vector valued function.

Theorem 2.2.4. Let $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$ be a function. Also let the component functions of f be given as follows.

$$f = \begin{bmatrix} f^1 \\ f^2 \\ \vdots \\ f^m \end{bmatrix}.$$

Then f is differentiable at a point $a \in A$ if and only if each f^i is so. Also,

$$Df(a) = \begin{bmatrix} Df^1(a) \\ Df^2(a) \\ \vdots \\ Df^m(a) \end{bmatrix} = \begin{bmatrix} \frac{\partial f^1}{\partial x_1}(a) & \frac{\partial f^1}{\partial x_2}(a) & \ldots & \frac{\partial f^1}{\partial x_n}(a) \\ \frac{\partial f^2}{\partial x_1}(a) & \frac{\partial f^2}{\partial x_2}(a) & \ldots & \frac{\partial f^2}{\partial x_n}(a) \\ \vdots & & & \\ \frac{\partial f^m}{\partial x_1}(a) & \frac{\partial f^m}{\partial x_2}(a) & \ldots & \frac{\partial f^m}{\partial x_n}(a) \end{bmatrix}. \tag{2.2.2}$$

Proof. Let there be an $m \times n$ matrix B. Let

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

which is defined for $0 < |h| < \epsilon$ for some ϵ. Now, $F(h)$ is a column matrix of size $m \times 1$. Its ith entry satisfies the equation

$$F^i(h) = \frac{f^i(a+h) - f^i(a) - (\text{row } i \text{ of } B \cdot h}{|h|}.$$

Let h approach 0. Then the matrix $F(h)$ approaches 0 if and only if each of its entries approaches 0. Hence if B is a matrix for which $F(h) \to 0$, then the ith row of B is a matrix for which $F^i(h) \to 0$. And conversely. Hence the theorem. □

The $m \times n$ matrix in equation (2.2.2) is the complete structure of the Jacobian matrix. As we saw earlier, the Jacobian matrix may or may not exist and it does not depend upon the differentiability of the function. However, if the function is differentiable at a, then the derivative of f at a is equal to the Jacobian matrix. Let us see a few examples.

Example 2.2.5. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined as follows.

$$f(x, y) = (\mathrm{e}^{x+y} + y, y^2 x).$$

Then $f^1(x, y) = \mathrm{e}^{x+y} + y$ and $f^2(x, y) = y^2 x$. Hence,

$$\frac{\partial f^1}{\partial x}(x, y) = \mathrm{e}^{x+y}, \quad \frac{\partial f^1}{\partial y}(x, y) = \mathrm{e}^{x+y} + 1,$$

and

$$\frac{\partial f^2}{\partial x}(x, y) = y^2, \quad \frac{\partial f^2}{\partial y}(x, y) = 2xy.$$

Thus, the Jacobian matrix is given by

$$J = \begin{bmatrix} \mathrm{e}^{x+y} & \mathrm{e}^{x+y} + 1 \\ y^2 & 2xy \end{bmatrix}.$$

Also, note that each of the component functions are differentiable and hence f is differentiable and $Df(x, y) = J$

Exercise 2.2.6. Find the Jacobian matrix of the function $f(x, y, z) = (z\,\mathrm{e}^x, -y\,\mathrm{e}^z)$.

We will state a few properties of the derivative. The proof of them are quite easy and left as exercise.

Theorem 2.2.7. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a constant function, then $Df(a) = 0$.

Theorem 2.2.8. If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then $Df(a) = f$.

Theorem 2.2.9. If $f, g : A \to \mathbb{R}$ are differentiable at a, then

$$\begin{aligned} D(f + g)(a) &= Df(a) + Dg(a), \\ D(f \cdot g)(a) &= g(a)Df(a) + f(a)Dg(a). \end{aligned}$$

Moreover, if $g(a) \neq 0$, then

$$D(f/g)(a) = \frac{g(a)Df(a) - f(a)Dg(a)}{[g(a)]^2}.$$

Few Probable Questions

1. Show that a function $f : A \to \mathbb{R}^n$, $A \subset \mathbb{R}^m$ is differentiable if and only if each of the component functions of f is differentiable.
2. Find the derivative of f at the point $(1, 0, 1)$ if $f : \mathbb{R}^3 \to \mathbb{R}^{\nVDash}$ is given by $f(x, y, z) = (xy^2, \sin(zy), \cos y + \mathrm{e}^x)$, if it exists.
3. Show that if a function f is differentiable at some point a, then all the partial derivatives of f exist. Is the converse true? Justify your answer.

Unit 3

Course Structure

- The chain rule and its matrix form. Mean value theorem for vector valued function. Mean value inequality.

3.1 Introduction

We saw that one of the major properties of differentiable functions is that their composites are also so. This was one of the properties that we were seeking while defining differentiability. For single variable case, this is given by the Chain rule. Chain rule is perhaps one of the principal properties exhibited by differentiable functions. Also, it is desirable that vector valued differentiable functions also exhibit the chain rule. Here in this unit, we will check the chain rule properties that are exhibited by vector valued functions, if any.

Objectives

After reading this unit, you will be able to

- deduce chain rule of differentiation for vector valued functions
- deduce mean value theorem for vector valued functions

3.2 Chain Rule

Theorem 3.2.1. Let $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$. Also let

$$f : A \to \mathbb{R}^m, \text{ and } g : B \to \mathbb{R}^p,$$

with $f(A) \subset B$. Suppose $f(a) = b$. If f is differentiable at a, and if g is differentiable at b, then the composite function $g \circ f$ is differentiable at a. Furthermore,

$$D(g \circ f)(a) = Dg(b) \cdot Df(a),$$

where the indicated product is matrix multiplication.

Proof. Let x be any arbitrary point of $\mathbb{R}^n$ and let y be any arbitrary point of $\mathbb{R}^m$. By hypothesis, g is defined in a neighbourhood of b. Choose $\epsilon > 0$ such that $g(y)$ is defined for $|y - b| < \epsilon$. Similarly, since f is defined in a neighbourhood of a and is continuous at a, we can choose a $\delta > 0$ such that $f(x)$ is defined and satisfies the condition

$$|f(x) - b| < \epsilon, \text{ for } |x - a| < \delta.$$

Then the composite function $(g \circ f)(x) = g(f(x))$ is defined for $|x - a| < \delta$.

Let us define

$$\Delta(h) = f(a + h) - f(a),$$

which is defined for $|h| < \delta$. We shall first show that the quotient $\dfrac{|\Delta(h)|}{|h|}$ is bounded for h in some deleted neighbourhood of 0. For this, let us introduce the function $F(h)$ defined by setting $F(0) = 0$ and

$$F(h) = \frac{[\Delta(h) - Df(a) \cdot h]}{|h|}, \text{ for } 0 < |h| < \delta.$$

Because f is differentiable at a, the function F is continuous at 0. Furthermore, one has the equation

$$\Delta(h) = Df(a) \cdot h + |h|F(h) \tag{3.2.1}$$

for $0 < |h| < \delta$, and also for $h = 0$ (trivially). The triangle inequality implies that

$$|\Delta(h)| \leq n|Df(a)||h| + |h||F(h)|.$$

Now, $|F(h)|$ is bounded for h in a neighbourhood of 0; in fact, it approaches 0 as h approaches 0. Thus, $\dfrac{|\Delta(h)|}{|h|}$ is bounded on a deleted neighbourhood of 0.

We repeat the same for the function g and define $G(k)$ by setting $G(0) = 0$ and

$$G(k) = \frac{g(b + k) - g(b) - Dg(b) \cdot k}{|k|}, \text{ for } 0 < |k| < \epsilon.$$

Since g is differentiable at b, the function G is continuous at 0. Furthermore, for $|k| < \epsilon$, G satisfies the equation

$$g(b + k) - g(b) = Dg(b) \cdot k + |k|G(k). \tag{3.2.2}$$

Now let h be any point in $\mathbb{R}^n$ with $|h| < \delta$. Then $|\Delta(h)| < \epsilon$, so we may substitute $\Delta(h)$ for k in equation (3.2.2). After this, $b + k$ becomes

$$b + \Delta(h) = f(a) + \Delta(h) = f(a + h),$$

and so the equation (3.2.2) takes the form

$$g(f(a + h)) - g(f(a)) = Dg(b) \cdot \Delta(h) + |\Delta(h)|G(\Delta(h)).$$

Now we use equation (3.2.1) to rewrite this equation in the form

$$\frac{1}{|h|}[g(f(a + h)) - g(f(a)) - Dg(b) \cdot Df(a) \cdot h] = Dg(b) \cdot F(h) + \frac{1}{|h|}|\Delta(h)|G(\Delta(h)).$$

This equation holds for $0 < |h| < \delta$. In order to show that $g \circ f$ is differentiable at a with derivative $Dg(b) \cdot Df(a)$, it suffices to show that the right side of this equation tends to zero as h tends to 0.

The matrix $Dg(b)$ is constant, while $F(h) \to 0$ as $h \to 0$ (because F is continuous at 0 and vanishes there). The factor $G(\Delta(h))$ also approaches zero as $h \to 0$; for it is the composite of two functions G and Δ, both of which are continuous at 0 and vanish there. Finally, $\dfrac{|\Delta(h)|}{|h|}$ is bounded in a deleted neighbourhood of 0 by previous discussion. Hence the theorem follows. □

Example 3.2.2. Let $g(x, y) = (x^2 + 1, y^2)$ and $f(u, v) = (u + v, v, v^2)$. Compute the derivative of $f \circ g$ at $(1, 1)$ using the chain rule.

Solution. First let us compute the derivative matrices of f and g.

$$Df(u,v) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 2v \end{bmatrix} \quad \text{and} \quad Dg(x,y) = \begin{bmatrix} 2x & 0 \\ 0 & 2y \end{bmatrix}.$$

When $(x, y) = (1, 1)$, we get $g(1, 1) = (2, 1)$. By chain rule,

$$\begin{aligned} D(f \circ g)(1,1) &= Df(2,1) \cdot Dg(1,1) \\ &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 2 \\ 2 & 0 \\ 0 & 4 \end{bmatrix}. \end{aligned}$$

■

Exercise 3.2.3. Compute $f \circ g$ and $D(f \circ g)$ for the functions f and g below at the indicated points.

1. $f(u, v) = (\tan(u - 1) - \mathrm{e}^v, u^2 - v^2)$ and $g(x, y) = (\mathrm{e}^{x-y}, x - y)$ at $(1, 1)$;
2. $f(u, v, w) = (\mathrm{e}^{u-w}, \cos(u + v) + \sin(u + v + w))$ and $g(x, y) = (\mathrm{e}^x, \cos(y - x), \mathrm{e}^{-y})$ at $(0, 0)$.

3.3 Mean Value Theorem

We are familiar with the mean value theorem of one-variable calculus. Let us restate it as follows.

Theorem 3.3.1. Suppose $f(x)$ is a function that satisfies both of the following.

1. f is continuous on the closed interval $[a, b]$;
2. f is differentiable on the open interval (a, b).

Then there is a number c such that $a < c < b$ and

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

In this section, we will explore the question whether this can be extended for vector valued functions as well. If it is a multivariable real-valued function, then we have the following theorem.

Theorem 3.3.2. Let A be open in $\mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be differentiable on A. If A contains the line segment with end points a and $a + h$, then there is a point $c = a + t_0 h$ with $0 < t_0 < 1$ of this line segment such that

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

Proof. Set $\phi(t) = f(a + th)$. Then ϕ is defined for t in an open interval about $[0, 1]$. Being the composite of differentiable functions, ϕ is differentiable and its derivative is given as follows.

$$\phi'(t) = Df(a + th) \cdot h.$$

The ordinary mean value theorem implies that

$$\phi(1) - \phi(0) = \phi'(t_0) \cdot 1$$

for some t_0 with $0 < t_0 < 1$. This equation can be rewritten in the form

$$f(a + h) - f(a) = Df(a + t_0 h) \cdot h.$$

□

Now, what will happen if the function is vector-valued? Let us consider the following example.

Example 3.3.3. Let $f : \mathbb{R} \to \mathbb{R}^2$ be defined as $f(t) = (\cos t, \sin t)$, $t \in \mathbb{R}$. Suppose a similar mean value theorem is true for vector valued functions. Then, we can have for x and y, there is a point $z = t_0 x + (1 - t_0) y$ for some $t_0 \in (0, 1)$ such that

$$f(y) - f(x) = Df(z)(y - x).$$

For this particular function in the example, if the above statement is true, then for $x, y \in \mathbb{R}$, $\exists z$ in between x and y such that

$$f(y) - f(x) = Df(z)(y - x).$$

If we take $y = 2\pi$ and $x = 0$. Then the left hand side of the above equation becomes $(0, 0)$ whereas the right hand side becomes $2\pi(-\cos z, \sin z)$. Now, since they are equal, so their norms will be equal which yields $0 = 2\pi$, a contradiction. Hence, the Mean Value Theorem does not hold good for vector valued functions.

So, for vector-valued function, we have the following theorem.

Theorem 3.3.4. Let $f : A \to \mathbb{R}^m$, A is an open and connected subset of $\mathbb{R}^n$, be a differentiable function on A and let $x, y \in A$ such that $tx + (1 - t)y \in U$ for all $t \in [0, 1]$. Fix any $a \in \mathbb{R}^m$. Then there exists $t_0 \in (0, 1)$ such that $z = t_0 x + (1 - t_0) y$ and

$$a \cdot (f(y) - f(x)) = a \cdot Df(z)(y - x).$$

Proof. Choose a $\delta > 0$ such that $tx + (1 - t)y \subseteq A$ for all $t \in (-\delta, 1 + \delta)$. Consider the function $F : (-\delta, 1 + \delta) \to \mathbb{R}$

$$F(t) = a \cdot f(x + th), \quad \text{where } h = y - x.$$

Applying MVT on the one-variable function F, we get a $t_0 \in (0, 1)$ such that

$$F(1) - F(0) = a \cdot F'(t_0)(1 - 0). \tag{3.3.1}$$

By Chain rule, $F'(t_0) = a \cdot Df(x + t_0 h)(y - x)$. Putting this and the function F in equation (3.3.1), we get

$$a \cdot (f(y) - f(x)) = a \cdot Df(z)(y - x)$$

where $z = z + t_0 h$. Hence the result. □

Corollary 3.3.5. Let $f : A \to \mathbb{R}^m$, A is an open and connected subset of $\mathbb{R}^n$, be a differentiable function on A and $Df(x) = 0$ for all $x \in A$. Then f is a constant function.

Proof. Since A is open connected, it is polygonally connected. That is, for $x, y \in A$, there exist $x = z_0, z_1, \ldots, z_k = y \in A$ such that the line $[z_{i-1}, z_i] \subset A$ for all $i = 1, \ldots, k$. Let $x, y \in A$. Without any loss of generality, we assume that the line joining x and y is completely contained in A that is, $tx + (1-t)y \in A$ for all $t \in [0, 1]$. Fix any $a \in \mathbb{R}^m$. Applying MVT, we get

$$a \cdot (f(y) - f(x)) = a \cdot Df(z)(y - x)$$

where $z = t_0 x + (1 - t_0)y$. Since $Df(z) = 0$, so the above equation becomes

$$a \cdot (f(y) - f(x)) = 0.$$

Take $u = f(y) - f(x)$. Then the above equation becomes

$$\|f(y) - f(x)\|^2 = 0$$

which implies $f(x) = f(y)$. Since x and y are arbitrary, so f is constant. □

Few Probable Questions

1. State and prove the Chain rule for multivariable function.
2. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ satisfy the conditions $f(0) = (1, 2)$ and

$$Df(0) = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

   Also let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $g(x, y) = (x + 2y + 1, 3xy)$. Find $D(g \circ f)(0)$.

Unit 4

Course Structure

- A sufficient condition for differentiability. A sufficient condition for mixed partial derivatives.

4.1 Introduction

In this unit, we will obtain a sufficient condition for differentiability. We know that the mere existence of partial derivatives or the existence of the Jacobian matrix does not guarantee the differentiability of f. If however, we impose mild additional condition that these partial derivatives are continuous, then we shall see that the differentiability is assured.

Objectives

After reading this unit, you will be able to

- learn a sufficient condition for differentiability
- learn a sufficient condition for mixed partial derivatives

4.2 Continuously Differentiable functions

We start by stating the theorem as follows.

Theorem 4.2.1. Let A be an open set in $\mathbb{R}^n$ and suppose that the partial derivatives $\dfrac{\partial f^i}{\partial x_j}$ of the component functions of f exist at each point x of A and are continuous on A. Then f is differentiable at each point of A.

Proof. We know that f is differentiable if and only if its component functions are so. Hence, we will only restrict ourselves to real-valued function $f : A \to \mathbb{R}$.

Let $a \in A$. We are given that for some ϵ, the partial $\dfrac{\partial f}{\partial x_j}$derivatives exist and are continuous for $|x-a| < \epsilon$. We wish to show that f is differentiable at a.

Let $h \in \mathbb{R}^n$ with $0 < |h| < \epsilon$ and let $h = (h_1, \dots, h_n)$. Consider the following sequence of points of $\mathbb{R}^n$:

$$\begin{aligned} p_0 &= a, \\ p_1 &= a + h_1 e_1, \\ p_2 &= a + h_1 e_1 + h_2 e_2 \\ &\dots \\ p_n &= a + h_1 e_1 + \dots + h_n e_n = a + h. \end{aligned}$$

The points p_i all belong to the closed cube C of radius $|h|$ centered at a. Since we are concerned with the differentiability of f, we shall need to deal with the difference $f(a+h) - f(a)$. We begin by writing it in the form

$$f(a+h) - f(a) = \sum_{j=1}^{m} [f(p_j) - f(p_{j-1})]. \tag{4.2.1}$$

Consider the general term of this summation. Let j be fixed, and define

$$\phi(t) = f(p_{j-1} + te_j).$$

Assume $h_j \neq 0$ for the moment. As t ranges over the closed interval I with end points 0 and h_j, the point $p_{j-1} + te_j$ ranges over the line segment from p_{j-1} to p_j; this line segment lies in C, and hence in A. Thus ϕ is defined for t in an open interval about I.

As t varies, only the jth component of the point $p_{j-1} + te_j$ varies. Hence because $\dfrac{\partial f}{\partial x_j}$ exists at each point of A, the function ϕ is differentiable on an open interval containing I. Applying the mean value-theorem to ϕ, we conclude that

$$\phi(h_j) - \phi(0) = \phi'(c_j) h_j$$

for some point c_j between 0 and h_j. (This argument applies whether h_j is positive or negative.) We can rewrite this equation in the form

$$f(p_j) - f(p_{j-1}) = \frac{\partial f}{\partial x_j}(q_j) h_j, \tag{4.2.2}$$

where q_j is the point $p_{j-1} + c_j e_j$ of the line segment from p_{j-1} to p_j, which lies in C. □

A function that satisfies the hypotheses of the above theorem are called continuously differentiable function, or of class C^1, on A.

Example 4.2.2. Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as

$$f(x, y) = \frac{\cos x + \mathrm{e}^{xy}}{x^2 + y^2}$$

is differentiable at all points $(x, y) \neq (0, 0)$.

Solution. The partial derivatives are

$$\begin{aligned} \frac{\partial f}{\partial x} &= \frac{(x^2 + y^2)(y\,\mathrm{e}^{xy} - \sin x) - 2x(\cos x + \mathrm{e}^{xy})}{(x^2 + y^2)^2} \\ \frac{\partial f}{\partial y} &= \frac{(x^2 + y^2)x\,\mathrm{e}^{xy} - 2y(\cos x + \mathrm{e}^{xy})}{(x^2 + y^2)^2} \end{aligned}$$

are continuous except when $x = 0$ and $y = 0$. So, it is differentiable at all points other than $(0, 0)$. ■

In practice, we usually deal only with functions that are of class C^1. While it is interesting to know there are functions that are differentiable but not of class C^1, such functions occur rarely enough that we need not be concerned with them.

Suppose f is a function mapping an open set A of $\mathbb{R}^n$ into $\mathbb{R}^n$, and suppose the partial derivatives $\dfrac{\partial f^i}{\partial x_j}$ of the component functions of f exist on A. These then are functions from A to $\mathbb{R}$, and we may consider their partial derivatives, which have the form $\dfrac{\partial}{\partial x_k}\left(\dfrac{\partial f^i}{\partial x_j}\right)$ and are called the second-order partial derivatives of f. Similarly, one defines the third-order partial derivatives of the functions f_i or more generally the partial derivatives of order r for arbitrary r.

If the partial derivatives of the functions f^i of order less than or equal to r are continuous on A, we say f is of class C^r on A. Then the function f is of class C^r on A, if and only if each of the functions $\dfrac{\partial f^i}{\partial x_j}$ is of class C^{r-1} on A. We say f is of class C^∞ on A, if the partials of the functions f^i of all orders are continuous on A.

As you may recall, for most functions the "mixed" partial derivatives

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f^i}{\partial x_j}\right) \quad \& \quad \frac{\partial}{\partial x_j}\left(\frac{\partial f^i}{\partial x_k}\right)$$

are equal. This result in fact holds under the hypothesis that the function f is of class C^2, as we now show.

Theorem 4.2.3. Let A be open in $\mathbb{R}^n$; let $f : A \to \mathbb{R}$ be a function of class C^2. Then for each $a \in A$, we have

$$\frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial x_j}\right)(a) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_k}\right)(a).$$

Proof. Since one calculates the partial derivatives in question by letting all variables other than x_k and x_j remain constant, it suffices to consider the case where f is a function merely of two variables. So we assume that A is open in $\mathbb{R}^2$, and that $f : A \to \mathbb{R}^2$ is of class C^2.

Let

$$Q = [a, a+h] \times [b, b+k]$$

be a rectangle contained in A. Define

$$\lambda(h,k) = f(a,b) - f(a+h,b) - f(a,b+k) + f(a+h,b+k).$$

Then λ is the sum, with appropriate signs, of the values of f at the four vertices of Q. We show that there are points p and q of Q such that

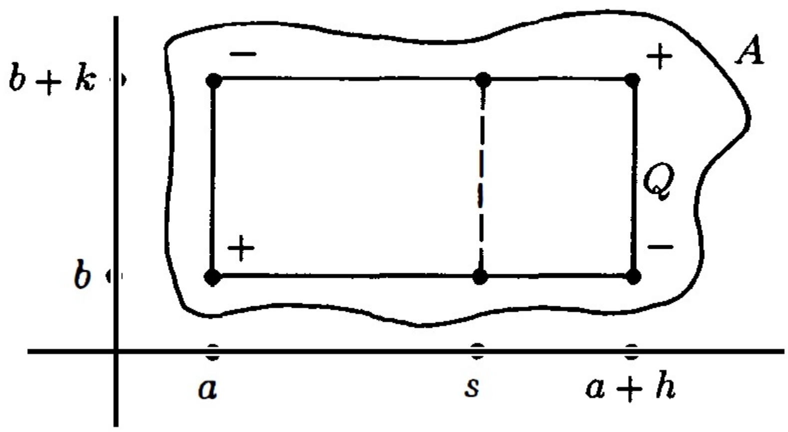
$$\lambda(h,k) = \frac{\partial}{\partial x_2}\left(\frac{\partial}{\partial x_1}\right) f(p).hk, \quad \& \quad \lambda(h,k) = \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2}\right) f(q).hk.$$

By symmetry, it suffices to prove the first of these equations. To begin, we define

$$\phi(s) = f(s, b+k) - f(s,b).$$

Then $\phi(a+h) - \phi(a) = \lambda(h,k)$. Because $\dfrac{\partial}{\partial x_1} f$ exists in A, the function ϕ is differentiable in an open interval containing $[a, a+h]$. The mean-value theorem implies that

$$\phi(a+h) - \phi(a) = \phi'(s_0).h$$



for some s_0 between a and $a+h$. This equation can be rewritten in the form

$$\lambda(h,k) = \left[\frac{\partial}{\partial x_1} f(s_0, b+k) - \frac{\partial}{\partial x_1} f(s_0, b)\right].h. \tag{4.2.3}$$

Now, s_0 is fixed, and we consider the function $\frac{\partial}{\partial x_1} f(s_0, t)$. Because $\frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f$ exists in A, this function is differentiable for t in an open interval about $[b, b+k]$. We apply the mean-value theorem once more to conclude that

$$\frac{\partial}{\partial x_1} f(s_0, b+k) - \frac{\partial}{\partial x_1} f(s_0, b) = \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f(s_0, t_0).k \tag{4.2.4}$$

for some t_0 between b and $b+k$. Combining (4.2.3) and (4.2.4), we get,

$$\lambda(h,k) = \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f(s_0, t_0).hk. \tag{4.2.5}$$

Now, we prove the theorem. Given the point $a = (a, b)$ of A and given $t > 0$, let Q_t be the rectangle

$$Q_t = [a, a+t] \times [b, b+t].$$

If t is sufficiently small, Q_t is contained in A. Then by (4.2.5), we get

$$\lambda(t,t) = \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f(p_t).t^2$$

for some point p_t in Q_t. If we let $t \to 0$, then $p_t \to a$. Because $\frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f$ is continuous, it follows that

$$\lambda(t,t)/t^2 \to \frac{\partial}{\partial x_2}\frac{\partial}{\partial x_1} f(a) \quad \text{as} \quad t \to 0.$$

A similar argument, using symmetry, gives

$$\lambda(t,t)/t^2 \to \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_2} f(a) \quad \text{as} \quad t \to 0.$$

Hence the theorem. $\square$

Few Probable Questions

1. Show that the function $f(x, y) = |xy|$ is differentiable at 0, but is not of class C^1 in any neighborhood of 0.

2. Define $f : \mathbb{R} \to \mathbb{R}$ by setting $f(0) = 0$ and

$$f(t) = t^2 \sin\left(\frac{1}{t}\right) \quad \text{if} \quad t \neq 0.$$

   (a) Show f is differentiable at 0, and calculate $f'(0)$.

   (b) Calculate $f'(t)$ if $t \neq 0$.

   (c) Show f' is not continuous at 0.

   (d) Conclude that f is differentiable on $\mathbb{R}$ but not of class C^1 on $\mathbb{R}$.

3. Show that if $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}$, and if the partials $D_j f$ exist and are bounded in a neighborhood of a, then f is continuous at a.

Unit 5

Course Structure

- Functions with non-zero Jacobian determinant, the inverse function theorem, the implicit function theorem as an application of Inverse function theorem.

5.1 Introduction

In mathematics, specifically differential calculus, the inverse function theorem gives a sufficient condition for a function to be invertible in a neighbourhood of a point in its domain: namely, that its derivative is continuous and non-zero at the point. The theorem also gives a formula for the derivative of the inverse function. In multivariable calculus, this theorem can be generalized to any continuously differentiable, vector-valued function whose Jacobian determinant is nonzero at a point in its domain, giving a formula for the Jacobian matrix of the inverse which we will explore here.

We will explore the implicit function theorem as an application of the inverse function theorem in this unit. In multivariable calculus, the implicit function theorem is a tool that allows relations to be converted to functions of several real variables. It does so by representing the relation as the graph of a function. There may not be a single function whose graph can represent the entire relation, but there may be such a function on a restriction of the domain of the relation. The implicit function theorem gives a sufficient condition to ensure that there is such a function.

Objectives

After reading this unit, you will be able to

- learn about the consequences of non-zero Jacobian determinant of vector valued functions
- learn the inverse function theorem and related theorems and lemmas
- apply the inverse function theorem in various examples
- learn the implicit function theorem as an application of the inverse function theorem
- apply the implicit function theorem in various problems

5.2 Functions with non-zero Jacobian determinant

We have read about the chain rule in the previous unit and the mean value theorem as an application of it. As yet another application of the chain rule, we consider the problem of differentiating an inverse function.

Recall the situation that occurs in single-variable analysis. Suppose $\phi(x)$ is differentiable on an open interval, with $\phi^{''}(x) > 0$ on that interval. Then ϕ is strictly increasing and has an inverse function ψ, which is defined by letting $\psi(y)$ be that unique number x such that $\phi(x) = y$. The function ψ is in fact differentiable, and its derivative satisfies the equation

$$\psi'(y) = \frac{1}{\phi'(x)},$$

where $y = \phi(x)$.

There is a similar formula for differentiating the inverse of a function f of several variables. In the present section, we do not consider the question whether the function f has an inverse, or whether that inverse is differentiable. We consider only the problem of finding the derivative of the inverse function.

Theorem 5.2.1. Let A be open in $\mathbb{R}^n$ and let $f : A \to \mathbb{R}^n$ such that $f(a) = b$. Suppose that g maps a neighbourhood of b into $\mathbb{R}^n$, such that $g(b) = a$ and $g(f(x)) = x$ for all x in a neighbourhood of a. If f is differentiable at a and if g is differentiable at b, then $Dg(b) = [Df(a)]^{-1}$.

Proof. Let $i : \mathbb{R}^n \to \mathbb{R}^n$ be the identity function; its derivative is the identity matrix I_n. We are given that $g(f(x)) = i(x)$ for all x in a neighbourhood of a. Then the chain rule implies that

$$Dg(b).Df(a) = I_n.$$

Thus, $Dg(b)$ is the inverse matrix to $Df(a)$. □

The preceding theorem implies that if a differentiable function f is to have a differentiable inverse, it is necessary that the matrix Df be non-singular. It is a somewhat surprising fact that this condition is also sufficient for a function f of class C^1 to have an inverse , at least locally.

Let us make a comment on notation. The usefulness of well-chosen notation can hardly be overemphasized. Arguments that are obscure, and formulas that are complicated, sometimes become beautifully simple once the proper notation is chosen. Our use of matrix notation for the derivative is a case in point . The formulas for the derivatives of a composite function and an inverse function could hardly be simpler. Nevertheless, a word may be in order for those who remember the notation used in calculus for partial derivatives, and the version of the chain rule proved there.

In advanced mathematics, it is usual to use either the functional notation ϕ' or the operator $D\phi$ for for the derivative of a real-valued function of a real variable. ($D\phi$ denotes a 1×1 matrix in this case!) In calculus, however, another notation is common. One often denotes the derivative $\phi'(x)$ by the symbol $d\phi/dx$.

5.3 The Inverse Function Theorem

Let A be open in $\mathbb{R}^n$ and let $f : A \to \mathbb{R}^n$ be of class C^1. We know that for f to have a differentiable inverse, it is necessary that the derivative $Df(x)$ of f be non-singular. We now prove that this condition is also sufficient for f to have a differentiable inverse, at least locally. This result is called the inverse function theorem.

We begin by showing that non-singularity of Df implies that f is locally one-to-one.

Lemma 5.3.1. Let A be open in $\mathbb{R}^n$ and let $f : A \to \mathbb{R}^n$ be of class C^1. If $Df(a)$ is non-singular, then there exists an $a > 0$ such that the inequality

$$|f(x_0) - f(x_1)| \geq a|x_0 - x_1|$$

holds for all x_0, x_1 in some open cube $C(a; \epsilon)$ centered at a. It follows that f is one-to-one on this open cube.

Proof. Let $E = Df(a)$. Then E is non-singular. We first consider the linear transformation that maps x to Ex. We compute

$$|x_0 - x_1| = |E^{-1}.(E.x_0 - E.x_1)| \leq |E^{-1}||E.x_0 - E.x_1|.$$

If we set $2a = 1/n|E^{-1}|$, then for all $x_0, x_1 \in \mathbb{R}^n$,

$$|E.x_0 - E.x_1| \geq 2a|x_0 - x_1|.$$

Now consider the function $H(x) = f(x) - E.x$. Then $DH(x) = Df(x) - E$, so that $DH(a) = 0$. Since H is of class C^1, we can choose $\epsilon > 0$ such that $|DH(x)| < a/n$ for x in the open cube $C = C(a;\epsilon)$. The mean-value theorem, applied to the ith component function of H, tells us that, given $x_0, x_1 \in C$, there is a $c \in C$ such that

$$|H_i(x_0) - H_i(x_1)| = |DH_i(c).(x_0 - x_1)| \neq n(a/n)|x_0 - x_1|.$$

Thus for $x_0, x_1 \in C$, we have

$$\begin{aligned} a|x_0 - x_1| &\geq |H(x_0) - H(x_1)| \\ &= |f(x_0) - E.x_0 - f(x_1) + E.x_1| \\ &\geq |E.x_1 - E.x_0| - |f(x_1) - f(x_0)| \\ &\geq 2a|x_1 - x_0| - |f(x_1) - f(x_0)|. \end{aligned}$$

Hence the result. □

We will now state a theorem which says that the non-singularity of Df, in the case where f is one-to-one, implies that the inverse function is differentiable.

Theorem 5.3.2. Let A be open in $\mathbb{R}^n$ and let $f : A \to \mathbb{R}^n$ be of class C^r. Let $B = f(A)$. If f is one-to-one on A and if $Df(x)$ is non-singular for $x \in A$, then the set B is open in $\mathbb{R}^n$ and the inverse function $g : B \to A$ is of class C^r.

We leave the proof of this theorem. We will finally prove the inverse function theorem.

Theorem 5.3.3. (**Inverse Function Theorem**) Let A be open in $\mathbb{R}^n$ and let $f : A \to \mathbb{R}^n$ be of class C^r. If $Df(x)$ is non-singular at the point $a \in A$, there is a neighbourhood U of the point a such that f carries U in a one-to-one fashion onto an open set V of $\mathbb{R}^n$ and the inverse function is of class C^r.

Proof. By lamma 5.3.1, there is a neighborhood U_0 of a on which f is one-to-one. Because $\det Df(x)$ is a continuous function of x, and $\det Df(a) \neq 0$, there is a neighbourhood U_1 of a such that $\det Df(x) \neq 0$ on U_1. If $U = U_0 \cap U_1$, then the hypotheses of the preceding theorem are satisfied for $f : U \to \mathbb{R}^n$. The theorem follows. □

This theorem is the strongest one that can be proved in general . While the non-singularity of Df on A implies that f is locally one-to-one at each point of A, it does not imply that f is one-to-one on all of A. Consider the following example:

Example 5.3.4. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by the equation

$$f(r,\theta) = (r\cos\theta, r\sin\theta).$$

Then

$$Df(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix},$$

so that $\det Df(r,\theta) = r$.

Let A be the open set $(0,1) \times (0,b)$ in the r-θ plane. Then Df is non-singular at each point of A. However, f is one-to-one on A only if $b \leq 2\pi$.

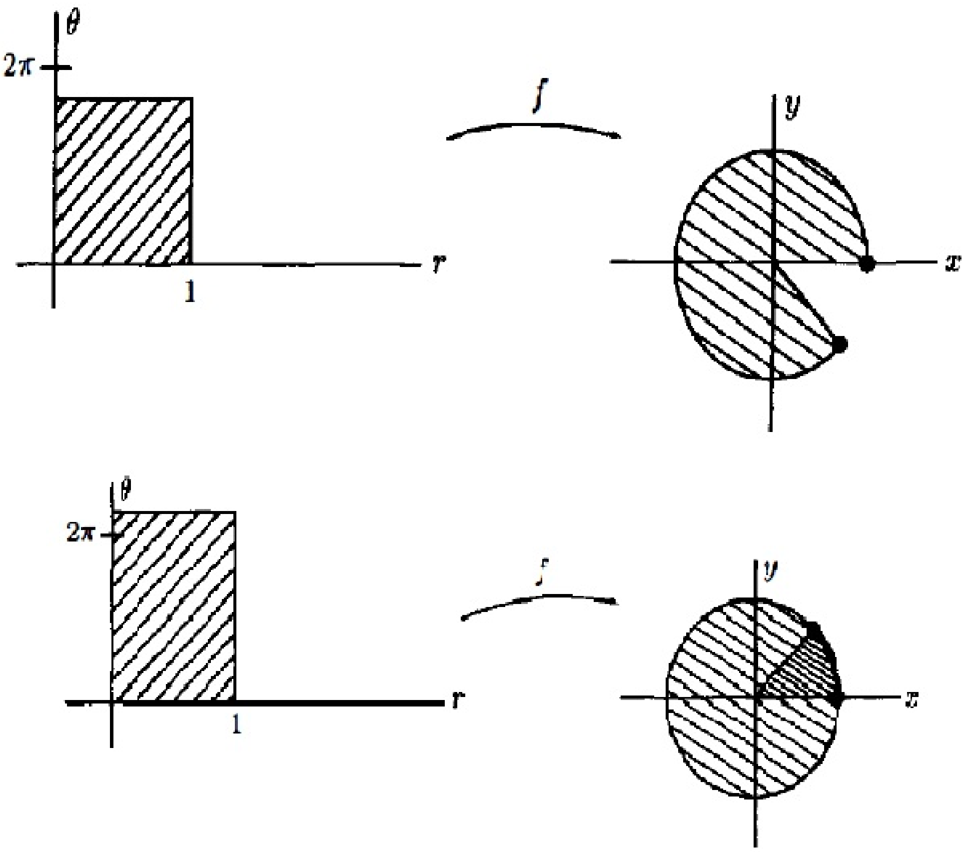


Figure 5.3.1: f in example 5.3.4

Exercise 5.3.5.

1. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by the equation

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

   (a) Show that f is one-to-one on the set A consisting of all (x, y) with $x > 0$.

   (b) What is the set $B = f(A)$?

   (c) If g is the inverse function, find $Dg(0, 1)$.

2. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by the equation

$$f(x, y) = (\mathrm{e}^x \cos y, \mathrm{e}^x \sin y).$$

   (a) Show that f is one-to-one on the set A consisting of all (x, y) with $0 < y < 2\pi$.

   (b) What is the set $B = f(A)$?

   (c) If g is the inverse function, find $Dg(0, 1)$.

5.4 Implicit Function Theorem

The topic of implicit differentiation is one that is probably familiar to you from calculus. Here is a typical problem:

Assume that the equation $x^3y + 2\,e^{xy} = 0$ determines y as a differentiable function of x. Find dy/dx.

One solves this calculus problem by "looking at y as a function of x, " and differentiating with respect to x. One obtains the equation

$$3x^2y + x^3\frac{dy}{dx} + 2\,e^{xy}\left(y + x\frac{dy}{dx}\right) = 0,$$

which one solves for dy/dx. The derivative dy/dx is of course expressed in terms of x and the unknown function y.

The case of an arbitrary function f is handled similarly. Supposing that the equation $f(x, y) = 0$ determines y as a differentiable function of x , say $y = g(x)$, the equation $f(x, g(x)) = 0$ is an identity. One applies the chain rule to calculate

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

which gives

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

where the partial derivatives are evaluated at the point $(x, g(x))$. Note that the solution involves a hypothesis not given in the statement of the problem. In order to find $g'(x)$, it is necessary to assume that $\partial f/\partial y$ is non-zero at the point in question.

It in fact turns out that the non-vanishing of $\partial f/\partial y$ is also sufficient to justify the assumptions we made in solving the problem. That is, if the function $f(x, y)$ has the property that $\partial f/\partial y \neq 0$ at a point (a, b) that is a solution of the equation $f(x, y) = 0$, then this equation does determine y as a function of x , for x near a , and this function of x is differentiable.

This result is a special case of a theorem called the implicit function theorem, which we prove in this section. The general case of the implicit function theorem involves a system of equations rather than a single equation. One seeks to solve this system for some of the unknowns in terms of the others. Specifically, suppose that $f : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is a function of class C^1. Then the vector equation

$$f(x_1, \ldots, x_{k+n}) = 0$$

is equivalent to a system of n scalar equations in $k + n$ unknowns. One would expect to be able to assign arbitrary values to k of the unknowns and to solve for the remaining unknowns in terms of these. One would also expect that the resulting functions would be differentiable, and that one could by implicit differentiation find their derivatives.

There are two separate problems here. The first is the problem of finding the derivatives of these implicitly defined functions, assuming they exist; the solution to this problem generalizes the computation of $g'(x)$ just given. The second involves showing that (under suitable conditions) the implicitly defined functions exist and are differentiable.

In order to state our results in a convenient form, we introduce a new notation for the matrix D f and its submatrices:

Definition 5.4.1. Let A be open in $\mathbb{R}^m$; let $f : A \to \mathbb{R}^n$ be differentiable and $f_1, \ldots, f_n$ be the component functions of f. We sometimes use the notation

$$Df = \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_m)}$$

for the derivative of f. On occasion we shorten this to the notation

$$Df = \frac{\partial f}{\partial x}.$$

More generally, we shall use the notation

$$\frac{\partial(f_{i_1}, \ldots, f_{i_k})}{\partial(x_{j_1}, \ldots, x_{j_l})}$$

to denote the $k \times l$ matrix that consists of the entries of Df lying in rows $i_1, \ldots, i_k$ and columns $j_1, \ldots, j_l$. The general entry of this matrix, in row p and column q, is the partial derivative $\partial f_{i_p}/\partial x_{j_q}$.

Now we deal with the problem of fin ding the derivative of an implicitly defined function, assuming it exists and is differentiable. For simplicity, we shall assume that we have solved a system of n equations in $k+n$ unknowns for the last n unknowns in terms of the first k unknowns.

Theorem 5.4.2. Let A be open in $\mathbb{R}^{k+n}$; let $f : A \to \mathbb{R}^n$ be differentiable. Write f in the form $f(x, y)$, for $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^n$; then Df has the form

$$Df = \begin{bmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{bmatrix}.$$

Suppose there is a differentiable function $g : B \to \mathbb{R}^n$ defined on an open set B in $\mathbb{R}^k$, such that

$$f(x, g(x)) = 0$$

for all $x \in B$. Then for $x \in B$,

$$\frac{\partial f}{\partial x}(x, g(x)) + \frac{\partial f}{\partial y}(x, g(x)).Dg(x) = 0.$$

This equation implies that if the $n \times n$ matrix $\partial f/\partial y$ is non-singular at the point $(x, g(x))$, then

$$Dg(x) = -\left[\frac{\partial f}{\partial y}(x, g(x))\right]^{-1} . \frac{\partial f}{\partial x}(x, g(x)).$$

Note that in the case $n = k = 1$, this is the same formula for the derivative that was derived earlier; the matrices involved are 1×1 matrices in that case.

Proof. Given g, let us define $h : B \to \mathbb{R}^{k+n}$ by the equation

$$h(x) = (x, g(x)).$$

The hypotheses of the theorem imply that the composite function

$$H(x) = f(h(x)) = f(x, g(x))$$

is defined and equals zero for all $x \in B$. The chain rule then implies that

$$\begin{aligned} 0 &= DH(x) = Df(h(x)).Dh(x) \\ &= \begin{bmatrix} \dfrac{\partial f}{\partial x}(h(x)) & \dfrac{\partial f}{\partial y}(h(x)) \end{bmatrix} \cdot \begin{bmatrix} I_k \\ Dg(x) \end{bmatrix} \\ &= \frac{\partial f}{\partial x}(h(x)) + \frac{\partial f}{\partial y}(h(x)).Dg(x), \end{aligned}$$

as desired. □

The preceding theorem tells us that in order to compute Dg, we must assume that the matrix $\partial f/\partial y$ is non-singular. Now we prove that the non-singularity of $\partial f/\partial y$ suffices to guarantee that the function g exists and is differentiable.

Theorem 5.4.3. (Implicit function theorem). Let A be open in $\mathbb{R}^{k+n}$; let $f : A \to \mathbb{R}^n$ be of class C^r. Write f in the form $f(x, y)$, for $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^n$. Suppose that (a, b) is a point of A such that $f(a, b) = 0$ and

$$\det \frac{\partial f}{\partial y}(a, b) \neq 0.$$

Then there is a neighbourhood B of a in $\mathbb{R}^k$ and a unique continuous function $g : B \to \mathbb{R}^n$ such that $g(a) = b$ and

$$f(x, g(x)) = 0, \ \ \forall x \in B.$$

The function g is in fact of class C^r.

Proof. We construct a function F to which we can apply the inverse function theorem. Define $F : A \to \mathbb{R}^{k+n}$ by the equation

$$F(x, y) = (x, f(x, y)).$$

Then F maps the open set A of $\mathbb{R}^{k+n}$ into $\mathbb{R}^k \times \mathbb{R}^n = \mathbb{R}^{k+n}$. Furthermore,

$$DF = \begin{bmatrix} I_k & 0 \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{bmatrix}.$$

Computing $\det DF$, we have

$$\det DF = \det \frac{\partial f}{\partial y}.$$

Thus DF is non-singular at the point (a, b). Now $F(a, b) = (a, 0)$. Applying the inverse function theorem to

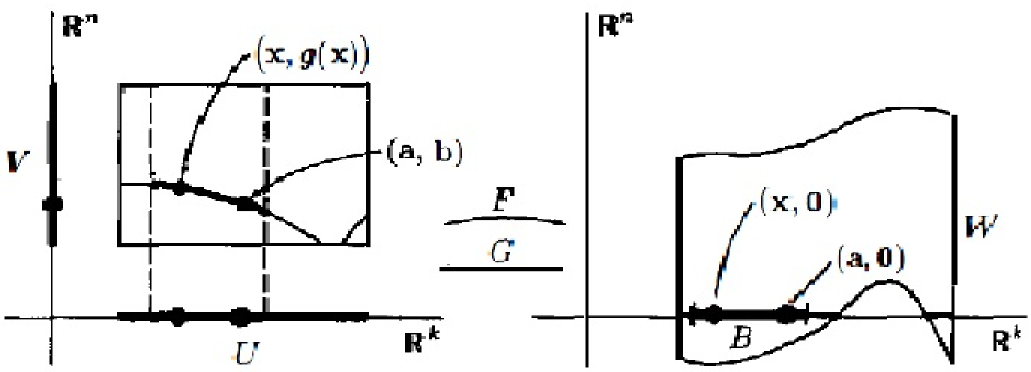


Figure 5.4.1: Implicit Function Theorem

the map F, we conclude that there exists an open set $U \times V$ of $\mathbb{R}^{k+n}$ about (a, b) (where U is open in $\mathbb{R}^k$ and V is open in $\mathbb{R}^n$) such that

1. F maps $U \times V$ in a one-to-one fashion onto an open set W in $\mathbb{R}^{k+n}$ containing $(a, 0)$.

2. The inverse function $G : W \to U \times V$ is of class C^r.

Note that since $F(x, y) = (x, f(x, y))$, we have

$$(x, y) = G(x, f(x, y)).$$

Thus G preserves the first k coordinates, as F does . Then we can write G in the form

$$G(x, z) = (x, h(x, z)), \text{ for } x \in \mathbb{R}^k \text{ and } x \in \mathbb{R}^n.$$

Here h is a function of class C^r mapping W into $\mathbb{R}^n$.

Let B be a connected neighbourhood of a in $\mathbb{R}^k$, chosen small enough that $B \times 0$ is contained in W. We prove existence of the function $g : B \to \mathbb{R}^n$. If $x \in B$, then $(x, 0) \in W$, so that we have

$$\begin{aligned} G(x, 0) &= (x, h(x, 0)), \\ (x, 0) &= F(x, h(x, 0)) = (x, f(x, h(x, 0))), \\ 0 &= f(x, h(x, 0)). \end{aligned}$$

We set $g(x) = h(x, 0)$, for $x \in B$; then g satisfies the equation $f(x, g(x)) = 0$, as desired. Further

$$(a, b) = G(a, 0) = (a, h(a, 0));$$

then $b = g(a)$, as desired.

Now we prove the uniqueness of g. Let $g_0 : B \to \mathbb{R}^n$ be a continuous function satisfying the conditions in the conclusion of our theorem. Then in particular, g_0 agrees with g at the point a. We show that if g_0 agrees with g at the point $a_0 \in B$, then g_0 agrees with g in a neighbourhood B_0 of a_0. This is easy. The map g carries a_0 into V. Since g_0 is continuous, there is a neighbourhood B_0 of a_0 contained in B such that g_0 also maps B_0 into V. The fact that $f(x, g_0(x)) = 0$ for $x \in B_0$ implies that

$$\begin{aligned} F(x, g_0(x)) &= (x, 0), \quad \text{so} \\ (x, g_0(x)) &= G(x, 0) = (x, h(x, 0)). \end{aligned}$$

Thus, g_0 and g agrees on B_0. It follows that g_0 and g agrees on all of B: The set of points of B for which $|g(x) - g_0(x)| = 0$ is open in B and so is the set of points of B for which $|g(x) - g_0(x)| > 0$ (by continuity of g and g_0). Since B is connected, the latter set must be empty. $\square$

In our proof of the implicit function theorem, there was of course nothing special about solving for the last n coordinates ; that choice was made simply for convenience. The same argument applies to the problem of solving for any n coordinates in terms of the others.

For example, suppose A is open in $\mathbb{R}^5$ and $f : A \to \mathbb{R}^2$ is a function of class C^r. Suppose one wishes to "solve" the equation $f(x, y, z, u, v) = 0$ for the two unknowns y and u in terms of the other three. In this case, the implicit function theorem tells us that if a is a point of A such that $f(a) = 0$ and

$$\det \frac{\partial f}{\partial(y, u)}(a) \neq 0,$$

then one can solve for y and u locally near that point, say $y = \phi(x, z, v)$ and $u = \psi(x, z, v)$. Furthermore, the derivatives of ϕ and ψ satisfy the formula

$$\frac{\partial(\phi, \psi)}{\partial(x, z, v)} = -\left[\frac{\partial f}{\partial(y, u)}\right]^{-1} \cdot \left[\frac{\partial f}{\partial(x, z, v)}\right].$$

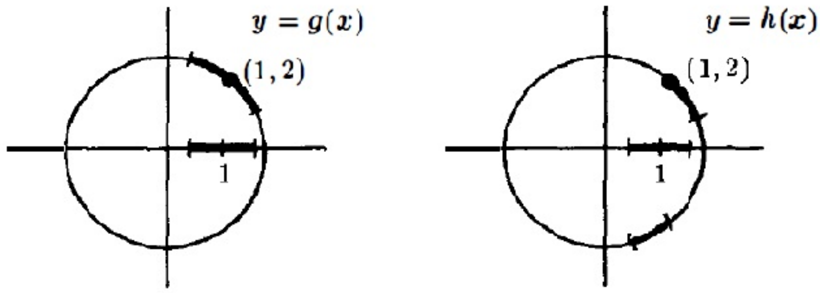


Figure 5.4.2: Example 5.4.4

Example 5.4.4. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by the equation $f(x, y) = x^2+y^2-5$. Then the point $(x, y) = (1, 2)$ satisfies the equation $f(x, y) = 0$. Both $\partial f/\partial x$ and $\partial f/\partial y$ are non-zero at $(1, 2)$, so we can solve this equation locally for either variable in terms of the other. In particular, we can solve for y in terms of x, obtaining the function

$$y = g(x) = [5 - x^2]^{1/2}.$$

Note that this solution is not unique in a neighbourhood of $x = 1$ unless we specify that g is continuous. For instance, the function

$$\begin{aligned} h(x) &= [5 - x^2]^{1/2}, \quad \text{for } x \geq 1, \\ &= -[5 - x^2]^{1/2}, \quad \text{for } x < 1. \end{aligned}$$

satisfies the same conditions, but is not continuous.

Example 5.4.5. The point $(x, y) = (\sqrt{5}, 0)$ also satisfies the equation $f(x, y) = 0$ for the function in example 5.4.4. The derivative $\partial f/\partial y$ vanishes at $(\sqrt{5}, 0)$, so we do not expect to be able to solve for y in terms of x near this point. And, in fact, there is no neighbourhood B of $\sqrt{5}$ on which we can solve for y in terms of x.

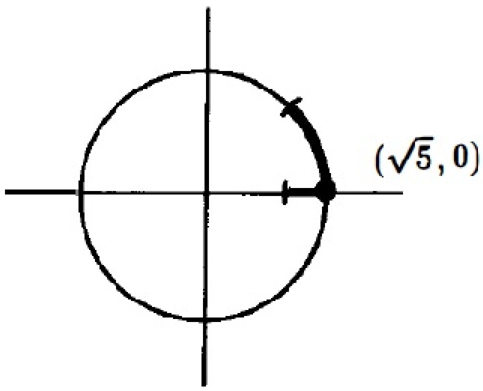


Figure 5.4.3: Example 5.4.5

Example 5.4.6. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by the equation $f(x, y) = x^2 - y^3$. Then $(0, 0)$ is a solution of the equation $f(x, y) = 0$. Because $\partial f/\partial y$ vanishes at $(0, 0)$, we do not expect to be able to solve this equation for y in terms of x near $(0, 0)$. But in fact, we can; and furthermore, the solution is unique! However, the function we obtain is not differentiable at $x = 0$.

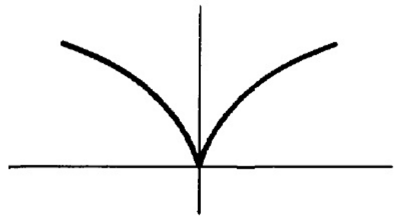


Figure 5.4.4: Example 5.4.6

Example 5.4.7. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by the equation $f(x, y) = y^2 - x^4$. Then $(0, 0)$ is a solution of the equation $f(x, y) = 0$. Because $\partial f/\partial y$ vanishes at $(0, 0)$, we do not expect to be able to solve for y in terms of x near $(0, 0)$. In fact , however, we can do so, and we can do so in such a way that the resulting function is differentiable. However, the solution is not unique. Now the point $(1, 2)$ also satisfies the equation

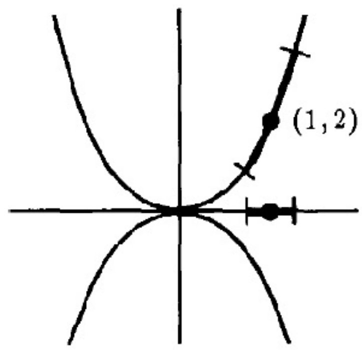


Figure 5.4.5: Example 5.4.7

$f(x, y) = 0$. Because $\partial f/\partial y$ is non-zero at $(1, 2)$, one can solve this equation for y as a continuous function of x in a neighbourhood of $x = 1$. One can in fact express y as a continuous function of x on a larger neighbourhood than the one pictured, but if the neighbourhood is large enough that it contains 0, then the solution is not unique on that larger neighbourhood.

Few Probable Questions

1. State and prove the inverse function theorem.
2. State and prove the implicit function theorem.
3. Let $f : \mathbb{R}^3 \to \mathbb{R}^2$ be of class C^1; write f in the form $f(x, y_1, y_2)$. Assume that $f(3, -1, 2) = 0$ and
$$Df(3, -1, 2) = \begin{bmatrix} 1 & 2 & 1 \\ 1 & -1 & 1 \end{bmatrix}.$$
   (a) Show that there exists a function $g : B \to \mathbb{R}^2$ of class C^1 defined on an open set B in $\mathbb{R}$ such that
$$f(x, g_1(x), g_2(x)) = 0, \quad \text{for } x \in B, \quad \text{and } g(3) = (-1, 2).$$

(b) Find $Dg(3)$.

(c) Discuss the problem of solving the equation $f(x, y_1, y_2) = 0$ for an arbitrary pair of the unknowns in terms of the third, near the point $(3, -1, 2)$.

4. Let $f : \mathbb{R}^5 \to \mathbb{R}^2$ be of class C^1. Let $a = (1, 2, -1, 3, 0)$. Suppose that $f(a) = 0$ and

$$Df(a) = \begin{bmatrix} 1 & 3 & 1 & -1 & 2 \\ 0 & 0 & 1 & 2 & -4 \end{bmatrix}.$$

(a) Show that there exists a function $g : B \to \mathbb{R}^2$ of class C^1 defined on an open set B in $\mathbb{R}^3$ such that

$$f(x_1, g_1(x), g_2(x), x_2, x_3) = 0, \text{ for } x = (x_1, x_2, x_3) \in B, \text{ and } g(1, 3, 0) = (2, -1).$$

(b) Find $Dg(1, 3, 0)$.

(c) Discuss the problem of solving the equation $f(x) = 0$ for an arbitrary pair of the unknowns in terms of the others, near the point a.

Unit 6

Course Structure

- Extremum problems with side conditions – Lagrange's necessary conditions as an application of Inverse function theorem.

6.1 Introduction

We are quite familiar of a function, often called the extrema of a function are probably one of the most elementary topics in the theory of real valued functions. We have come across both global minima or maxima as well as local minima or maxima (together called the local extrema). In order to find the extrema, we saw that the derivatives play a significant role. However, it is also correct to think that the same applies for multivariable real-valued functions. In this unit, we shall concentrate on finding the conditions of extrema of functions of the form $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$. One of the main applications of the concept of maxima and minima is to solve optimization problems arising in economics such as expenditure minimization problem, profit maximization problem, utility maximization problem. Most of these problems are concerned with maximizing and minimizing real-valued n-variable function called objective function and there are some constraints also attached with the problem which are again represented as a functional relationship. Such problems can be solved by a method called Lagrange Multiplier method.

Objectives

After reading this unit, you will be able to

- define critical points, stationary points, saddle points, local maxima and local minima;
- state a necessary condition for functions to have local extrema and apply it;
- state and prove the theorem known as "second derivative test" which gives a sufficient condition for finding local maxima and minima;
- use Hessian for classifying local maxima and local minima;
- apply Lagrange's multiplier method for finding the stationary points when the variables are subject to some constraints.

6.2 Local Maxima and Local Minima

Definition 6.2.1. Let $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$, be a function. A point $a \in A$ is called a maximum point (resp. minimum point) with respect to A if $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$) for all $x \in A$.

If $a \in A$ is either maximum or minimum point w.r.t. A, then that point is called an extreme point or point of extrema. Now we define local extrema.

Definition 6.2.2. Let $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$, be a function, where A is an open set. A point $a \in A$ is called a local maximum (resp. local minimum) of f if there exists a neighbourhood N_a of a such that $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$) for all $x \in N_a$.

Example 6.2.3. Let us consider the function given by

$$f(x, y) = (x + 1)^2 + (y - 3)^2 - 1.$$

We first note that $f(-1, 3) = -1$. Also, $f(x, y) \geq f(-1, 3)$ for all $(x, y) \in \mathbb{R}^2$. This shows that the function has a minimum at $(-1, 3)$ and the minimum value is $f(-1, 3) = -1$.

Let us consider another example.

Example 6.2.4. Let $f(x, y) = \dfrac{1}{2} - \sin(x^2 + y^2)$. Here, $f(0, 0) = \dfrac{1}{2}$. Let us consider the neighbourhood U of $(0, 0)$ given by

$$U = \left\{(x, y) \in \mathbb{R}^2 : \, x^2 + y^2 < \frac{\pi}{6}\right\}.$$

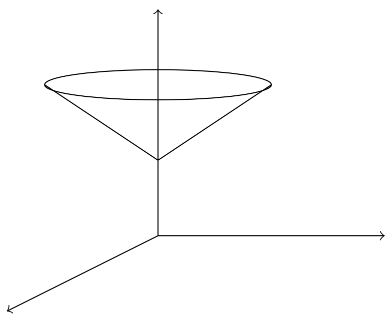
Then for any $(x, y) \in U$, we have, $\sin(x^2 + y^2) > 0$ and therefore

$$f(x, y) = \frac{1}{2} - \sin(x^2 + y^2) < \frac{1}{2} = f(0, 0).$$

Thus, $f(x, y) \leq f(0, 0)$ for all $(x, y) \in U$ in the disc. Note that $f(x, y)$ can be greater than $\dfrac{1}{2}$ for $(x, y) \notin U$. Hence, f has a local minimum at $(0, 0)$.

In the above two examples, we can see that the partial derivatives at the extrema exists and are 0. However, this may not be the case always. Let us see the example below.

Example 6.2.5. Let $f(x, y) = 1 + \sqrt{x^2 + y^2}$. Clearly, from the figure below, we can say that $(0, 0)$ is a minimum point of f. However, the partial derivatives don't exist there.



Now we state a result which shows that if all the first order partial derivatives of f exists at a point $a \in A$, where A is an open set, then they necessarily vanish at the points of extrema.

Theorem 6.2.6. Let $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$ is an open set. Suppose that all the first order partial derivatives of the function f exists at a point $a \in A$. Then a necessary condition for the function to have a local extremum at the point a is that $\dfrac{\partial f}{\partial x_i}(a) = 0$ for $i = 1, \ldots, n$.

Proof. Suppose that f has a local extrema at the point $a = (a_1, a_2, \ldots, a_n)$. Let us consider the real-valued function ϕ defined by

$$\phi(t) = f(t, a_2, \ldots, a_n).$$

Since a is an extreme point of f, we get that a_1 is an extreme point of ϕ. Then from one-variable calculus we know that

$$\phi'(a_1) = \frac{\partial f}{\partial x_1}(a_1, a_2, \ldots, a_n) = 0.$$

In this way, we show that $\dfrac{\partial f}{\partial x_j}(a_1, a_2, \ldots, a_n) = 0$ for each $j = 1, \ldots, n$. Hence the result. □

Now we make the following definition.

Definition 6.2.7. Let $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$, be a function. A point $a \in A$ is called a critical point of f if either

1. the partial derivatives of f do not exist at a, or
2. $\dfrac{\partial f}{\partial x_i}(a) = 0$ for $i = 1, 2, \ldots, n$.

The points for which the condition 2 is satisfied are called stationary points.

You may recall that all stationary points of a function need not be its point of local extrema. Such points are called saddle points. Note that a point $a \in A$ is called a saddle point if every neighbourhood N_a of a contains point $x \in A$ such that $f(x) > f(a)$ and other points $y \in N_a$ such that $f(y) < f(a)$.

Let us consider an example.

Example 6.2.8. Let us consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x, y) = (y - x^2)(y - 2x^2).$$

Here we have $\dfrac{\partial f}{\partial x}(0, 0) = \dfrac{\partial f}{\partial y}(0, 0) = 0$. Thus, $(0, 0)$ is a stationary point. Now, the graph of the function f given below shows that $(0, 0)$ is not a point of local extrema. Note that the function f assumes both positive and negative values in every neighbourhood of $(0, O)$. Therefore $(0, O)$ is a saddle point for the function f.

Next we discuss a sufficient condition in terms of second order partial derivatives to check whether a point is an local extremum point.

Theorem 6.2.9. (Second-derivative test for extrema) Let $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$. Assume that the second-order partial derivative $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$ exist in an open ball $B(a)$ and are continuous at $a \in \mathbb{R}^n$, where a is a stationary point of f. Let

$$Q(x) = \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} f(a) x_i x_j \tag{6.2.1}$$

where $x = (x_1, \ldots, x_n)$. Then

1. if $Q(x) > 0$ for all $x \neq 0$, f has a relative minima at a;

2. if $Q(x) < 0$ for all $x \neq 0$, f has a relative maxima at a;

3. if $Q(x)$ takes both positive and negative values, then f has a saddle point at a.

A real-valued function Q defined on $\mathbb{R}^n$ by an equation of the type

$$Q(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

where $x = (x_1, x_2, \ldots, x_n)$ and a_{ij} are real called a quadratic form. The form is called symmetric if $a_{ij} = a_{ji}$ for all i and j. It is called positive definite if $x \neq 0$ implies $Q(x) > 0$, and negative definite if $x \neq 0$ implies $Q(x) < 0$.

In general, it is not easy to determine whether a quadratic form is positive or negative definite. One criterion, involving determinants, can be described as follows. Let $D =$ determinant of the matrix $[a_{ij}]$ and let D_k denote the determinant of the $k \times k$ matrix obtained by deleting the last $n - k$ rows and columns of the matrix $[a_{ij}]$. Also put $D_0 = 1$. From the theory of quadratic forms it is known that a necessary and sufficient condition for a symmetric form to be positive definite is that the $n + 1$ numbers $D_0, D_1, \ldots, D_n$ be positive. The form is negative definite if and only if, the same $n + 1$ numbers are alternately positive and negative. The quadratic form which appears in equation (6.2.1) s symmetric because the mixed partial derivatives $\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f(a)$ and $\frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i} f(a)$ are equal. Therefore, under the conditions of the above theorem, we see that f has a local minimum at a if the $(n + 1)$ numbers $D_0, D_1, \ldots, D_n$ of the corresponding Jacobian matrix for f are all positive, and a local maximum if these numbers are alternately positive and negative.

We have the following result.

Theorem 6.2.10. If $f : A \to \mathbb{R}$, where A is an open subset of $\mathbb{R}^n$, has continuous first and second-order partial derivatives at a where a is a critical point of f, and Hf is the Hessian of f given as follows:

$$Hf = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

and it is evaluated at a; also let D_k denote the determinant o0f the $k \times k$ matrix obtained by deleting the last $(n - k)$ rows and column of the matrix. Then the following hold:

1. If $D_{2k} < 0$ for some k then a is a saddle point of f;

2. If $D_n \neq 0$ then

   (a) f has a local minimum at a if and only if $D_k > 0$ for all k,

   (b) f has a local maximum at a if and only if $(-1)^k D_k > 0$ for all k;

3. If $D_n = 0$ we call it a degenerate case and the test cannot be applied.

The case $n = 2$ can be handled directly and gives the following criterion.

Theorem 6.2.11. Let f be a real-valued function with continuous second-order partial derivatives at a stationary point a in $\mathbb{R}^2$. Let

$$A = \frac{\partial^2}{\partial x_1^2} f(a), \quad B = \frac{\partial^2}{\partial x_1 \partial x_2} f(a), \quad C = \frac{\partial^2}{\partial x_2^2} f(a)$$

and let

$$D = \det \begin{bmatrix} A & B \\ B & C \end{bmatrix} = AC - B^2.$$

Then we have

1. If $D > 0$ and $A > 0$, f has a local minimum at a.
2. If $D > 0$ and $A < 0$, f has a local maximum at a.
3. If $D < 0$, f has a saddle point at a.

If $D = 0$, then the result fails. Let us consider some examples.

Example 6.2.12. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be defined as follows.

$$f(x, y, z) = x^2y^2 + z^2 + 2x - 4y + z.$$

We check f for extrema. First,

$$Df = (2xy^2 + 2, 2x^2y - 4, 2z + 1).$$

If a is a critical point of f, then a satisfies the following system of equations.

$$\begin{aligned} 2xy^2 + 2 &= 0 \\ 2x^2y - 4 &= 0 \\ 2z + 1 &= 0 \end{aligned}$$

Solving, we get $a = \left(-2^{2/3}, 2^{-1/3}, -1/2\right)$, which is the only critical point of f. Now, we see that the Hessian matrix of f is given as

$$Hf = \begin{bmatrix} 2y^2 & 4xy & 0 \\ 4xy & 2x^2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

and

$$Hf\left(-2^{2/3}, 2^{-1/3}, -1/2\right) = \begin{bmatrix} 2^{1/3} & -4 \cdot 2^{1/3} & 0 \\ -4 \cdot 2^{1/3} & 2 \cdot 2^{4/3} & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Note that,

$$D_1 = 2^{1/3} > 0$$

and

$$D_2 = \begin{vmatrix} 2^{1/3} & -4 \cdot 2^{1/3} \\ -4 \cdot 2^{1/3} & 2 \cdot 2^{4/3} \end{vmatrix} = 2 \cdot 2^{5/3} - 16 \cdot 2^{2/3} = 4 \cdot 2^{2/3} - 16 \cdot 2^{2/3} < 0.$$

Hence, by theorem 6.2.10, the critical point $\left(-2^{2/3}, 2^{-1/3}, -1/2\right)$ is a saddle point of f.

Exercise 6.2.13. Find the critical points of $f(x, y, z) = (x^2y + y^2z + z^2 - 2x)$ and check whether they are extreme points of f.

6.3 Lagrange Multiplier

Lagrange multiplier method is a technique for finding a maximum or minimum of a function $f(x_1, \ldots, x_n)$ subject to a constraint (also called side condition) of the form $g(x_1, \ldots, x_n) = 0$. Let us see a practical situation to understand this.

Suppose that $f(xy, z)$ represents the temperature at the point (x, y, z) in space and we want to find the maximum or minimum value of the temperature on a certain surface. If the equation of the surface is given explicitly in the form $z = h(x, y)$, then in the expression for $f(x, y, z)$ we can replace z by $h(x, y)$ to obtain the temperature on the surface as a function of x and y alone, say $F(x, y) = f(x, y, h(x, y))$. The problem is then reduced to finding the extreme value of F. However, in practice, certain difficulties arise. The equation of the surface might be given in an implicit form, say $g(x, y, z) = 0$ and it may be impossible, in practice, to solve this equation explicitly for z in terms of x and y, or even for x or y in terms of the remaining variables. The problem might be further complicated by asking for the extreme values of the temperature at those points which lie on a given curve in space. Such a curve can be the intersection of two surfaces, say $g_1(x, y, z) = 0$ and $g_2(x, y, z) = 0$. If we could solve these two equations simultaneously, say for x and y in terms of z, then we could introduce these expressions into f and obtain a new function of z alone, whose extrema we would then seek. In general, however, this procedure cannot be carried out and a more practicable method need to be sought. An elegant and useful method for solving such problems was developed by Lagrange.

Lagrange's method provides a necessary condition for a point to be an extreme point which we shall explain now.

Let $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$ be an open set, be a function whose extreme values are sought when the variables are restricted by a certain number of side conditions, say $g_1(x_1, \ldots, x_n) = 0, \ldots, g_m(x_1, \ldots, x_n) = 0$. Let us form a new function as follows.

$$L(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - \lambda_1 g_1(x_1, \ldots, x_n) - \ldots - \lambda_m g_m(x_1, \ldots, x_n) \tag{6.3.1}$$

where $\lambda_1, \ldots, \lambda_m$ are m constants. We then differentiate L with respect to each coordinate and consider the following system of $n + m$ equations:

$$\frac{\partial L}{\partial x_i} = 0, \ \ i = 1, 2, \ldots, n, \tag{6.3.2}$$

$$g_k(x_1, \ldots, x_n) = 0, \ \ k = 1, 2, \ldots, m. \tag{6.3.3}$$

Lagrange proved that if $(x_1, \ldots, x_n)$ is a point of extrema for f, then it will also satisfy this system of $(n+m)$ equations. In practice, we solve for the $n + m$ unknowns $\lambda_1, \ldots, \lambda_m$. The point $(x_1, x_2, \ldots, x_n)$ so obtained is a stationary point. According to the Lagrange's theorem this point can then be tested for maximum or minimum point by the already known methods.

The numbers $\lambda_1, \ldots, \lambda_m$ are introduced only to help to solve the system for $x_l, x_2, \ldots, x_n$ and they are called Lagrange's multipliers. One multiplier is introduced for each side condition.

Theorem 6.3.1. Let $f : A \subset \mathbb{R}^n \to \mathbb{R}, A$ an open set in $\mathbb{R}^n$, be such that the partial derivatives of f exists and are continuous on A. Let $g_1, \ldots, g_m$ be m real-valued functions defined on A such that partial derivatives of g_i exists and are continuous on A for $i = 1, \ldots, m$. Let us assume that $m < n$. Let X_0 be that subset of A on which each g_i vanishes for $i = 1, \ldots, m$, that is,

$$\mathrm{X_0 = \{x \in E, g_i(x) = 0 \text{ for } i = 1, \ldots, m\}}.$$

Assume that $x_0 \in X_0$ and assume that there exists a ball $B(x_0)$ in $\mathbf{R}^{\mathrm{n}}$ such that $f(x) \leq f(x_0)$ for all x in $X_0 \cap B(x_0)$ or such that $f(x) \geq f(x_0)$ for all x in $\mathrm{X_0} \cap \mathbf{B}(\mathrm{x_0})$. Assume also that the m-rowed determinant $\det[\mathrm{D_j g_i}(\mathrm{x_0})] \neq 0$. Then there exist m real numbers $\lambda_1, \ldots, \lambda_{\mathrm{m}}$ such that they satisfy following n equations:

$$\frac{\partial f}{\partial x_i}(x_0) - \sum_{k=1}^{m} \lambda_k \frac{\partial g_k}{\partial x_i}(\mathbf{x}_0) = 0 \quad (i = 1, 2, \ldots, n).$$

Let us illustrate the steps involved in finding extrems using Lagrange's method.

1. Form the Lagrangian function given in Equation (6.3.1)
2. Form the Lagrangian equations given in Equations (6.3.2) and (6.3.3). The solution thus obtained is a stationary point.
3. Check the stationary point for extrema by the methods already discussed in the preceding section.

Here we state a sufficient condition for checking extrema when we have a single constraint. In this case the Equation (6.3.1) reduces to

$$L(x_1, x_2, \dots, x_n, \lambda) = f(x_1, x_2, \dots, x_n) - \lambda g(x_1, x_2, \dots, x_n). \tag{6.3.4}$$

To check that the stationary point obtained by Lagrange method is local b maximum or local minimum, we need to compute the value of $n-1$ principal minors of the following determinant

$$D_{n+1} = \begin{vmatrix} 0 & \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2} & \cdots & \frac{\partial g}{\partial x_n} \\ \frac{\partial g}{\partial x_1} & \frac{\partial^2 f}{\partial x_1^2} - \lambda \frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} - \lambda \frac{\partial^2 g}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} - \lambda \frac{\partial^2 g}{\partial x_1 \partial x_n} \\ \frac{\partial g}{\partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} - \lambda \frac{\partial^2 g}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_1^2} - \lambda \frac{\partial^2 g}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} - \lambda \frac{\partial^2 g}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial g}{\partial x_n} & \frac{\partial^2 f}{\partial x_n \partial x_1} - \lambda \frac{\partial^2 g}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} - \lambda \frac{\partial^2 g}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} - \lambda \frac{\partial^2 g}{\partial x_n^2} \end{vmatrix}.$$

If the signs of minors D_3, D_4 D_5 are alternatively positive and negative, then extreme point is a local maximum. But if sign of all minors D_3, D_4 D_5 are negative, the extreme point is a local minimum.

Example 6.3.2. Suppose we want to find the extreme values of the function

$$Z = 2x_1^2 + x_2^2 + 3x_3^2 + 10x_1 + 8x_2 + 6x_3 - 100$$

subject to the constraint

$$x_1 + x_2 + x_3 = 20, \quad x_1, x_2, x_3 \geq 0.$$

Solution. Here $n = 3$ and $m = 1$. Let $g(x_1, x_2, x_3) = x_1 + x_2 + x_3 - 20$. Lagrangian function can be formulated as:

$$L(x, \lambda) = 2x_1^2 + x_2^2 + 3x_3^2 + 10x_1 + 8x_2 + 6x_3 - 100 - \lambda(x_1 + x_2 + x_3 - 20).$$

To obtain the stationary points, we solve the following system of equations.

$$\frac{\partial L}{\partial x_1} = 4x_1 + 10 - \lambda = 0; \quad \frac{\partial L}{\partial x_2} = 2x_2 + 8 - \lambda = 0$$
$$\frac{\partial L}{\partial x_3} = 6x_3 + 6 - \lambda = 0; \quad g(x_1, x_2, x_3) = x_1 + x_2 + x_3 - 20 = 0.$$

Putting the value of x_1, x_2, x_3 in the last equation $g(x_1, x_2, x_3) = 0$, and solving for λ, we get $\lambda = 30$. Substituting the value of λ in the other three equations, we get the stationary point $(5, 11, 4)$. To prove the sufficient condition whether the stationary point gives maximum or minimum value of the function we evaluate 2 principal minors.

$$D_3 = \begin{vmatrix} 0 & \frac{\partial g}{\partial x_1} & \frac{\partial g}{\partial x_2} \\ \frac{\partial g}{\partial x_1} & \frac{\partial^2 f}{\partial x_1^2} - \lambda \frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} - \lambda \frac{\partial^2 g}{\partial x_1 \partial x_2} \\ \frac{\partial g}{\partial x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} - \lambda \frac{\partial^2 g}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} - \lambda \frac{\partial^2 g}{\partial x_2^2} \end{vmatrix}_{(5,11,4)} = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 2 \end{vmatrix} = -6$$

$$D_4 = \begin{vmatrix} 0 & 1 & 1 & 1 \\ 1 & 4 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 1 & 0 & 0 & 6 \end{vmatrix} = 48.$$

Since the signs of D_3 and D_4 are alternative, the stationary point is a local maximum. At this point the value of the function is, $Z = 281$. ■

Exercise 6.3.3.

1. Find and clarify the extreme values of the following functions subject to the constraints given along side.

   (a) $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$ subject to the constraint $4x_1 + x_2^2 + 2x_3 = 14,\ x_1, x_2, x_3 \geq 0$.

   (b) $f(x_1, x_2) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2$ subject to the constraint $x_1 + 2x_2 = 2,\ x_1, x_2 \geq 0$.

2. A rectangular box without a lid is to be made from 27 m^2 of cardboard. Find the maximum volume of such a box.

Few Probable Questions

1. Suppose a function $f : A \to \mathbb{R}$, $A \subset \mathbb{R}^n$ is an open set such that all the first order partial derivatives of f exists at a point $a \in A$. If f has a local extrema at a, then show that $\dfrac{\partial f}{\partial x_i}(a) = 0$ for all $i = 1, 2, \ldots, n$.

2. Use the Second derivative test for multivariable functions to find the relative extrema and saddle points, if they exists, of the function $f(x, y) = 4y^3 + x^2 - 12y^2 - 36y + 2$.

3. Find the maximum and minimum of $f(x, y, z) = 4y - 2z$ subject to the constraints $2xyz = 2$ and $x^2 + y^2 = 1$.

Unit 7

Course Structure

- Integration on $\mathbb{R}^n$: Integral of $f : A \to \mathbb{R}$ when $A \subset \mathbb{R}^n$ closed rectangle.

7.1 Introduction

The multiple integral is a definite integral of a function of more than one real variable, for example, $f(x, y)$ or $f(x, y, z)$. Integrals of a function of two variables over a region in $\mathbb{R}^2$ are called double integrals, and integrals of a function of three variables over a region of $\mathbb{R}^3$ are called triple integrals.

Just as the definite integral of a positive function of one variable represents the area of the region between the graph of the function and the x-axis, the double integral of a positive function of two variables represents the volume of the region between the surface defined by the function (on the three-dimensional Cartesian plane where $z = f(x, y)$ and the plane which contains its domain. If there are more variables, a multiple integral will yield hypervolumes of multidimensional functions.

Objectives

After reading this unit, you will be able to

- define the partition of a rectangle in $\mathbb{R}^n$
- define the upper and lower sums of a bounded function defined on a closed rectangle and their relationships with respect to refinements
- define the integral of a bounded function defined on a closed rectangle, if it exists
- learn a necessary and sufficient condition for the existence of the integral of a bounded function over a closed rectangle
- apply the theorems in various problems

7.2 Integral Over a Closed Rectangle

We begin by defining the volume of a rectangle. Let

$$Q = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

be a rectangle in $\mathbb{R}^n$. Each of the intervals $[a_i, b_i]$ is called the component interval of Q. The maximum of the numbers $b_1 - a_1, \ldots, b_n - a_n$ is called the width of Q . Their product

$$v(Q) = (b_1 - a_1)(b_2 - a_2) \ldots (b_n - a_n)$$

is called the volume of Q.

In the case $n = 1$, the volume and the width of the (1-dimensional) rectangle $[a, b]$ are the same, namely, the number $b - a$. This number is also called the length of $[a, b]$.

Definition 7.2.1. Given a closed interval $[a, b]$ of $\mathbb{R}$, a partition of $[a, b]$ is a finite collection P of points of $[a, b]$ that includes the points a and b. We usually index the elements of P in increasing order, for notational convenience, as

$$a = t_0 < t_1 < \cdots < t_k = b;$$

each of the intervals $[t_{i-1}, t_i]$, for $i = 1, \ldots, k$, is called a subinterval determined by P, of the interval $[a, b]$. More generally, given a rectangle

$$Q = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

in $\mathbb{R}^n$, a partition P of Q is an n-tuple $(P_1, \ldots, P_n)$ such that P_j is a partition of $[a_j, b_j]$ for each j. If for each j, I_j is one of the subintervals determined by P_j of the interval $[a_j, b_j]$, then the rectangle

$$R = I_1 \times \cdots \times I_n$$

is called a subrectangle determined by P, of the rectangle Q. The maximum width of these subrectangles is called the mesh of P.

Definition 7.2.2. Let Q be a rectangle in $\mathbb{R}^n$ and let $f : Q \to \mathbb{R}$ be a bounded function. Let P be a partition of Q. For each subrectangle R determined by P, let

$$m_R(f) = \inf\{f(x) : \ x \in R\}, \quad M_R(f) = \sup\{f(x) : \ x \in R\}.$$

We define the lower sum and the upper sum, respectively, of f, determined by P, by the equations

$$\begin{aligned} L(f, P) &= \sum_R m_R(f).v(R), \\ U(f, P) &= \sum_R M_R(f).v(R). \end{aligned}$$

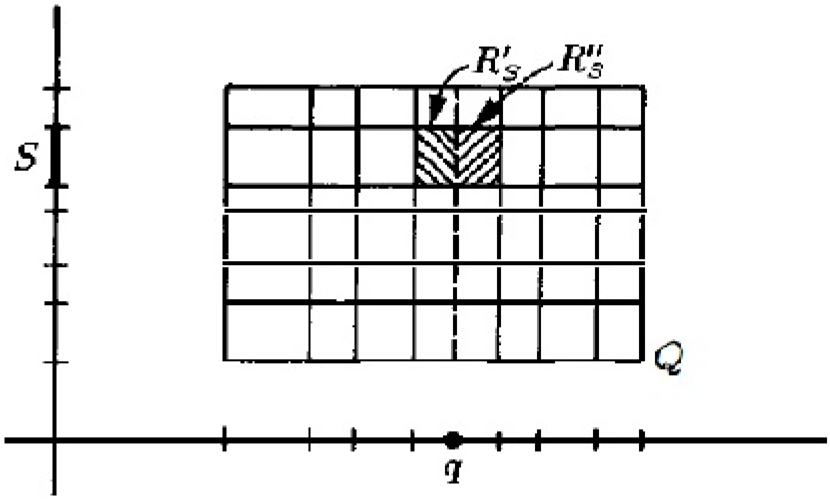
where the summations extend over all subrectangles R determined by P.

Let $P = (P_1, \ldots, P_n)$ be a partition of the rectangle Q. If P'' partition of Q obtained from P by adjoining additional points to some or all of the partitions $P_1, \ldots, P_n$, then P" is called a refinement of P. Given two partitions P and $P' = (P_1^{'}, \ldots, P_n^{'})$ of Q, the partition

$$P'' = (P_1 \cup P_1^{'}, \ldots, P_n \cup P_n^{'})$$

is a refinement of both P and P'; it is called their common refinement.

Passing from P to a refinement of P of course affects lower sums and upper sums; in fact, it tends to increase the lower sums and decrease the upper sums as we have seen in the case of one-dimensional upper and lower sums. That is the substance of the following lemma:



Lemma 7.2.3. Let P be a partition of the rectangle Q and let $f : Q \to \mathbb{R}$ be a bounded function. If P'' is a refinement of P, then

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

Proof. Let Q be the rectangle

$$Q = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

It suffices to prove the lemma when P'' is obtained by adjoining a single additional point to the partition of one of the component intervals of Q. Suppose, to be definite, that P is the partition $(P_1, \ldots, P_n)$ and that P'' is obtained by adjoining the point q to the partition P_1. Further, suppose that P_1 consists of the points

$$a_1 = t_0 < t_1 < \cdots < t_k = b_1$$

and that q lies interior to the subinterval $[t_{i-1}, t_i]$. We first compare the lower sums $L(f,P)$ and $L(f,P'')$. Most of the subrectangles determined by P are also subrectangles determined by P''. An exception occurs for a subrectangle determined by P of the form

$$R_S = [t_{i-1}, t_i] \times S$$

where S is one of the subrectangles of $[a_2, b_2] \times \cdots \times [a_n, b_n]$ determined by $(P_2, \ldots, P_n)$. The term involving the subrectangle R_S disappears from the lower sum and is replaced by the terms involving the two subrectangles

$$R_S^{'} = [t_{i-1}, q] \times S \quad \text{and} \quad R_S^{''} = [q, t_i] \times S,$$

which are determined by P''.

Now since $m_{R_S}(f) \leq f(x)$ for each $x \in R_S^{'}$ and for each $x \in R_S^{''}$, it follows that

$$m_{R_S}(f) \leq m_{R'_S}(f) \quad \text{and} \quad m_{R_S}(f) \leq m_{R''_S}(f).$$

Because $v(R_S) = v(R_S^{'}) + v(R_S^{''})$ by direct computation , we have

$$m_{R_S}(f)v(R_S) \leq m_{R'_S}(f)v(R_S^{'}) + m_{R''_S}(f)v(R_S^{''}).$$

Since this inequality holds for each subrectangle of the form R_S, it follows that

$$L(f,P) \leq L(f,P'').$$

A similar argument applies to show that $U(f,P'') \leq U(f,P)$. $\square$

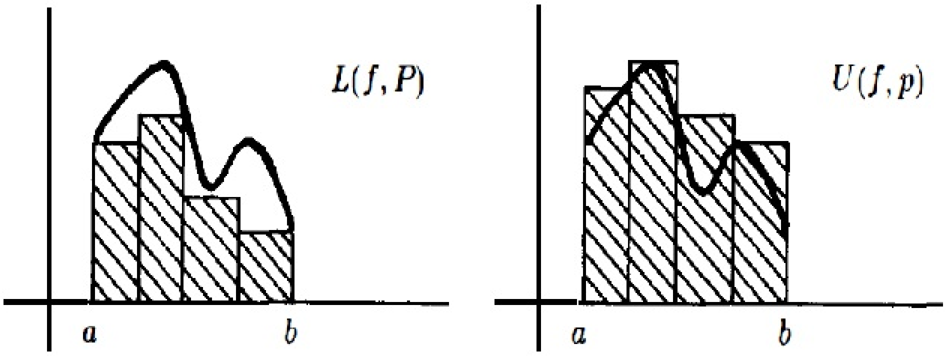


Figure 7.2.1: For one-dimensional f in example 7.2.6

Now we explore the relation between upper sums and lower sums. We have the following result:

Lemma 7.2.4. Let Q be a rectangle and $f : Q \to \mathbb{R}$ be a bounded function. If P and P' are any two partitions of Q, then

$$L(f, P) \leq U(f, P').$$

Proof. In the case where $P = P'$, the result is obvious: For any subrectangle R determined by P, we have $m_R(f) \leq M_R(f)$. Multiplying by $v(R)$ and summing gives the desired inequality.

In general, given partitions P and P' of Q , let P'' be their common refinement. Using the preceding lemma, we conclude that

$$L(f, P) \leq L(f, P'') \leq U(f, P'') \leq U(f, P').$$

□

We are now in a position to define the integral.

Definition 7.2.5. Let Q be a rectangle and $f : Q \to \mathbb{R}$ be a bounded function. As P ranges over all partitions of Q, define

$$\underline{\int_Q} f = \sup_P \{L(f, P)\} \quad \text{and} \quad \overline{\int_Q} f = \inf_P \{U(f, P)\}.$$

These numbers are called the lower integral and upper integral, respectively, of f over Q. They exist because the numbers $L(f, P)$ are bounded above by $U(f, P')$ where P' is any fixed partition of Q; and the numbers $U(f, P)$ are bounded below by $L(f, P')$. If the upper and lower integrals of f over Q are equal, we say that f is integrable over Q , and we define the integral of f over Q as the common value of the upper and lower integrals. We denote the integral of f over Q by either of the symbols

$$\int_Q f \quad \text{or} \quad \int_{x \in Q} f(x).$$

Example 7.2.6. Let $f : [a, b] \to \mathbb{R}$ be a non-negative bounded function. If P is a partition of $I = [a, b]$, then $L(f, P)$ equals the total area of a bunch of rectangles inscribed in the region between the graph of I and the x-axis, and $U(f, P)$ equals the total area of a bunch of rectangles circumscribed about this region as shown in the figure.

The lower integral represents the so-called "inner area" of this region, computed by approximating the region by inscribed rectangles, while the upper integral represents the so-called "outer area," computed by

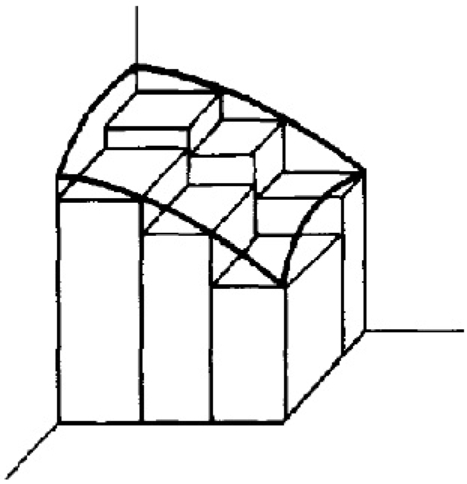


Figure 7.2.2: For two-dimensional f in example 7.2.6

approximating the region by circumscribed rectangles. If the "inner" and "outer" areas are equal, then f is integrable.

Similarly, if Q is a rectangle in $\mathbb{R}^2$ and $f : Q \to \mathbb{R}$ is non-negative and bounded, one can picture $L(f, P)$ as the total volume of a bunch of boxes inscribed in the region between the graph of f and the xy-plane, and $U(f, P)$ as the total volume of a bunch of boxes circumscribed about this region.

Example 7.2.7. Let $I = [0, 1]$. Let $f : I \to \mathbb{R}$ be defined by setting

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

We show that f is not integrable over I.

Let P be a partition of f. If R is any subinterval determined by P, then $m_R(f) = 0$ and $M_R(f) = 1$, since R contains both rational and irrational numbers. Then

$$L(f, P) = \sum_R 0.v(R) = 0, \quad \text{and} \quad U(f, P) = \sum_R 1.v(R) = 1.$$

Since P is arbitrary, it follows that the lower integral of f over I equals 0, and the upper integral equals 1 . Thus f is not integrable over I.

Theorem 7.2.8. (The Riemann condition) . Let Q be a rectangle and $f : Q \to \mathbb{R}$ is a bounded function. Then

$$\underline{\int_Q} f \leq \overline{\int_Q} f;$$

equality holds if and only if given $\epsilon > 0$, there exists a partition P of Q for which

$$U(f, P) - L(f, P) < \epsilon.$$

Proof. Let P' be a fixed partition of Q. It follows from the fact that $L(f, P) \leq U(f, P)$ for every partition P of Q, that

$$\underline{\int_Q} f \leq U(f, P').$$

Now we use the fact that P' is arbitrary to conclude that

$$\underline{\int_Q} f \leq \overline{\int_Q} f.$$

Suppose now that the upper and lower integrals are equal and let $\epsilon > 0$ be arbitrary. So, there exist a partitions P and P' so that

$$\underline{\int_Q} f - \frac{\epsilon}{2} < L(f, P) \leq \underline{\int_Q} f = \int_Q f.$$

and

$$\int_Q f = \overline{\int_Q} f \leq U(f, P') < \overline{\int_Q} f + \frac{\epsilon}{2}.$$

Let $P'' = P \cup P'$. Then both the above inequalities simultaneously hold for P''. Thus, we get

$$\underline{\int_Q} f - \frac{\epsilon}{2} < L(f, P) \leq L(f, P'') \leq \int_Q f \leq U(f, P'') \leq U(f, P) < \overline{\int_Q} f + \frac{\epsilon}{2},$$

since P'' is the common refinement of P and P'. Thus, we get

$$U(f, P'') - L(f, P'') < \epsilon.$$

Conversely, suppose the upper and lower integrals are not equal. Let

$$\epsilon = \overline{\int_Q} f - \underline{\int_Q} f > 0.$$

Let P be any partition of Q. Then

$$L(f, P) \leq \underline{\int_Q} f < \overline{\int_Q} f \leq U(f, P);$$

which implies that

$$U(f, P) - L(f, P) \leq \overline{\int_Q} f - \underline{\int_Q} f = \epsilon$$

and the Riemann condition does not hold. □

Here is an easy application of this theorem.

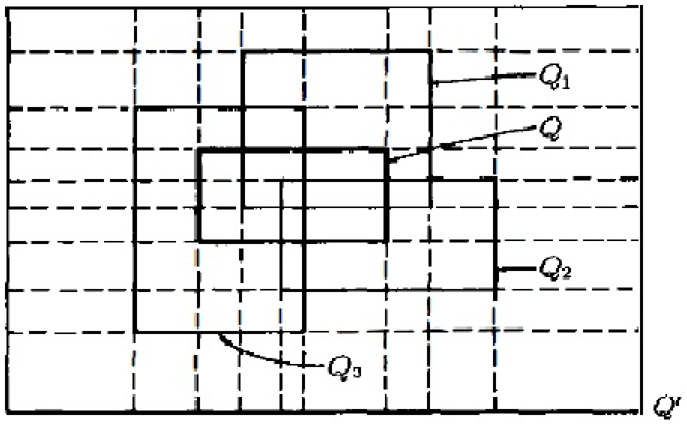
Theorem 7.2.9. Every constant function $f(x) = c$ is integrable. Indeed, if Q is a rectangle and if P is a partition of Q, then

$$\int_Q c = c.v(Q) = c \sum_R v(R),$$

where the summation extends over all subrectangles determined by P.

Proof. If R is a subrectangle determined by P, then $m_R(f) = c = M_R(f)$. It follows that

$$L(f, P) = c \sum_R v(R) = U(f, P),$$



so the Riemann condition holds trivially. Thus $\int_Q c$ exists; since it lies between $L(f, P)$ and $U(f, P)$, it must be equal to $c \sum_R v(R)$.

This result holds for any partition P. In particular, if P is the trivial partition whose only subrectangle is Q itself, then

$$\int_Q c = c.v(Q).$$

□

Corollary 7.2.10. Let Q be a rectangle in $\mathbb{R}^n$. Let $\{Q_1, \ldots, Q_k\}$ be a finite collection of rectangles that covers Q. Then

$$v(Q) \leq \sum_{i=1}^{k} v(Q_i).$$

Proof. Choose a rectangle Q' containing all the rectangles $Q_1, \ldots, Q_k$. Use the end points of the component intervals of the rectangles $Q, Q_1, \ldots, Q_k$ to define a partition P of Q'. Then each of the rectangles $Q, Q_1, \ldots, Q_k$ is a union of sub rectangles determined by P.

From the preceding theorem, we conclude that

$$v(Q) = \sum_{R \subset Q} v(R),$$

where the summation extends over all sub rectangles contained in Q. Because each such subrectangle R is contained in at least one of the rectangles $Q_1, \ldots, Q_k$, we have

$$\sum_{R \subset Q} v(R) \leq \sum_{i=1}^{k} \sum_{R \subset Q_i} v(R).$$

By the preceding theorem, we get

$$\sum_{R \subset Q_i} v(R) = v(Q_i),$$

and the corollary follows. □

In the case of $n = 1$, Q is a closed interval $[a, b]$ in $\mathbb{R}$ and we denote the integral of f over $[a, b]$ by one of the symbols

$$\int_a^b f \quad \text{or} \quad \int_{x=a}^{x=b} f(x)$$

instead of $\int_{[a,b]} f$.

Theorem 7.2.11. Let Q be a rectangle and $f, g : Q \to \mathbb{R}$ be bounded functions such that $f(x) \leq g(x)$ for $x \in Q$. Then

$$\underline{\int_Q} f \leq \underline{\int_Q} g \quad \text{and} \quad \overline{\int_Q} f \leq \overline{\int_Q} g.$$

Proof. Left as exercise. □

Few Probable Questions

1. Suppose $f : Q \to \mathbb{R}$ is continuous. Show that f is integrable over Q. Is the converse true? Justify.
2. State and prove the necessary and sufficient condition for integrability of a bounded function f, defined on a closed rectangle Q.
3. Show that any constant function f defined on a closed rectangle Q is always integrable.
4. Show that the function $f : [a, b] \to \mathbb{R}$ is not integrable over $[a, b]$ where

$$\begin{aligned} f(x) &= 0; \text{ if } x \text{ is rational} \\ &= 1; \text{ if } x \text{ is irrational} \end{aligned}$$

5. Let $I = [0, 1]^2 = [0, 1] \times [0, 1]$ and $f : I \to \mathbb{R}$ be defined by

$$\begin{aligned} f(x) &= 0; \text{ if } y \neq x \\ &= 1; \text{ if } y = x. \end{aligned}$$

   Show that f is integrable over I.

6. Let $f : \mathbb{R} \to \mathbb{R}$ be defined as

$$\begin{aligned} f(x) &= \frac{1}{q}; \text{ if } x = \frac{p}{q}, \text{ where } p \ \& \ q \text{ are positive integers having no common factor} \\ &= 0; \text{ otherwise} \end{aligned}$$

   Show that f is integrable over $[0, 1]$.

Unit 8

Course Structure

- Conditions of integrability. Integrals of $f : C \to \mathbb{R}$, $C \subset \mathbb{R}^n$ is not a rectangle, concept of Jordan measurability of a set in $\mathbb{R}$.

8.1 Introduction

Integration and measure zero sets are related in a very crucial way. We know that, in the one-dimensional case, a function f defined on a closed interval $[a, b]$ is integrable (due to Riemann) if and only if the set of discontinuities of f is of measure zero. We will try to find an analogous theorem for the multivariable case. First, we will define measure zero sets in $\mathbb{R}^n$ and then will move on to derive the necessary and sufficient condition of integrability of a bounded function f defined on a closed rectangle in connection to the measure zero sets.

Also, we so far have dealt with the integration of a bounded function f defined on a closed rectangle. We will see that, with the help of the closed rectangles we can define integrability of a bounded function, on any set, say C in $\mathbb{R}^n$. Let's explore!

Objectives

After reading this unit, you will be able to

- define measure zero sets in $\mathbb{R}^n$
- learn the characteristics of measure zero sets and see certain examples
- learn some more conditions of integrability of a bounded function f, defined on a closed rectangle Q in $\mathbb{R}^n$
- apply them in problems
- define the integration of a bounded function on any set C in $\mathbb{R}^n$, other than a closed rectangle
- learn certain related properties

8.1.1 Measure zero sets in $\mathbb{R}^n$

Definition 8.1.1. Let A be a subset of $\mathbb{R}^n$. We say that A has measure zero in $\mathbb{R}^n$ if for every $\epsilon > 0$, there is a cover $Q_1, Q_2, \ldots$ of A by countably many closed rectangles such that

$$\sum_{i=1}^{\infty} v(Q_i) < \epsilon.$$

If this inequality holds, we often say that the total volume of the rectangles $Q_1, Q_2, \ldots$ is less than ϵ.

A set with only finitely many points clearly has measure 0. If A has infinitely many points which can be arranged in a sequence $a_1, a_2, \ldots$, then A also has measure 0, since for $\epsilon > 0$, we can choose Q_i to be a closed rectangle containing a_i with

$$v(Q_i) < \frac{\epsilon}{2^i}.$$

Then,

$$\sum_{i=1}^{\infty} v(Q_i) < \sum_{i=1}^{\infty} \frac{\epsilon}{2^i} = \epsilon.$$

We derive some properties of sets of measure zero.

Theorem 8.1.2.

1. If $B \subset A$ and A has measure zero in $\mathbb{R}^n$, then so does B.
2. Let A be the union of the countable collection of sets $A_1, A_2, \ldots$ If each A_i has measure zero in $\mathbb{R}^n$, then so does A.
3. A set A has measure zero in $\mathbb{R}^n$ if and only if for every $\epsilon > 0$, there is a countable covering of A by open rectangles $\text{Int}Q_1, \text{Int}Q_2, \ldots$ such that

$$\sum_{i=1}^{\infty} v(Q_i) < \epsilon.$$

4. If Q is a rectangle in $\mathbb{R}^n$, then $\text{Bd}Q$ has measure zero in $\mathbb{R}^n$ but Q does not ($\text{Bd}Q$ is the boundary of Q).

Proof.

1. Let $\epsilon > 0$. Since A is measure zero set, so for the given ϵ, there is a cover $Q_1, Q_2, \ldots$ of A by countably many closed rectangles such that

$$\sum_{i=1}^{\infty} v(Q_i) < \epsilon.$$

Since $B \subset A$, so B satisfies the definition of zero measure in $\mathbb{R}^n$.

2. To prove 2,we cover the set A_j, for each j, by countably many rectangles

$$Q_{1j}, Q_{2j}, \ldots$$

of total volume less than $\epsilon/2^j$. Then the collection of rectangles $\{Q_{ij}\}$ is countable, that covers A, having total volume

$$\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} v(Q_{ij}) < \sum_{j=1}^{\infty} \frac{\epsilon}{2^j} = \epsilon.$$

Hence A is of measure zero.

3. If the open rectangles $\text{Int}Q_1, \text{Int}Q_2, \ldots$ cover A, then so do the rectangles $Q_1, Q_2, \ldots$ Thus the given condition implies that A has measure zero. Conversely, suppose A has measure zero. Cover A by rectangles $Q_1^{'}, Q_2^{'}, \ldots$ of total volume less than $\epsilon/2$. For each i, choose a rectangle Q_i such that
$$Q_i^{'} \subset \text{Int}\, Q_i \quad \text{and} \quad v(Q_i) \leq 2v(Q_i^{'}).$$
This is possible because $v(Q)$ is a continuous function of the end points of the component intervals of Q. Then the open rectangles $\text{Int}Q_1, \text{Int}Q_2, \ldots$ cover A, and $\sum v(Q_i) < \epsilon$.

4. Let
$$Q = [a_1, b_1] \times \cdots \times [a_n, b_n].$$
The subset of Q consisting of those points x of Q for which $x_i = a_i$ is called one of the ith faces of Q. The other ith face consists of those x for which $x_i = b_i$. Each face of Q has measure zero in $\mathbb{R}^n$; for instance, the face for which $x_i = a_i$ can be covered by the single rectangle
$$[a_1, b_1] \times \cdots \times [a_i, a_i + \delta] \times \cdots \times [a_n, b_n],$$
whose volume may be made as small as desired by taking δ small. Now $\text{Bd}Q$ is the union of the faces of Q, which are finite in number. Therefore $\text{Bd}Q$ has measure zero in $\mathbb{R}^n$.

   Now we suppose Q has measure zero in $\mathbb{R}^n$, and derive a contradiction . Set $\epsilon = v(Q)$. By 3, we can cover Q by open rectangles $\text{Int}Q_1, \text{Int}Q_2, \ldots$ with $\sum v(Q_i) < \epsilon$. Because Q is compact, we can cover Q by finitely many of these open sets, say $\text{Int}Q_1, \text{Int}Q_2, \ldots \text{Int}Q_k$. But
$$\sum_{i=1}^{k} v(Q_i) < \epsilon,$$
which is a contradiction to a previous corollary we read in the previous unit.

□

By the third point of the above theorem, we can easily say that open rectangles may be used instead of closed rectangles in the definition of measure zero sets.

Definition 8.1.3. Let A be a subset of $\mathbb{R}^n$. We say that A has measure zero in $\mathbb{R}^n$ if for every $\epsilon > 0$, there is a cover $Q_1, Q_2, \ldots Q_n$ of A by finitely many closed rectangles such that
$$\sum_{i=1}^{n} v(Q_i) < \epsilon.$$

If A has content 0, then A clearly has measure 0. Again, open rectangles could be used instead of closed rectangles in the definition.

Theorem 8.1.4. If $a < b$, then $[a, b] \subset \mathbb{R}$ does not have content 0. In fact, if $Q_1, Q_2, \ldots Q_n$ is a finite cover of $[a, b]$ by closed intervals, then
$$\sum_{i=1}^{n} v(Q_i) \geq b - a.$$

Proof. Clearly we can assume that each $Q_i \subset [a, b]$. Let $a = t_0 < t_1 < t_2 < \cdots < t_k = b$ be all endpoints of all Q_i. Then, each $v(Q_i)$ is the sum of certain $t_j - t_{j-1}$. Moreover, each $[t_{j-1}, t_j]$ lies in at least one Q_i (namely, any one which contains an interior point of $[t_{j-1}, t_j]$), so that
$$\sum_{i=1}^{n} v(Q_i) \geq \sum_{j=1}^{k} (t_j - t_{j-1}) = b - a.$$

□

If $a < b$, it is also true that $[a, b]$ does not have measure 0. This follows from

Theorem 8.1.5. If A is compact and has measure 0, then A has content 0.

Proof. Let $\epsilon > 0$. Since A has measure 0, there is a cover $\{Q_1, Q_2, \ldots\}$ of A by open rectangles such that

$$\sum_{i=1}^{\infty} v(Q_i) < \epsilon.$$

Since A is compact, a finite subcover $\{Q_1, Q_2, \ldots Q_n\}$ of A for which

$$\sum_{i=1}^{n} v(Q_i) < \epsilon.$$

□

The conclusion of the above theorem is false if A is not compact. For example, let A be the set of rational numbers between 0 and 1; then A has measure 0. Suppose, however, that $\{[a_1, b_1], \ldots, [a_n, b_n]\}$ covers A. Then A is contained in the closed set $[a_1, b_1] \cup \cdots \cup [a_n, b_n]$, and hence $[0, 1] \subset [a_1, b_1] \cup \cdots \cup [a_n, b_n]$. Thus, we get

$$\sum_{i=1}^{n} (b_i - a_i) \geq 1$$

for any such cover, and consequently A does not have content 0.

Recall that $o(f, x)$ denotes the oscillation of f at x.

Lemma 8.1.6. Let Q be a closed rectangle and let $f : Q \to \mathbb{R}$ be a bounded function such that $o(f, x) < \epsilon$ for all $x \in Q$. Then there is a partition P of Q such that $U(f, P) - L(f, P) < \epsilon.v(Q)$.

Proof. For each $x \in A$, there is a closed rectangle Q_x containing x in its interior, such that $M_{Q_x}(f) - m_{Q_x}(f) < \epsilon$. Since Q is compact, there exists a finite number $Q_{x_1}, \ldots, Q_{x_n}$ of the sets Q_x that cover Q. Let P be a partition for Q such that each subrectangle S of P is contained in some Q_{x_i}. Then $M_S(f) - m_S(f) < \epsilon$ for each subrectangle S of P, so that

$$U(f, P) - L(f, P) = \sum_S [M_S(f) - m_S(f)].v(S) < \epsilon.v(A).$$

□

Theorem 8.1.7. Let Q be a closed rectangle and let $f : Q \to \mathbb{R}$ be a bounded function. Let $B = \{x : \ f$ is not continuous at $x\}$. Then f is integrable if and only if B is a set of measure 0.

Proof. Suppose first that B has measure 0. Let $\epsilon > 0$ and let $B_\epsilon = \{x : \ o(f, x) \geq \epsilon\}$. Then $B_\epsilon \subset B$, so that B_ϵ has measure zero. Since B_ϵ is compact, it has content zero. Thus, there exist a finite collection $Q_1, \ldots, Q_n$ of closed rectangles, whose interiors cover B_ϵ, such that $\sum\limits_{i=1}^{n} v(Q_i) < \epsilon$. Let P be a partition of Q such that every subrectangle S of P is in one of two groups

1. S_1, which consists of subrectangles S, such that $S \subset Q_i$ for some i.
2. S_2, which consists of subrectangles S with $S \cap B_\epsilon = \emptyset$.

Let $|f(x)| < M$ for $x \in Q$. Then $M_S(f) - m_S(f) < 2M$ for every S. Hence

$$\sum_{S \subset S} [M_S(f) - m_S(f)].v(S) < 2M \sum_{i=1}^{n} v(Q_i) < 2M\epsilon.$$

Now, if $S \in S_2$, then $o(f, x) < \epsilon$ for $x \in S$. The previous lemma implies that there is a refinement P' of P such that

$$\sum_{S' \subset S} [M_{S'}(f) - m_{S'}(f)].v(S') < \epsilon.v(S)$$

for $S \in S_2$. Then

$$\begin{aligned} U(f, P') - L(f, P') &= \sum_{S' \subset S \in S_1} [M_{S'}(f) - m_{S'}(f)].v(S') + \sum_{S' \subset S \in S_2} [M_{S'}(f) - m_{S'}(f)].v(S') \\ &< 2M\epsilon + \sum_{S \in S_2} \epsilon.v(S) \\ &\leq 2M\epsilon + \epsilon.v(Q). \end{aligned}$$

Since M and $v(Q)$ are fixed, this shows that we can find a partition P' with $U(f, P') - L(f, P')$ as small as desired. Thus f is integrable.

Suppose, conversely, that f is integrable. Since $B = B_1 \cup B_{1/2} \cup B_{1/3} \cup \cdots$, it suffices to prove that each $B_{1/n}$ has measure 0. In fact we will show that each $B_{1/n}$ has content zero (since $B_{1/n}$ is compact, this is actually equivalent).

Let $\epsilon > 0$, and let P be a partition of Q such that

$$U(f, P) - L(f, P) < \epsilon/n.$$

Let $\mathcal{S}$ be the collection of subrectangles S of P which intersect $B_{1/n}$. Then $\mathcal{S}$ is a cover of $B_{1/n}$. Now, if $S \in \mathcal{S}$, then $M_S(f) - m_S(f) \geq 1/n$. Thus

$$\begin{aligned} \frac{1}{n} \sum_{S \in \mathcal{S}} v(S) &\leq \sum_{S \in \mathcal{S}} [M_S(f) - m_S(f)].v(S) \\ &\leq \sum_{S} [M_S(f) - m_S(f)].v(S) \\ &< \frac{\epsilon}{n}, \end{aligned}$$

and so

$$\sum_{S \in \mathcal{S}} v(S) < \epsilon.$$

$\square$

Exercise 8.1.8. 1. Show that any finite set in $\mathbb{R}^n$ has measure zero.

8.1.2 Integrals of functions on sets other than rectangles

We have thus far dealt only with the integrals of functions over rectangles. Integrals over other sets are easily reduced to this type. If $C \in \mathbb{R}^n$, the characteristic function χ_C of C is defined by

$$\begin{aligned} \chi_C(x) &= 0, \ c \notin C, \\ &= 1, \ x \in C. \end{aligned}$$

If $C \subset Q$ for some closed rectangle Q and $f : A \to \mathbb{R}$ bounded, then $\int_C f$ is defined as $\int_A f.\chi_C$ is integrable. This certainly occurs if f and χ_C are integrable.

Theorem 8.1.9. The function $\chi_C : Q \to \mathbb{R}$ i8 integrable if and only if the boundary of C has measure zero (and hence content zero).

Proof. If x is in the interior of C, then there is an open rectangle U with $x \in U \subset C$. Thus, $\chi_C = 1$ on U and χ_C is clearly continuous at x. Similarly, if x is in the exterior of C, there is an open rectangle U with $x \in U \subset \mathbb{R}^n \setminus C$. Hence $\chi_C = 0$ on U and χ_C is continuous at x. Finally, if x is in the boundary of C, then for every open rectangle U containing x, there is $y_1 \in U \cap C$, so that $\chi_C(y_1) = 1$ and there is $y_2 \in U \cap (\mathbb{R}^n \setminus C)$, so that $\chi_C(y_2) = 0$. Hence χ_C is not continuous at x. Thus, $\{x : \ \chi_C$ is not continuous at $x\}$ =boundary of C and the result follows by the previous theorem. □

A bounded set C whose boundary has measure 0 is called Jordan-measurable. The integral $\int_C 1$ is called the n-dimensional content of C, or the n-dimensional volume of C. Naturally one-dimensional volume is often called length, and two-dimensional volume, area.

Few Probable Questions

1. Define measure zero set in $\mathbb{R}^n$. Show that a countable set in $\mathbb{R}^n$ has measure zero.
2. Deduce a necessary and sufficient condition for a bounded function defined on a closed rectangle to be integrable.
3. Define content zero sets. Show that a content zero set is of measure zero.
4. Deduce a necessary and sufficient condition for a bounded function defined on a bounded set C of $\mathbb{R}^n$ to be integrable.

Unit 9

Course Structure

- Fubini's theorem for integral of $f : A \times B \to \mathbb{R},\ A \subset \mathbb{R}^n,\ B \subset \mathbb{R}^n$ are closed rectangles, Fubini's theorem for $f : C \to \mathbb{R},\ C \subset A \times B$

9.1 Introduction

Given that a function $f : Q \to \mathbb{R}$ is integrable, how does one evaluate its integral?

Even in the case of a function $f : [a, b] \to \mathbb{R}$ of a single variable, the problem is not easy. One tool is provided by the fundamental theorem of calculus, which is applicable when f is continuous. This theorem is familiar to you from single-variable analysis. We restate it over here.

Theorem 9.1.1. (Fundamental theorem of calculus).

1. If f is continuous on $[a, b]$, and if
$$F(x) = \int_a^x f$$
for $x \in [a, b]$, then $F'(x)$ exists and equals $f(x)$.

2. If f is continuous on $[a, b]$, and if g is a function such that $g'(x) = f(x)$ for $x \in [a, b]$, then
$$\int_a^b f = g(b) - g(a).$$

To summarise, we need to find the antiderivative of f, that is, a function g such that $g' = f$. For the n-dimensional case, we use the Fubini's theorem. Fubini's theorem, named after Guido Fubini, is a result which gives conditions under which it is possible to compute a double integral using iterated integrals. As a consequence it allows the order of integration to be changed in iterated integrals.

Objectives

After reading this unit, you will be able to

- learn Fubini's theorem and its consequences

9.2 Fubini's Theorem

The problem of calculating integrals is solved, in some sense, by Fubini's theorem, which reduces the computation of integrals over a closed rectangle in $\mathbb{R}^n$, $n > 1$, to the computation of integrals over closed intervals in $\mathbb{R}$.

The idea behind the theorem is best illustrated for a positive continuous function $f : [a, b] \times [c, d] \to \mathbb{R}$. Let $t_0, t_1, \ldots, t_n$ be a partition of $[a, b]$ and divide $[a, b] \times [c, d]$ into n strips by means of the line segments $\{t_i\} \times [c, d]$. If g_x is defined by $g_x(y) = f(x, y)$, then the area of the region under the graph of f and above $\{x\} \times [c, d]$ is

$$\int_c^d g_x = \int_c^d f(x, y)dy.$$

The volume of the region under the graph of f and above $[t_{i-1}, t_i] \times [c, d]$ is therefore approximately equal to $(t_i - t_{i-1}) \cdot \int_c^d f(x, y)dy$ for any $x \in [t_{i-1}, t_i]$. Thus,

$$\int_{[a,b]\times[c,d]} f = \sum_{i=1}^{n} \int_{[t_{i-1},t_i]\times[c,d]} f$$

is approximately

$$\sum_{i=1}^{n} (t_i - t_{i-1}) \cdot \int_c^d f(x, y)dy,$$

with x_i in $[t_{i-1}, t_i]$. On the other hand, sums similar to these appear in the definition of $\int_a^b \left(\int_c^d f(x, y)dy \right) dx$. Thus, if h is defined by

$$h(x) = \int_a^b g_x = \int_c^d f(x, y)dy,$$

it is reasonable to hope that h is integrable on $[a, b]$ and that

$$\int_{[a,b]\times[c,d]} f = \int_a^b h = \int_a^b \left(\int_c^d f(x, y)dy \right) dx.$$

This will indeed turn out to be true when f is continuous, but in the general case difficulties arise.

We will require a bit of terminology. If $f : A \to \mathbb{R}$ is a bounded function on a closed rectangle, then, whether or not f is integrable, the least upper bound of all lower sums, and the greatest lower bound of all upper sums, both exist. They are called the lower and upper integrals of f on A, and denoted by

$$L \int_A f \text{ and } L \int_A f$$

respectively.

Theorem 9.2.1. (Fubini's Theorem) Let $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ be closed rectangles, and let $f : A \times B \to \mathbb{R}$ be integrable. For $x \in A$ let $g_x : B \to \mathbb{R}$ be defined by $g_x(y) = f(x, y)$ and let

$$\mathcal{L}(x) = L \int_B g_x = L \int_B f(x, y)$$
$$\mathcal{U}(x) = U \int_B g_x = U \int_B f(x, y).$$

Then $\mathcal{L}$ and $\mathcal{U}$ are integrable on A and

$$\int_{A\times B} f = \int_A \mathcal{L} = \int_A \left(L \int_B f(x,y)dy \right) dx,$$
$$\int_{A\times B} f = \int_A \mathcal{U} = \int_A \left(U \int_B f(x,y)dy \right) dx.$$

(The integrals on the right side are called iterated integrals for f.)

Proof. Let P_A be a partition of A and P_B a partition of B. Together they give a partition P of $A \times B$ for which any subrectangle S is of the form $S_A \times S_B$, where S_A is a subrectangle of the partition P_A, and S_B is a subrectangle of the partition P_B. Thus

$$\begin{aligned} L(f,P) = \sum_S m_S(f) \cdot v(S) &= \sum_{S_A, S_n} m_{S_A \times S_B}(f) \cdot v\left(S_A \times S_B\right) \\ &= \sum_{S_A} \left(\sum_{S_B} m_{S_A \times S_B}(f) \cdot v\left(S_B\right) \right) \cdot v\left(S_A\right). \end{aligned}$$

Now, if $x \in S_A$, then clearly $m_{S_A \times S_B}(f) \leq m_{S_B}\left(g_x\right)$. Consequently, for $x \in S_A$ we have

$$\sum_{S_B} m_{S_A \times S_B}(f) \cdot v\left(S_B\right) \leq \sum_{S_B} m_{S_B}\left(g_x\right) \cdot v\left(S_B\right) \leq \mathbf{L} \int_B g_x = \mathfrak{L}(x).$$

Therefore

$$\sum_{S_A} \left(\sum_{S_B} m_{S_A \times S_B}(f) \cdot v\left(S_B\right) \right) \cdot v\left(S_A\right) \leq L\left(\mathfrak{L}, P_A\right).$$

We thus obtain

$$L(f,P) \leq L\left(\&, P_A\right) \leq U\left(\&, P_A\right) \leq U\left(\mathcal{U}, P_A\right) \leq U(f,P),$$

where the proof of the last inequality is entirely analogous to the proof of the first. Since f is integrable, $\sup\{L(f,P)\} = \inf\{U(f,P)\} = \int_{A\times B} f$. Hence

$$\sup\left\{L\left(\mathscr{L}, P_A\right)\right\} = \inf\left\{U\left(\&, P_A\right)\right\} = \int_{A\times B} f.$$

In other words, $\mathcal{L}$ is integrable on A and $\int_{A\times B} f = \int_A \mathcal{L}$. The assertion for U follows similarly from the inequalities

$$L(f,P) \leq L\left(\&, P_A\right) \leq L\left(\sqcap, P_A\right) \leq U\left(\mathcal{U}, P_A\right) \leq U(f,P).$$

$\square$

Remarks. 1. A similar proof shows that

$$\int_{\mathbf{A}\times \boldsymbol{B}} f = \int_{\boldsymbol{B}} \left(\mathbf{L} \int_{\boldsymbol{A}} f(x,y)dx \right) dy = \int_{\boldsymbol{B}} \left(\mathbf{U} \int_{\boldsymbol{A}} f(x,y)dx \right) dy.$$

These integrals are called iterated integrals for f in the reverse order from those of the theorem. As several problems show, the possibility of interchanging the orders of iterated integrals has many consequences. 2. In practice it is often the case that each g_x is integrable, so that $\int_{A\times B} f = \int_A \left(\int_B f(x,y)dy\right) dx$. This certainly occurs if f is continuous. 3. The worst irregularity commonly encountered is that g_x is not integrable for a finite number of $x \in A$. In this case $\mathcal{L}(x) = \int_B f(x,y)dy$ for all but these finitely many x. Since $\int_A \mathcal{L}$ remains unchanged if & is redefined at a finite number of points, we can still write $\int_{A\times B} f = \int_A \left(\int_B f(x,y)dy\right) dx$, provided that $\int_B f(x,y)dy$ is defined arbitrarily, say as 0 , when it does not exist. 4. There are cases when this will not work and Theorem 3-10 must be used as stated. Let $f : [0,1] \times [0,1] \to \mathbf{R}$ be defined by

$$f(x,y) = \begin{cases} 1 & \text{if } x \text{ is irrational,} \\ 1 & \text{if } x \text{ is rational and } y \text{ is irrational,} \\ 1 - 1/q & \text{if } x = p/q \text{ in lowest terms and } y \text{ is} \\ & \text{rational.} \end{cases}$$

Then f is integrable and $\int_{[0,1]\times[0,1]} f = 1$. Now $\int_0^1 f(x,y)dy = 1$ if x is irrational, and does not exist if x is rational. Therefore h is not integrable if $h(x) = \int_0^1 f(x,y)dy$ is set equal to 0 when the integral does not exist. 5. If $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$ and $f : A \to \mathbf{R}$ is sufficiently nice, we can apply Fubini's theorem repeatedly to obtain

$$\int_A f = \int_{a_n}^{b_n} \left(\cdots \left(\int_{a_1}^{b_1} f\left(x^1, \ldots, x^n\right) dx^1 \right) \cdots \right) dx^n.$$

6. If $C \subset A \times B$, Fubini's theorem can be used to evaluate $\int_{cf}$, since this is by definition $\int_{A\times B} \chi cf$. Suppose, for example, that

$$C = [-1,1] \times [-1,1] - \{(x,y) : |(x,y)| < 1\}.$$

Then

$$\int_C f = \int_{-1}^{1} \left(\int_{-1}^{1} f(x,y) \cdot \chi_C(x,y)dy \right) dx.$$

Unit 10

Course Structure

- Formula for change of variables in an integral in $\mathbb{R}^n$.

In this section we introduce a tool of extreme importance in the theory of integration. 3-11 Theorem. Let $A \subset \mathbf{R}^n$ and let $\mathcal{O}$ be an open cover of A. Then there is a collection Φ of C^∞ functions φ defined in an open set containing A, with the following properties: (1) For each $x \in A$ we have $0 \leq \varphi(x) \leq 1$. (2) For each $x \in A$ there is an open set V containing x such that all but finitely many $\varphi \in \Phi$ are 0 on V. (3) For each $x \in A$ we have $\Sigma_{\varphi \in \uparrow \varphi}(x) = 1$ (by (2) for each x this sum is finite in some open set containing x). (4) For each $\varphi \in \Phi$ there is an open set U in $\mathcal{O}$ such that $\varphi = 0$ outside of some closed set contained in U. (A collection Φ satisfying (1) to (3) is called a C^∞ partition of unity for A. If Φ also satisfies (4), it is said to be subordinate to the cover $\mathcal{O}$. In this chapter we will only use continuity of the functions φ.) Proof. Case 1. A is compact. Then a finite number $U_1, \ldots, U_n$ of open sets in $\odot$ cover A. It clearly suffices to construct a partition of unity subordinate to the cover $\{U_1, \ldots, U_n\}$. We will first find compact sets $D_i \subset U_i$ whose interiors cover A. The sets D_i are constructed inductively as follows. Suppose that $D_1, \ldots, D_k$ have been chosen so that interior $D_1, \ldots$, interior D_k, $U_{k+1}, \ldots, U_n\}$ covers A. Let

$$C_{k+1} = A - (\text{ int } D_1 \cup \cdots \cup \text{int } D_k \cup U_{k+2} \cup \cdots \cup U_n).$$

Then $C_{k+1} \subset U_{k+1}$ is compact. Hence (Problem 1-22) we can find a compact set D_{k+1} such that

$$C_{k+1} \subset \text{ interior } D_{k+1} \text{ and } D_{k+1} \subset U_{k+1}.$$

Having constructed the sets $D_1, \ldots, D_n$, let ψ_i be a nonnegative C^∞ function which is positive on D_i and 0 outside of some closed set contained in U_i (Problem 2-26). Since $\{D_1, \ldots, D_n\}$ covers A, we have $\psi_1(x) + \cdots + \psi_n(x) > 0$ for all x in some open set U containing A. On U we can define

$$\varphi_i(x) = \frac{\psi_i(x)}{\psi_1(x) + \cdots + \psi_n(x)}.$$

If $f : U \to [0,1]$ is a C^∞ function which is 1 on A and 0 outside of some closed set in U, then $\Phi = \{f \cdot \varphi_1, \ldots, f \cdot \varphi_n\}$ is the desired partition of unity.

Case 2. $A = A_1 \cup A_2 \cup A_3 \cup \cdots$, where each A_i is compact and $A_i \subset$ interior A_{i+1}.

For each i let $\mathcal{O}_i$ consist of all $U \cap$ (interior $A_{i+1} - A_{i-2}$) for U in $\mathcal{O}$. Then $\mathcal{O}_i$ is an open cover of the compact set $B_i = A_i -$ interior A_{i-1}. By case 1 there is a partition of unity Φ_i for B_i, subordinate to Θ_i. For each $x \in A$ the sum

$$\sigma(x) = \sum_{\varphi \in \text{ Ail}^{\text{ all }} i} \varphi(x)$$

is a finite sum in some open set containing x, since if $x \in A_i$ we have $\varphi(x) = 0$ for $\varphi \in \Phi_j$ with $j \geq i+2$. For each φ in each Φ_i, define $\varphi'(x) = \varphi(x)/\sigma(x)$. The collection of all φ' is the desired partition of unity. Case 3. A is open. Let $A_i = \{x \in A : |x| \leq i$ and distance from x to boundary $A \geq 1/i\}$, and apply case 2 . Case 4. A is arbitrary. Let B be the union of all U in O. By case 3 there is a partition of unity for B; this is also a partition of unity for A. An important consequence of condition (2) of the theorem should be noted. Let $C \subset A$ be compact. For each $x \in C$ there is an open set V_x containing x such that only finitely many $\varphi \in \Phi$ are not 0 on V_x. Since C is compact, finitely many such V_x cover C. Thus only finitely many $\varphi \in \Phi$ are not 0 on C.

One important application of partitions of unity will illustrate their main role - piecing together results obtained locally. An open cover $\mathcal{O}$ of an open set $A \subset \mathbf{R}^n$ is admissible if each $U \in \mathcal{O}$ is contained in A. If Φ is subordinate to θ, $f : A \to \mathbf{R}$ is bounded in some open set around each point of A, and $\{x : f$ is discontinuous at $x\}$ has measure 0 , then each $\int_A \varphi \cdot |f|$ exists. We define f to be integrable (in the extended sense) if $\Sigma_{\varphi \in \Phi} \int_A \varphi \cdot |f|$ converges (the proof of Theorem 3-11 shows that the φ 's may be arranged in a sequence). This implies convergence of $\Sigma_{\varphi \in \Phi} \left|\int_A \varphi \cdot f\right|$, and hence absolute convergence of $\Sigma_{\varphi \in \Phi} \int_A \varphi \cdot f$, which we define to be $\int_A f$. These definitions do not depend on $\mathcal{O}$ or Φ (but see Problem 3-38). 3-12 Theorem. (1) If Ψ is another partition of unity, subordinate to an admissible cover $\mathcal{O}'$ of A, then $\Sigma_{\psi \in \Psi} \int_A \psi \cdot |f|$ also converges, and

$$\sum_{\varphi \in \Phi} \int_A \varphi \cdot f = \sum_{\nu \in \Psi} \int_A \psi \cdot f$$

(2) If A and f are bounded, then f is integrable in the extended sense. (3) If A is Jordan-measurable and f is bounded, then this definition of $\int_A f$ agrees with the old one. Proof (1) Since $\varphi \cdot f = 0$ except on some compact set C, and there are only finitely many ψ which are non-zero on C, we can write

$$\sum_{\varphi \in \Phi} \int_A \varphi \cdot f = \sum_{\varphi \in \Phi} \int_A \sum_{\psi \in \Psi} \psi \cdot \varphi \cdot f = \sum_{\varphi \in \Phi} \sum_{\psi \in \Psi} \int_\Lambda \psi \cdot \varphi \cdot f$$

This result, applied to $|f|$, shows the convergence of $\Sigma_{\varphi \in \Phi} \Sigma_{\psi \in \Psi} \int_A \psi \cdot \varphi \cdot |f|$, and hence of $\Sigma_{\varphi \in \Phi} \Sigma_{\psi \in \Psi} \left|\int_A \psi \cdot \varphi \cdot f\right|$. This absolute convergence justifies interchanging the order of summation in the above equation; the resulting double sum clearly equals $\Sigma_{\psi \in \Psi} \int_{\boldsymbol{A}} \psi \cdot f$. Finally, this result applied to $|f|$ proves convergence of $\Sigma_{\psi \in \Psi} \int_A \psi \cdot |f|$. (2) If A is contained in the closed rectangle B and $|f(x)| \leq M$ for $x \in A$, and $F \subset \Phi$ is finite, then

$$\sum_{\varphi \in F} \int_A \varphi \cdot |f| \leq \sum_{\varphi \in F} M \int_A \varphi = M \int_A \sum_{\varphi \in F} \varphi \leq M v(B),$$

since $\Sigma_{\varphi \in F} \varphi \leq 1$ on A.

(3) If $\varepsilon > 0$ there is (Problem 3-22) a compact Jordan-measurable $C \subset A$ such that $\int_{A-C} 1 < \varepsilon$. There are only finitely many $\varphi \in \Phi$ which are non-zero on C. If $F \subset \Phi$ is any finite collection which includes these, and $\int_A f$ has its old meaning, then

$$\left|\int_{\boldsymbol{A}} f - \sum_{\varphi \in \in_F} \int_{\boldsymbol{A}} \varphi \cdot f\right| \leq \int_{\boldsymbol{A}} \left| f - \sum_{\varphi \in \mathcal{E}_F} \varphi \cdot f\right|$$

$$\leq M \int_A \left(1 - \sum_{\varphi \in \mathcal{F}_F} \varphi\right)$$

$$= M \int_{\boldsymbol{A}} \sum_{\varphi \in \Phi - F} \varphi \leq M \int_{A-C} 1 \leq M\varepsilon.$$

Unit 11

Course Structure

- Interval Arithmetic: Interval numbers, arithmetic operations on interval numbers,
- Distance between intervals, two level interval numbers

11.1 Introduction

Interval arithmetic is the arithmetic of quantities that lie within specified ranges (i.e., intervals) instead of having definite known values. Interval arithmetic can be especially useful when working with data that is subject to measurement errors or uncertainties. It can be considered a rigorous version of significance arithmetic (a.k.a., automatic precision control).

Interval arithmetic, interval mathematics, interval analysis, or interval computation, is a method developed by mathematicians since the 1950s and 1960s, as an approach to putting bounds on rounding errors and measurement errors in mathematical computation and thus developing numerical methods that yield reliable results. Very simply put, it represents each value as a range of possibilities. For example, instead of estimating the height of someone using standard arithmetic as 2.0 metres, using interval arithmetic we might be certain that that person is somewhere between 1.97 and 2.03 metres.

This concept is suitable for a variety of purposes. The most common use is to keep track of and handle rounding errors directly during the calculation and of uncertainties in the knowledge of the exact values of physical and technical parameters. The latter often arise from measurement errors and tolerances for components or due to limits on computational accuracy. Interval arithmetic also helps find reliable and guaranteed solutions to equations (such as differential equations) and optimization problems.

Mathematically, instead of working with an uncertain real x we work with the two ends of the interval $[a, b]$ that contains x. In interval arithmetic, any variable x lies between a and b, or could be one of them. A function f when applied to x is also uncertain. In interval arithmetic f produces an interval $[c, d]$ that is all the possible values for $f(x)$ for all $x \in [a, b]$.

Objectives

After reading this unit you will be able to

- define interval numbers

- define set operations on intervals numbers and see certain examples related to them
- define arithmetic operations on intervals numbers and see certain examples related to them
- define algebraic properties of interval numbers
- define distance between intervals

11.2 Interval Number System

We are familiar with the closed intervals in the real line, which is denoted by

$$[a, b] = \{x \in \mathbb{R} : \; a \leq x \leq b\}.$$

Here, we will mainly refer to the closed intervals as intervals.

We will denote the endpoints of an interval I as $\underline{I}$ and $\overline{I}$, where these both represent the lower and upper endpoints respectively, that is,

$$I = [\underline{I}, \overline{I}]$$

and two intervals I and J are said to be equal if they are the same sets, that is

$$I = J \;\; \& \;\; \underline{I} = \underline{J}, \;\; \overline{I} = \overline{J}.$$

We say that an interval I is *degenerate* if $\underline{I} = \overline{I}$. Such an interval contains a single real number x. By convention, we agree to identify a degenerate interval $[x, x]$ with the real number x.

11.2.1 Certain Important Definitions

The intersection of two intervals I and J is empty if either $\overline{J} < \underline{I}$ or $\overline{I} < \underline{J}$. In this case, we let $\emptyset$ denote the empty set and write

$$I \cap J = \emptyset,$$

which indicates that I and J have no points in common. We may otherwise define the intersection $I \cap J$ as the interval

$$\begin{aligned} I \cap J &= \{z : \;\; z \in I \;\; \& \;\; z \in J\} \\ &= [\max\{\underline{I}, \underline{J}\}, \min\{\overline{I}, \overline{J}\}]. \end{aligned}$$

In this latter case, the union of I and J is also an interval

$$\begin{aligned} I \cup J &= \{z : \;\; z \in I \;\text{ or }\; z \in J\} \\ &= [\min\{\underline{I}, \underline{J}\}, \max\{\overline{I}, \overline{J}\}]. \end{aligned}$$

In general, the union of two intervals is not an interval. However, the interval hull of two intervals, defined by

$$I \underline{\cup} J = [\min\{\underline{I}, \underline{J}\}, \max\{\overline{I}, \overline{J}\}],$$

is always an interval and can be used in interval computations. We have

$$I \cup J \subseteq I \underline{\cup} J,$$

for any two intervals I and J.

Example 11.2.1. If $I = [-1, 0]$ and $J = [1, 2]$, then $I \underline{\cup} J = [-1, 2]$. $I \cup J$ is a disconnected set and hence is not an interval. But this is not the case if we consider $I \underline{\cup} J$ and $I \cup J$ is still a subset of $I \underline{\cup} J$.

Intersection plays a key role in interval analysis. If we have two intervals containing a result of interest — regardless of how they were obtained — then the intersection, which may be narrower, also contains the result.

Example 11.2.2. Suppose two people make independent measurements of the same physical quantity q. One finds that $q = 10.3$ with a measurement error less than 0.2. The other finds that $q = 10.4$ with an error less than 0.2. We can represent these measurements as the intervals $I = [10.1, 10.5]$ and $J = [10.2, 10.6]$, respectively. Since q lies in both, it also lies in $I \cup J = [10.2, 10.5]$. An empty intersection would imply that at least one of the measurements is wrong.

Definition 11.2.3. 1. As the name suggests, the **width** of an interval I is defined as

$$w(I) = \overline{I} - \underline{I}.$$

2. The **absolute value** of I, denoted as $|I|$, is the maximum of the absolute values of its endpoints

$$|I| = \max\{|\underline{I}|, |\overline{I}|\}.$$

Note that, $|x| \leq |I|$ for every $x \in I$.

3. The **midpoint** of I is given by

$$m(I) = \frac{1}{2}(\underline{I} + \overline{I}).$$

Example 11.2.4. Let $I = [0, 2]$ and $J = [-1, 1]$. Then the intersection and union of I and J are the intervals

$$I \cap J = [0, 1], \quad I \cup J = [-1, 2].$$

We have, $w(I) = w(J) = 2$ and

$$|I| = 2, \quad \& \quad |J| = 1.$$

The midpoint of I and J are 1 and 0 respectively.

The real numbers are ordered by the relation $<$. A corresponding relation can be defined for the intervals as follows

$$I < J \implies \overline{I} < \underline{J}.$$

For example, $[3, 4] < [6, 8]$ and we also have the transitivity relation which says that

$$A < B \;\&\; B < C \implies A < C.$$

We can also define $I > 0$ and $I < 0$. That is, $I > 0$ if $x > 0$ for all $x \in I$ and $I < 0$ if $x < 0$ for all $x \in I$.

We can also define another relation on the set of intervals as the set inclusion relation which says that

$$I \subseteq J \quad \text{iff} \quad \underline{J} \leq \underline{I} \;\&\; \overline{I} \leq \overline{J}.$$

For example, $[1, 2] \subseteq [0, 2]$. This is a partial ordering. This has to be noted that not every pair of intervals is comparable under this relation.

The notion of the degenerate interval permits us to regard the system of closed intervals as an extension of the real number system. Indeed, there is an obvious one-to-one pairing $[x, x] \mapsto x$ between the elements of the two systems. We will next investigate into the arithmetic operations of the intervals.

11.3 Arithmetic Operations on Intervals

We are about to define the basic arithmetic operations between intervals. The key point in these definitions is that computing with intervals is computing with sets. For example, when we add two intervals, the resulting interval is a set containing the sums of all pairs of numbers, one from each of the two initial sets. By definition then, the sum of two intervals I and J is

$$I + J = \{i + j : \ i \in I \ \& \ j \in J\}.$$

We will return to an operational description of addition momentarily (that is, to the task of obtaining a formula by which addition can be easily carried out). But let us define the remaining three arithmetic operations. The difference of two intervals I and J is the set

$$I - J = \{i - j : \ i \in I \ \& \ j \in J\}.$$

The product of I and J is given by

$$I.J = \{ij : \ i \in I \ \& \ j \in J\}.$$

Finally the quotient I/J is defined as

$$I/J = \{i/j : \ i \in I \ \& \ j \in J\}.$$

provided that $0 \notin J$.

We have seen the purpose of introducing the interval number system. So it is redundant to talk about arithmetic operations in terms of the terms in the interval. So, we will find a way to write it in terms of intervals.

1. Addition : Since $i \in I$ and $j \in J$ implies that

   $$\underline{I} \leq i \leq \overline{I} \quad \& \quad \underline{J} \leq j \leq \overline{J},$$

   we see by addition of inequalities that the sum $i + j \in I + J$ must satisfy

   $$\underline{I} + \underline{J} \leq i + j \leq \overline{I} + \overline{J}.$$

   Hence the formula

   $$I + J = [\underline{I} + \underline{J}, \overline{I} + \overline{J}].$$

   Example 11.3.1. Let $I = [0, 2]$ and $J = [-1, 2]$. Then

   $$I + J = [-1, 3].$$

   This is not the same as $I \cup J = [-1, 2]$

2. Subtraction : Since $i \in I$ and $j \in J$ implies that

   $$\underline{I} \leq i \leq \overline{I} \quad \& \quad -\overline{J} \leq -j \leq -\underline{J},$$

   gives

   $$\underline{I} - \overline{J} \leq i - j \leq \overline{I} - \underline{J}.$$

   It follows that

   $$I - J = [\underline{I} - \overline{J}, \overline{I} - \underline{J}].$$

Note that

$$I - J = I + (-J),$$

where, $-J$ is defined as

$$-J = [-\overline{J}, -\underline{J}] = \{y : \quad -y \in Y\}.$$

Note the reversal of endpoints that occurs when we find the negative of an interval.

Example 11.3.2. If $I = [-1, 0]$ and $J = [1, 2]$, then

$$-J = [-2, -1], \quad \& \quad I - J = [-3, -1].$$

What happens for $I - I$? Is it necessary that $I - I = 0$ as in the case of any real number? Consider $I = [2, 3]$. Then, as we have seen the definition of interval subtraction,

$$I - I = [2 - 3, 3 - 2] = [-1, 1].$$

In fact, for any interval $I = [\underline{I}, \overline{J}]$, we have

$$I - I = [\underline{I} - \overline{I}, \overline{I} - \underline{I}]$$

which is equal to 0 if and only if I is a degenerate interval.

3. Multiplication : The multiplication of intervals is given in terms of the minimum and maximum of four products of endpoints. Actually, by testing for the signs of the endpoints $\underline{I}$, $\overline{I}, \underline{J}, \overline{J}$. The formula for the endpoints of the interval product can be broken into nine special cases. In eight of these, only two products need be computed. We may write it as follows.

$$I \cdot J = \left[\min\{\underline{I}\underline{J}, \underline{I}\overline{J}, \overline{I}\underline{J}, \overline{I}\overline{J}\}, \max\{\underline{I}\underline{J}, \underline{I}\overline{J}, \overline{I}\underline{J}, \overline{I}\overline{J}\}\right]$$

Example 11.3.3. Let $I = [3, 4]$ and $J = [2, 2]$. Then

$$I \cdot J = [6, 8].$$

4. Division : The division of intervals are similarly found using minimum and maximum of the quotient of the endpoints of the intervals where the second interval J does not contain the term 0. So

$$I/J = \left[\min\{\underline{I}/\underline{J}, \underline{I}/\overline{J}, \overline{I}/\underline{J}, \overline{I}/\overline{J}\}, \max\{\underline{I}/\underline{J}, \underline{I}/\overline{J}, \overline{I}/\underline{J}, \overline{I}/\overline{J}\}\right]$$

Example 11.3.4. Let $I = [4, 10]$ and $J = [1, 2]$. Then

$$I/J = [2, 10].$$

Exercise 11.3.5. 1. Find $I \cap J$ and $I \cup J$ for the following intervals

(a) $I = [3, 4]$ and $J = [5, 7]$
(b) $I = [1, 2]$ and $J = [0, 3]$
(c) $I = [1, 4]$ and $J = [2, 6]$

2. Find $I + J$ and $I \cup J$ if $I = [5, 7]$ and $J = [-2, 6]$.

3. Find $I - J$ if $I = [5, 6]$ and $J = [-2, 4]$.

11.4 Algebraic Properties of Interval Numbers

We will now study certain algebraic properties related to the interval numbers as follows.

1. **Commutative and Associative Properties:** It is easy to show that the interval addition and multiplication are commutative and associative. That is, for any three intervals I, J, K,

$$\begin{array}{ll} I+J=J+I, & I+(J+K)=(I+J)+K, \\ IJ=JI, & I(JK)=(IJ)K. \end{array}$$

2. **Additive and Multiplicative elements:** The degenerate intervals 0 and 1 are additive and multiplicative identity elements in the system of intervals

$$0+I=0+I=I, \quad 1.I=I.1=I, \quad 0.I=I.0=0$$

for any interval I.

3. **Nonexistence of Inverse Elements:** We note that $-I$ is not an additive inverse for I. We have

$$I+(-I)=[\underline{I},\overline{I}]+[-\overline{I},-\underline{I}]=[\underline{I}-\overline{I},\overline{I}-\underline{I}],$$

and this is zero only if $\underline{I}=\overline{I}$. If I does not have zero width, then

$$I-I=w(I)[-1,1].$$

Similarly, $I/I=1$ only if $w(I)=0$. In general,

$$\begin{array}{rcl} I/I & = & [\underline{I}/\overline{I},\overline{I}/\underline{I}];\ \ 0<\underline{I}, \\ & = & [\overline{I}/\underline{I},\underline{I}/\overline{I}];\ \ \overline{I}<0. \end{array}$$

We don't have additional additive or multiplicative inverses except for degenerate intervals. However, we always have the inclusions $0\in I-I$ and $1\in I/I$.

4. **Subdistributivity:** The distributive law

$$x(y+z)=xy+xz$$

of ordinary arithmetic also fails to hold for intervals. An easy counterexample can be obtained by taking $I=[1,2],\ \ J=[1,2],\ \ K=[-1,1]$ which gives

$$I(J+K)=[1,2].([1,1]-[1,1])=[1,2].[0,0].$$

Also,

$$IJ+IK=[1,2].[1,1]-[1,2].[1,1]=[-1,1].$$

However, the subdistributive law says that

$$I(J+K)\subseteq IJ+IK.$$

We can see this in the example above. Full distributivity does hold in certain special cases. In particular, for any real number x we have

$$x(J+K)=xJ+xK.$$

Interval multiplication can be distributed over a sum of intervals as long as those intervals have the same sign:

$$I(J+K)\subseteq IJ+IK, \text{ provided that } JK>0.$$

5. **Cancellation Law:** The cancellation law

$$I + K = J + K \implies I = J$$

holds for interval addition.

We should emphasize that, with the identification of degenerate intervals and real numbers, interval arithmetic is an extension of real arithmetic. It reduces to ordinary real arithmetic for intervals of zero width.

Exercise 11.4.1. 1. Verify the distributive law for the intervals $I = [1, 2],\ J = [-3, -2],\ K = [-5, -1]$.

2. Prove the Cancellation law. Show that multiplicative cancellation does not hold in interval arithmetic, that is, $IK = JK$ does not imply $I = J$.

Symmetric Intervals

An interval I is said to be symmetric if $\underline{I} = -\overline{I}$. For example, $[-1, 1]$ is symmetric and $[-1, 5]$ is not. Any symmetric interval has midpoint 0. If I is symmetric, then

$$|I| = \frac{1}{2}w(I), \qquad I = |I|[-1, 1].$$

The rules of interval arithmetic are slightly simpler when symmetric intervals are involved. If I, J, K are all symmetric, then

$$\begin{aligned} I + J = I - J &= (|I| + |J|)[-1, 1], \\ IJ &= |I||J|[-1, 1], \\ I(J \pm K) = IJ + JK &= |I|(|J| + |K|)[-1, 1]. \end{aligned}$$

If J is symmetric and I is any interval, then

$$IJ = |I|J.$$

It follows that if J and K are symmetric, then

$$I(J + K) = IJ + IK$$

for any interval I.

Inclusion Isotonicity of Interval Arithmetic

Let $\odot$ stand for interval addition, subtraction, multiplication, or division. If A, B, C and D are intervals such that

$$A \subseteq C \quad \text{and} \quad B \subseteq D,$$

then

$$A \odot B \subseteq C \odot D.$$

These relations follow directly from the definitions given previously. Interval arithmetic is said to be inclusion isotonic. We will now extend the concept of interval expressions to include functions such as $\sin x$ and e^x.

11.5 Interval Functions

Let f be a real-valued function of a single real variable x. Ultimately, we would like to know the precise range of values taken by $f(x)$ as x varies through a given interval I. In other words, we would like to be able to find the image of the set I under the mapping f, which is, $f(I) = \{f(x) :\ x \in I\}$. More generally, given a function $f = f(x_1, \ldots, x_n)$ of several variables, we will wish to find the image set

$$f(I_1, \ldots, I_n) = \{f(x_1, \ldots, x_n) :\ x_1 \in I_1, \ldots, x_n \in I_n\}$$

where $I_1, \ldots, I_n$ are specified intervals.

Definition 11.5.1. Let $g : M_1 \to M_2$ be a mapping between sets M_1 and M_2, and denote by $S(M_1)$ and $S(M_2)$ the families of subsets of M_1 and M_2, respectively. The united extension of g is the set-valued mapping $\overline{g} : S(M_1) \to S(M_2)$ such that

$$\overline{g}(I) = \{g(x) :\ x \in I,\ I \in S(M_1)\}.$$

The mapping $\overline{g}$ is sometimes of interest as a single-valued mapping on $S(M_1)$ with values in $S(M_2)$. For our purposes, however, it is merely necessary to note that

$$\overline{g}(I) = \cup_{x \in I}\{g(x)\},$$

that is, $\overline{g}(I)$ contains precisely the same elements as the set image $g(I)$. For this reason, and because the usage is common, we shall apply the term united extension to set images such as those described previously.

Elementary Functions of Interval Arguments

For some functions, the image set is easy to compute. For example, consider $f(x) = x^2$, $x \in \mathbb{R}$. If $I = [\underline{I}, \overline{I}]$, it is evident that the set

$$f(I) = \{x^2 :\ x \in I\}$$

can be expressed as

$$\begin{aligned} f(I) &= [\underline{I}^2, \overline{I}^2], \quad 0 \le \underline{I} \le \overline{I}, \\ &= [\overline{I}^2, \underline{I}^2], \quad \underline{I} \le \overline{I} \le 0, \\ &= [0, \max\{\underline{I}^2, \overline{I}^2\}], \quad \underline{I} < 0 < \overline{I}. \end{aligned}$$

Note that I^2 is not the same as $I.I$. For example

$$[-1, 1]^2 = [0, 1], \qquad [-1, 1].[-1, 1] = [-1, 1].$$

We will use the definition of I^2 for $f(I)$. However, $[-1, 1]$ does contain $[0, 1]$. The overestimation when we compute a bound on the range of I^2 as $I.I$ is due to the phenomenon of interval dependency. Namely, if we assume x is an unknown number known to lie in the interval I, then, when we form the product $x.x$, the x in the second factor, although known only to lie in I must be the same as the x in the first factor, whereas, in the definition of the interval product $I.I$, it is assumed that the values in the first factor and the values in the second factor vary independently.

Interval dependency is a crucial consideration when using interval computations. It is a major reason why simply replacing floating point computations by intervals in an existing algorithm is not likely to lead to satisfactory results.

The reasoning is particularly straightforward with functions f (x) that happen to be monotonic, i.e., either increasing or decreasing with increasing x. Note that, an increasing function f maps an interval $I = [\underline{I}, \overline{I}]$ into the interval $f(I) = [f(\underline{I}), f(\overline{I})]$.

11.6 Interval-Valued Extensions of Real Functions

Let us begin with an example. Consider the real-valued function f given by $f(x) = 1 - x$, $x \in \mathbb{R}$. Note carefully that a function is defined by two things: (1) a domain over which it acts, and (2) a rule that specifies how elements of that domain are mapped under the function. Both of these are specified in the definition of f. The elements of Dom$f$ are real numbers x, and the mapping rule is $x \mapsto 1 - x$. Taken in isolation, the entity $f(x) = 1 - x$ is a formula—not a function. Often this distinction is ignored; in many elementary math books, for example, we would interpret the entity as a function whose domain should be taken as the largest possible set over which the formula makes sense (in this case, all of $\mathbb{R}$). However, we will understand that Dom$f$ is just as essential to the definition of f as is the formula $f(x)$.

Now suppose we take the formula that describes the given function f and apply it to interval arguments. The resulting interval-valued function

$$F(I) = 1 - I, \quad I = [\underline{I}, \overline{I}],$$

is an extension of the function f. we have enlarged the domain to include nondegenerate intervals I as well as the degenerate intervals $x = [x, x]$.

Definition 11.6.1. We say that F is an interval extension of f , if for degenerate interval arguments, F agrees with f, that is, $F([x, x]) = f(x)$.

Let us compare $F(I)$ with the set image $f(I)$. We have according to the laws of interval arithmetic,

$$F(I) = [1, 1] - [\underline{I}, \overline{I}] = [1, 1] + [-\overline{I}, -\underline{I}] = [1 - \overline{I}, 1 - \underline{I}].$$

On the other hand, as x increases through the interval $[\underline{I}, \overline{I}]$, the the value of $f(x)$ given by $1 - x$ decreases from $1 - \overline{I}$ to $1 - \underline{I}$. So by definition, $f(I) = [1 - \overline{I}, 1 - \underline{I}]$. In this example, we have $F(I) = f(I)$; this particular extension of f obtained by the formula $f(x) = 1 - x$ directly to interval arguments, yields the desired set image $f(I)$. In other words, we have found the united extension of f, which is, $f(I) = 1 - I$. Although the situation is not always so simple, but we will leave it for the time being and move on to the definition of distance between intervals.

11.7 Distance between Intervals

We are very much accustomed with the idea of metric and the basic point set theory, the convergence, completeness, etc. We will now attempt to define metric for the interval numbers.

Definition 11.7.1. If I and J are two intervals, then the distance between them is defined by

$$d(I, J) = \max\{|\underline{I} - \underline{J}|, |\overline{I} - \overline{J}|\}.$$

We can define the concepts of convergence, continuity with the help of the above definition.

Definition 11.7.2. Let $\{I_k\}$ be a sequence of intervals. We say that it converges if there exists an interval I^* such that for every $\epsilon > 0$, there is a natural number $N = N(\epsilon)$ such that $d(I_k, I^*) < \epsilon$ whenever $k > N$. As in the case of real sequences, we write

$$I^* = \lim_{k \to \infty} I_k.$$

We know that the interval number system represents an extension of the real number system. In fact, the correspondence $[x, x] \leftrightarrow x$ can be regarded as a function or mapping between the two systems. This mapping preserves distances between corresponding objects. We have

$$d([x, x], [y, y]) = \max\{|x - y|, |x - y|\} = |x - y|$$

for any real x and y. For this reason, it is called an isometry, and we say that the real line is "isometrically embedded" in the metric space of interval numbers.

Exercise 11.7.3. 1. Show that the definition of distance given between two intervals satisfy the metric axioms.

2. Find the distance between the intervals $I = [1, 2]$ and $J = [3, 5]$.

3. For any intervals I, J, K prove that

   (a) $d(I + K, Y + K) = d(I, J)$;

   (b) $d(I, J) \leq w(J)$ when $I \subseteq J$;

   (c) $d(I, 0) = |I|$.

Few Probable Questions

1. Define symmetric interval. Show that any interval I can be expressed as the sum of a real number (i.e., degenerate interval) and a symmetric interval:

$$I = m + W, \quad \text{where} \quad m = m(I) \quad \text{and} \quad W = \frac{1}{2}w(I)[-1, 1].$$

2. Show that $I_k \to I$ if and only if $\underline{I}_k \to \underline{I}$ and $\overline{I}_k \to \overline{I}$.

Unit 12

Course Structure

- Basic concepts of fuzzy sets: Types of fuzzy sets, α-cuts and its properties, representations of fuzzy sets.
- Support, convexity, normality, cardinality, standard set-theoretic operations on fuzzy sets

---

12.1 Introduction

In mathematics, fuzzy sets (also known as uncertain sets) are somewhat like sets whose elements have degrees of membership. Fuzzy sets were introduced independently by Lotfi A. Zadeh and Dieter Klaua in 1965 as an extension of the classical notion of set. At the same time, Salii (1965) defined a more general kind of structure called an L-relation, which he studied in an abstract algebraic context. Fuzzy relations, which are used now in different areas, such as linguistics (De Cock, Bodenhofer & Kerre 2000), decision-making (Kuzmin 1982), and clustering (Bezdek 1978), are special cases of L-relations when L is the unit interval $[0, 1]$.

In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition — an element either belongs or does not belong to the set. By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of a membership function valued in the real unit interval $[0, 1]$. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of the membership functions of fuzzy sets, if the latter only take values 0 or 1. In fuzzy set theory, classical bivalent sets are usually called crisp sets. The fuzzy set theory can be used in a wide range of domains in which information is incomplete or imprecise, such as bioinformatics.

Objectives

After reading this unit, you will be able to

- define fuzzy sets and its types
- define α-cuts of fuzzy sets and related properties
- learn various representations of fuzzy sets
- define the set theoretic operations on fuzzy sets and see various related examples

12.1.1 Fuzzy Sets

A classical (crisp) set is normally defined as a collection of elements or objects $x \in X$ that can be finite, countable, or uncountable. Each single element can either belong to or not belong to a set A, $A \subseteq X$. In the former case, the statement "x belongs to A" is true, whereas in the latter case this statement is false.

Such a classical set can be described in different ways: one can either enumerate (list) the elements that belong to the set; describe the set analytically, for instance, by stating conditions for membership ($A = \{x : x \leq 5\}$); or define the member elements by using the characteristic function, in which 1 indicates membership and 0 nonmembership. For a fuzzy set, the characteristic function allows various degrees of membership for the elements of a given set.

Definition 12.1.1. If X is a collection of objects denoted generically by x, then a fuzzy set $\tilde{A}$ in X is a set of ordered pairs

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) : x \in X\}.$$

$\mu_{\tilde{A}}(x)$ is called the membership function or grade of membership (also degree of compatibility or degree of truth) of $x \in \tilde{A}$ that maps X to the membership space M (When M contains only the two points 0 and 1, A is nonfuzzy and $\mu_{\tilde{A}}(x)$ is identical to the characteristic function of a nonfuzzy set). The range of the membership function is a subset of the nonnegative real numbers whose supremum is finite. Elements with a zero degree of membership are normally not listed. The set X is called the universal set and let us denote the set of all fuzzy sets on X by $\mathscr{F}(X)$.

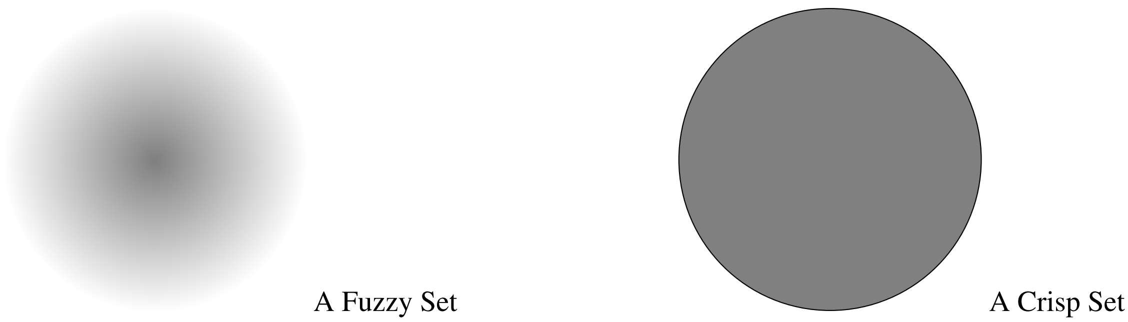


Figure 12.1.1: A visual comparison between Fuzzy and Crisp Sets

Fuzzy sets are represented in different ways.

1. A fuzzy set is denoted by an ordered set of pairs, the first element of which denotes the element and the second the degree of membership.

   Example 12.1.2. A realtor wants to classify the house he offers to his clients. One indicator of comfort of these houses is the number of bedrooms in it. Let $X = \{1, 2, \ldots, 10\}$ be the set of available types of houses described by x =number of bedrooms in a house. Then the fuzzy set "comfortable type of house for a four-person family" may be described as

   $$\tilde{A} = \{(1, 0.2), (2, 0.5), (3, 0.8), (4, 1), (5, 0.7), (6, 0.3)\}.$$

   Example 12.1.3. Let $\tilde{A}$ = real numbers "considerably" larger than 10. Then in this case, the numbers less than or equal to 10 automatically falls out and we must define $\mu_{\tilde{A}}(x)$ in such a way that as x goes farther away from 10, the membership function increases. We define $\mu_{\tilde{A}}(x)$ as

   $$\begin{aligned} \mu_{\tilde{A}}(x) &= 0, \quad x \leq 10 \\ &= \frac{1}{1 + \frac{1}{(x-10)^2}}, \quad x > 10 \end{aligned}$$

Example 12.1.4. Let $\tilde{A}$ = real numbers close to 10. Then

$$\tilde{A} = \left\{ (x, \mu_{\tilde{A}}(x)) : \ \mu_{\tilde{A}}(x) = \frac{1}{1 + (x - 100)} \right\}.$$

If we plot the graph of the membership function against the members set elements, then we will get somewhat as given in the figure.

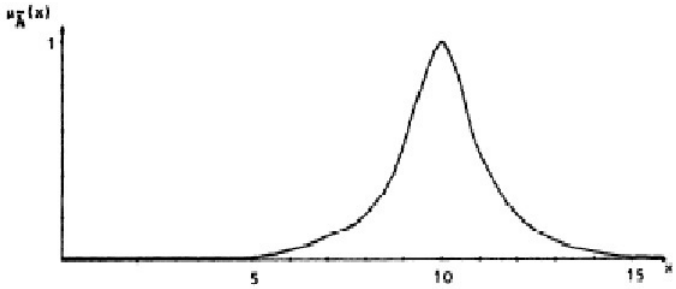


Figure 12.1.2: Real numbers close to 10

2. A fuzzy set is represented can be sometimes solely by stating its membership function.
3.

$$\begin{aligned} \tilde{A} = \mu_{\tilde{A}}(x_1)/x_1 + \mu_{\tilde{A}}(x_2)/x_2 + \cdots &= \sum_{i=1}^{n} \mu_{\tilde{A}}(x_i)/x_i \\ &\text{or } \int_x \mu_{\tilde{A}}(x)/x. \end{aligned}$$

Example 12.1.5. If $\tilde{A}$ =integers close to 10, then

$$\tilde{A} = 0.1/7 + 0.5/8 + 0.8/9 + 1/10 + 0.8/11 + 0.5/12 + 0.1/13.$$

Also, if $\tilde{A}$ =real numbers close to 10, then

$$\tilde{A} = \int_{\mathbb{R}} \frac{1}{1 + (x - 10)^2} / x.$$

It has already been mentioned that the membership function is not limited to values between 0 and 1. However, the most commonly used range of values of membership functions is the unit interval $[0, 1]$.

Definition 12.1.6. A fuzzy set $\tilde{A}$ is called normal if $\sup_x \mu_{\tilde{A}}(x) = 1$.

A non-empty fuzzy set $\tilde{A}$ can always be normalized by dividing $\mu_{\tilde{A}}(x)$ by $\sup_x \mu_{\tilde{A}}(x)$. For the representation of fuzzy sets, we will use the notation 1.

A fuzzy set is obviously a generalization of a classical set and the membership function a generalization of the characteristic function. Since we are generally referring to a universal (crisp) set X, some elements of a fuzzy set may have the degree of membership zero. Often it is appropriate to consider those elements of the universe that have a nonzero degree of membership in a fuzzy set.

Definition 12.1.7. The support of a fuzzy set $\tilde{A}$, $S(\tilde{A})$, is the crisp set of all $x \in X$ such that $\mu_{\tilde{A}}(x) > 0$.

Example 12.1.8. For example (12.1.2), the support of $\tilde{A}$ is $S(\tilde{A}) = \{1, 2, 3, 4, 5, 6\}$. The elements $\{7, 8, 9, 10\}$ are not part of the support of $\tilde{A}$.

12.1.2 Types of Fuzzy Sets

So far we have considered fuzzy sets with crisply defined membership functions or degrees of membership. Such kind of fuzzy sets are also called ordinary fuzzy sets. However, there may be some situations in which the appropriate membership functions are needed to be specified only approximately. So, for every point on the universal set X, one may assign a sub-interval in $[0, 1]$. Thus, we arrive at a more general kind of fuzzy set, called the interval-valued fuzzy set. These sets are defined formally as follows.

Definition 12.1.9. An interval-valued fuzzy set defined over a universal set X is a set whose membership function is of the form

$$\mu : X \to \epsilon([0, 1])$$

where $\epsilon([0, 1])$ denotes the family of all closed intervals of real numbers in $[0, 1]$.

These fuzzy sets can further be generalised by allowing the intervals of the membership functions to be fuzzy. Each interval now becomes an ordinary fuzzy set. This brings us to the type 2 fuzzy sets which are formally defined below.

Definition 12.1.10. A type 2 fuzzy set is a fuzzy set whose membership values are ordinary fuzzy sets on $[0, 1]$, that is, if X is the universal set, then a type 2 fuzzy set on X is a fuzzy set whose membership function is of the form

$$\mu : X \to \mathscr{F}([0, 1]).$$

Motivated from this definition, we can also say that ordinary fuzzy sets are type 1 fuzzy sets. However, if we further generalise the above definition and say that each point on the universal set is assigned a type 2 fuzzy set for its value of the membership function, then such type of set is called a type 3 fuzzy set. Other higher types of fuzzy set can be recursively defined.

Definition 12.1.11. A type m fuzzy set is a fuzzy set in X whose membership values are type $m - 1$ ($m > 1$) fuzzy sets on $[0, 1]$.

From a practical point of view, such type m fuzzy sets for large m (even for $m \geq 3$) are hard to deal with, and it will be extremely difficult or even impossible to measure them or to visualize them. We will, therefore, not even try to define the usual operations on them.

There is another direction of generalising the fuzzy sets. There may arise certain situations, in which the elements of the universal set can not be certainly specified. Hence, in that case, it would be more efficient to assign membership values to fuzzy sets of X instead of the elements of X. Such kind of fuzzy sets are called level 2 fuzzy sets.

Definition 12.1.12. A fuzzy set of level 2 is a set universal set is a fuzzy set; that is, whose membership function has the form

$$\mu : \mathscr{F}(X) \to [0, 1].$$

Level 2 fuzzy sets can also be generalised into level 3 fuzzy set by using a universal set whose elements are fuzzy level 2 sets. Higher level fuzzy sets can be recursively defined. We can also define fuzzy sets that are of type 2 and level 2. The membership function is of the form

$$\mu : \mathscr{F}(X) \to \mathscr{F}([0, 1]).$$

Another definition was given by Hirota which is given below.

Definition 12.1.13. A probabilistic set A on X is defined by a defining function μ_A,

$$\mu_A : X \times \Omega \text{ defined as } (x, \omega) \mapsto \mu_A(x, \omega) \in \Omega_C$$

where $\mu_A(x, \dot{)}$ is the (B, B_C)-measurable function for each fixed $x \in X$.

For Hirota, a probabilistic set A with the defining function $\mu_A(x,\omega)$ is contained in a probabilistic set B with $\mu_B(x,\omega)$ if for each $x \in X$ there exists an $E \in B$ such that $P(E) = 1$ and $\mu_A(x,\omega) \leq \mu_B(x,\omega)$ for all $\omega \in E$. (Ω, B, P) is called the parameter space.

Further attempts at representing vague and uncertain data with different types of fuzzy sets were made by Atanassov and Stoeva and by Pawlak which are given below.

Definition 12.1.14. Given an underlying set X of objects, an intuitionistic fuzzy set (IFS) A is a set of ordered triples,

$$A = \{(x, \mu_A(x), \nu_A(x)) : \ x \in X\}$$

where $\mu_A(x)$ and $\nu_A(x)$ are functions mapping from X into $[0,1]$. For each $x \in X$, $\mu_A(x)$ represents the degree of membership of the element x to the subset A of X, and $\nu_A(x)$ gives the degree of nonmembership. For the functions $\mu_A(x)$ and $\nu_A(x)$ mapping into $[0,1]$, the condition $0 \leq \mu_A(x) + \nu_A(x) \leq 1$ holds.

Ordinary fuzzy sets over X may be viewed as special intuitionistic fuzzy sets with the nonmembership function $\nu_A(x) = 1 - \mu_A(x)$.

Definition 12.1.15. Let U denote a set of objects called universe and let $R \subset U \times U$ be an equivalence relation on U. The pair $A = (U, R)$ is called an approximation space. For $u, v \in U$ and $(u,v) \in R$, u and v belong to the same equivalence class, and we say that they are indistinguishable in A. Hence the relation R is called an indiscernibility relation. Let $[x]_R$ denote an equivalence class (elementary set of A) R containing element x; then the lower and upper approximations for a subset $X \subseteq U$ in A-denoted by $\underline{A}(X)$ and $\overline{A}(X)$ respectively, are defined as follows.

$$\underline{A}(X) = \{x \in U : \ [x]_R \subset X\} \quad \text{and} \quad \overline{A}(X) = \{x \in U : \ [x]_R \cap X \neq \theta\}.$$

If an object x belongs to the lower approximation space of X in A, then “x surely belongs to X in A," $x \in \overline{A}(X)$ means that “x possibly belongs to X in A."

For the subset $X \subseteq U$ representing a concept of interest, the approximation space $A = (U, R)$ can be characterized by three distinct regions of X in A: the so-called positive region $\underline{A}(X)$, the boundary region $\overline{A}(X) - \underline{A}(X)$, and the negative region $U - \overline{A}(X)$.

The characterization of objects in X by the indiscernibility relation R is not precise enough if the boundary region $\overline{A}(X) - \underline{A}(X)$ is not empty. For this case it may be impossible to say whether an object belongs to X or not, and so the set X is said to be nondefinable in A, and X is a rough set.

12.1.3 Basic Set-Theoretic Operations for Fuzzy Sets

The membership function is obviously the crucial component of a fuzzy set. It is therefore not surprising that operations with fuzzy sets are defined via their membership functions. We shall first present the concepts suggested by Zadeh in 1965. They constitute a consistent framework for the theory of fuzzy sets. They are, however, not the only possible way to extend classical set theory consistently.

Definition 12.1.16. The membership function $\mu_{\tilde{C}}(x)$ of the intersection $\tilde{C} = \tilde{A} \cap \tilde{B}$ is pointwise defined by

$$\mu_{\tilde{C}}(x) = \min\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\} \ \ x \in X.$$

Definition 12.1.17. The membership function $\mu_{\tilde{D}}(x)$ of the union $\tilde{D} = \tilde{A} \cup \tilde{B}$ is pointwise defined by

$$\mu_{\tilde{D}}(x) = \max\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\} \ \ x \in X.$$

Definition 12.1.18. The membership function of the complement of a normalized fuzzy set $\tilde{A}$ (denoted by $C\tilde{A}$), $\mu_{C\tilde{A}}(x)$ is defined by

$$\mu_{C\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x), \ \ x \in X.$$

Example 12.1.19. Let $\tilde{A}$ be the fuzzy set in the example (12.1.2) and $\tilde{B}$ be the fuzzy set "large type of house" defined as

$$\tilde{B} = \{(3, 0.2), (4, 0.4), (5, 0.6), (6, 0.8), (7, 1), (8, 1)\}$$

The intersection $\tilde{C} = \tilde{A} \cap \tilde{B}$ is then

$$\tilde{C} = \{(3, 0.2), (4, 0.4), (5, 0.6), (6, 0.3)\}$$

and the union $\tilde{D} = \tilde{A} \cup \tilde{B}$

$$\tilde{D} = \{(l, 0.2), (2, 0.5), (3, 0.8), (4, 1), (5, 0.7), (6, 0.8), (7, 1), (8, 1)\}$$

The complement $C\tilde{B}$, which might be interpreted as "not large type of house," is

$$C\tilde{B} = \{(1, 1), (2, 1), (3, 0.8), (4, 0.6), (5, 0.4), (6, 0.2), (9, 1), (l0, 1)\}.$$

It has already been mentioned that min and max are not the only operators that could have been chosen to model the intersection or union, respectively, of fuzzy sets. The question arises, why those and not others? Bellman and Giertz addressed this question axiomatically in 1973. They argued from a logical point of view, interpreting the intersection as "logical and,"the union as "logical or," and the fuzzy set $\tilde{A}$ as the statement "The element x belongs to the set $\tilde{A}$" which can be accepted as more or less true. It is very instructive to follow their line of argument, which is an excellent example for an axiomatic justification of specific mathematical models. We shall therefore sketch their reasoning: Consider two statements, S and T, for which the truth values are μ_S and μ_T respectively, where $\mu_S, \mu_T \in [0, 1]$. The truth value of the "and" and "or" combination of these statements, $\mu(S \text{ and } T)$ and $\mu(S \text{ or } T)$, both from the interval $[0, 1]$, are interpreted as the values of the membership functions of the intersection and union , respectively, of S and T. We are now looking for two real-valued functions f and g such that

$$\begin{aligned} \mu_{S \text{ and } T} &= f(\mu_S, \mu_T) \\ \mu_{S \text{ or } T} &= g(\mu_S, \mu_T). \end{aligned}$$

Bellman and Giertz feel that the following restrictions are reasonably imposed on f and g:

1. f and g are nondecreasing and continuous in μ_S and μ_T.
2. f and g are symmetric, that is,

$$\begin{aligned} f(\mu_S, \mu_T) &= f(\mu_T, \mu_S) \\ g(\mu_S, \mu_T) &= g(\mu_T, \mu_S). \end{aligned}$$

3. $f(\mu_S, \mu_S)$ and $g(\mu_S, \mu_S)$ are strictly increasing in μ_S.
4. $f(\mu_S, \mu_T) \leq \min(\mu_S, \mu_T)$ and $g(\mu_S, \mu_T) \geq \max(\mu_S, \mu_T)$. This implies that accepting the truth of the statement "S and T" requires more, and accepting the truth of the statement "S or T" less than accepting S or T alone as true.
5. $f(1, 1) = 1$ and $g(0, 0) = 0$.
6. Logically equivalent statements must have equal truth values, and fuzzy sets with the same contents must have the same membership functions, that is,

$$S_1 \text{ and } (S_2 \text{ or } S_3)$$

   is equivalent to

$$(S_1 \text{ and } S_2) \text{ or } (S_1 \text{ and } S_3)$$

   and therefore must be equally true.

Bellman and Giertz now formalize the above assumptions as follows : Using the symbols $\wedge$ for "and" and $\vee$ for "or", these assumptions amount to the following seven restrictions, to be imposed on the two commutative and associative binary compositions $\wedge$ and $\vee$ on the closed interval $[0, 1]$, which distributive with respect to one another.

1. $\mu_S \wedge \mu_T = \mu_T \wedge \mu_S$ and $\mu_S \vee \mu_T = \mu_T \vee \mu_S$.
2. $(\mu_S \wedge \mu_T) \wedge \mu_U = \mu_S \wedge (\mu_T \wedge \mu_U)$ and $(\mu_S \vee \mu_T) \vee \mu_U = \mu_S \vee (\mu_T \vee \mu_U)$.
3. $\mu_S \wedge (\mu_T \vee \mu_U) = (\mu_S \wedge \mu_T) \vee (\mu_S \wedge \mu_U)$ and $\mu_S \vee (\mu_T \wedge \mu_U) = (\mu_S \vee \mu_T) \wedge (\mu_S \vee \mu_U)$.
4. $\mu_S \wedge \mu_T$ and $\mu_S \vee \mu_T$ are continuous and nondecreasing in each component.
5. $\mu_S \wedge \mu_T$ and $\mu_S \vee \mu_T$ are are strictly increasing in μ_S.
6. $\mu_S \wedge \mu_T \leq \min(\mu_S, \mu_T)$ and $\mu_S \vee \mu_T \leq \max(\mu_S, \mu_T)$.
7. $1 \wedge 1 = 1$ and $0 \vee 0 = 0$.

Bellman and Giertz then prove mathematically that $\mu_{S\wedge T} = \min(\mu_S, \mu_T)$ and $\mu_{S\vee T} = \max(\mu_S, \mu_T)$.

For the complement, it would be reasonable to assume that if statement "S" is true, its complement "non S" is false, or if $\mu_S = 1$, then $\mu_{\text{non } S} = 0$ and vice versa.

12.1.4 α-cuts and strong α-cuts

A more general and even more useful notion is that of an α-level set.

Definition 12.1.20. The (crisp) set of elements that belong to the fuzzy set $\tilde{A}$ at least to the degree α is called the α-level set or α-cut

$$A_\alpha = \{x \in X : \ \mu_{\tilde{A}}(x) \geq \alpha\}$$

$A'_\alpha = \{x \in X : \ \mu_{\tilde{A}}(x) > \alpha\}$ is called strong α-level set or strong α-cut.

Any property generalised from the classical set theory, if preserved for all α-cuts from $\alpha \in (0, 1]$ in the classical sense, is called a cutworthy property and if it is preserved for all strong α-cuts, then it is called a strong cutworthy property.

Example 12.1.21. Again we refer to the example (12.1.2). We list a possible α-level sets.

$$\begin{aligned} A_{0.2} &= \{1, 2, 3, 4, 5, 6\} \\ A_{0.5} &= \{2, 3, 4, 5\} \\ A_{0.8} &= \{3, 4\} \\ A_1 &= \{4\}. \end{aligned}$$

The strong 0.8-level set is $A'_{0.8} = \{4\}$.

Let us discuss some properties of α-cuts and strong α-cuts.

Theorem 12.1.22. Let $\tilde{A}$ and $\tilde{B}$ be be two fuzzy sets on a universal set X. Then for all $a, b \in [0, 1]$,

1. ${A_a}' \subseteq A_a$;
2. $a \leq b$ implies that $A_b \subseteq A_a$ and $A'_b \subseteq A'_a$;
3. $(A \cap B)_a = A_a \cap B_a$ and $(A \cup B)_a = A_a \cup B_a$;

4. $(A \cap B)'_a = A'_a \cap B'_a$ and $(A \cup B)'_a = A'_a \cup B'_a$.

5. $(CA)_a = X \setminus A'_{1-a}$.

Proof. 1. By definition, $A'_a = \{x \in X : \mu_{\tilde{A}}(x) > a\} \subseteq \{x \in X : \mu_{\tilde{A}}(x) \geq a\} = A_a$.

2. Let $a \leq b$. Then, $A_b = \{x \in X : \mu_{\tilde{A}}(x) \geq b\} \subseteq \{x \in X : \mu_{\tilde{A}}(x) \geq a\} = A_a$. We can similarly show the result for the strong cuts.

3. For $x \in (A \cap B)_a$, we have, $\mu_{\tilde{A}\cap\tilde{B}}(x) \geq a$ and hence $\min\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\} \geq a$. This means that $\mu_{\tilde{A}}(x) \geq a$ and $\mu_{\tilde{B}}(x) \geq a$ and hence $x \in (A_a \cap B_a)$ and hence $(A \cap B)_a \subseteq A_a \cap B_a$. Conversely, for any $x \in A_a \cap B_a$, we have $x \in A_a$ and $x \in B_a$, that is, $\mu_{\tilde{A}}(x) \geq a$ and $\mu_{\tilde{B}}(x) \geq a$. Hence, $\min\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\} \geq a$ which means that $\mu_{A\tilde{\cap}B}(x) \geq a$. Hence, $x \in (A \cap B)_a$ and consequently, we have $(A \cap B)_a \supseteq A_a \cap B_a$. Thus, we have $(A \cap B)_a = A_a \cap B_a$.

   For the second equality, let $x \in (A \cup B)_a$, we have, $\max\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\} \geq a$ and hence, $\mu_{\tilde{A}}(x) \geq a$ and $\mu_{\tilde{B}}(x) \geq a$. This implies that $x \in A_a \cup B_a$ and thus $(A \cup B)_a \subseteq (A_a \cup B_a)$. Conversely, for any $x \in A_a \cup B_a$, we have, $x \in A_a$ and $x \in B_a$; that is, $\mu_{\tilde{A}}(x) \geq a$ or $\mu_{\tilde{B}}(x) \geq a$. Hence $\max\{\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)\} \geq a$, which means that $\mu_{A\tilde{\cup}B}(x) \geq a$. This means that $x \in (A \cup B)_a$ and hence, $A_a \cup B_a \subseteq (A \cup B)_a$. Hence the result.

4. Left as an exercise.

5. Note that for any $x \in (CA)_a$, $1 - \mu_A(x) = \mu_{CA}(x) \geq a$. This implies that $\mu_A(x) \leq 1 - a \Rightarrow x \notin A'_{1-a}$ and hence $x \in X \setminus A'_{1-a}$. Clearly, $(CA)_a \subseteq X \setminus A'_{1-a}$. For the opposite inequality, let us take an element x from $X \setminus A'_{1-a}$. This means that $x \notin \setminus A'_{1-a}$ due to which $\mu_A(x) \leq 1 - a$ or, $1 - \mu_A(x) \geq a$. Thus, $\mu_{CA}(x) \geq a$ which means that $x \in (CA)_a$, finally yielding $X \setminus A'_{1-a} \subseteq (CA)_a$. Hence the result.

□

Let us examine the significance of the properties stated in the previous theorem. Property 1 is trivial, expressing that the strong α-cut is always included in the α-cut of any fuzzy set and for any $a \in [0, 1]$; the property follows directly from the definitions of the two types of α-cuts. Property 2 means that the set sequences $\{A_a : a \in [0, 1]\}$ and $\{A'_a : a \in [0, 1]\}$ of a-cuts and strong a-cuts, respectively are always monotonic decreasing with respect to a; consequently, they are nested families of sets. Properties 3 and 4 show that the standard fuzzy intersection and fuzzy union are both cutworthy and strong cutworthy when applied to two fuzzy sets or, due to the associativity of min and max, to any finite number of fuzzy sets. However, property 5 shows that standard fuzzy complement is neither cutworthy or strong cutworthy. The following result shows the behaviour of the α-cuts and strong α-cuts for any number of fuzzy sets.

Theorem 12.1.23. Let A^i be fuzzy sets over the universal set X for all $i \in I$, where I is an index set. Then,

1. $\displaystyle\bigcup_{i\in I} A^i_a \subseteq \left(\bigcup_{i\in I} A^i\right)_a$ and $\displaystyle\bigcap_{i\in I} A^i_a = \left(\bigcap_{i\in I} A^i\right)_a$;

2. $\displaystyle\bigcup_{i\in I} A^{i'}_a = \left(\bigcup_{i\in I} A^i\right)'_a$ and $\displaystyle\bigcap_{i\in I} A^{i'}_a \subseteq \left(\bigcap_{i\in I} A^i\right)'_a$.

Proof. 1. Let $x \in \bigcup_{i\in I} A^i_a$. Then $x \in A^{i_0}_a$ for some $i_0 \in I$. Then

$$\begin{aligned} & \mu_{A^{i_0}}(x) \geq a \\ \Rightarrow\ & \sup_{i\in I} \mu_{A^i}(x) \geq a \\ \Rightarrow\ & \mu_{\bigcup_{i\in I} A^i}(x) \geq a \\ \Rightarrow\ & x \in \left(\bigcup_{i\in I} A^i\right)_a. \end{aligned}$$

For the second part of the statement, we see that

$$\begin{aligned} & x \in \bigcap_{i\in I} A^i_a \\ \Leftrightarrow\ & x \in A^i_a \ \ \forall i \in I \\ \Leftrightarrow\ & \mu_{A^i}(x) \geq a \ \ \forall i \in I. \end{aligned}$$

This implies that $\inf_{i\in I} \mu_{A^i}(x) \geq a$, or, $\mu_{\bigcap_{i\in I} A^i}(x) \geq a$. Thus, the result follows.

2. For all $x \in X$,

$$x \in \bigcup_{i\in I} A^{i'}_a$$

if and only if there exists some $i_0 \in I$ such that $x \in A^{i'_0}_a$ (that is, $\mu_{A^{i_0}}(x) > a$). This inequality is satisfied iff

$$\sup_{i\in I} \mu_{A^i}(x) > a,$$

which is equivalent to

$$\mu_{\bigcup\limits_{i\in I} A^i}(x) > a.$$

That is,

$$x \in \left(\bigcup_{i\in I} A^i\right)'_a.$$

Hence the equality in 2 is satisfied.

We now prove the second proposition in 2. For all

$$x \in \left(\bigcap_{i\in I} A^i\right)'_a,$$

we have

$$\mu_{\bigcap\limits_{i\in I} A^i}(x) > a;$$

that is,

$$\inf_{i\in I} \mu_{A^i}(x) > a.$$

Hence, for any $i \in I$, $\mu_{A^i}(x) > a$ which means that $x \in A^{i'}_a$. Hence

$$x \in \bigcap_{i\in I} A^{i'}_a,$$

which concludes the proof.

□

The inequalities in the above theorem can't be replaced by equalities.

Example 12.1.24. Consider the fuzzy set A^i in the universal set X defined as

$$\mu_{A^i}(x) = 1 - \frac{1}{i}$$

for all $x \in X$ and $i \in \mathbb{N}$. Then for any $x \in X$,

$$\mu_{\bigcup\limits_i A^i}(x) = \sup_i \mu_{A^i}(x) = \sup_i \left(1 - \frac{1}{i}\right) = 1.$$

Let $a = 1$. Then

$$\left(\bigcup_i A^i\right)_1 = X.$$

However, for any $i \in \mathbb{N}$, $A^i_1 = \emptyset$ because, for any $x \in X$,

$$\mu_{A^i}(x) = 1 - \frac{1}{i} < 1.$$

Hence

$$\bigcup_i A^i_1 = \bigcup_i \emptyset = \emptyset \neq X = \left(\bigcup_i A^i\right)_1.$$

This shows that equality is not possible always in case of property 1 of the above theorem. A similar example can be used to show the same for property 2.

Theorem 12.1.25. Let A and B be two fuzzy sets in the universal set X. Then for all $a \in [0, 1]$,

1. $A \subseteq B$ iff $A_a \subseteq B_a$ and $A \subseteq B$ iff $A^{'}_a \subseteq B^{'}_a$;
2. $A = B$ iff $A_a = B_a$ and $A = B$ iff $A^{'}_a = B^{'}_a$

Proof.

1. To prove the first proposition, we assume that there exists $a_0 \in [0, 1]$ such that $A_{a_0} \nsubseteq B_{a_0}$, that is, there exists $x_0 \in X$ such that $x_0 \in A_{a_0}$ but $x_0 \notin B_{a_0}$. Then, $\mu_A(x_0) \geq a_0$ and $\mu_B(x_0) < a_0$. Hence, $\mu_B(x_0) < \mu_A(x_0)$, which contradicts that $A \subseteq B$. Now assume that $A \nsubseteq B$; that is, there exists $x_0 \in X$ such that $\mu_B(x_0) < \mu_A(x_0)$. Let $a = \mu_A(x_0)$. Then $x_0 \in A_a$ and $x_0 \notin B_a$, which demonstrates that $A_a \subseteq B_a$ is not satisfied for all $a \in [0, 1]$.

   Now we prove the second proposition. The first part is similar to the previous proof. For the second part, assume that $A \nsubseteq B$. Then there exists $x_0 \in X$ such that $\mu_A(x_0) > \mu_B(x_0)$. Let a be any number between $\mu_A(x_0)$ and $\mu_B(x_0)$. Then $x_0 \in A^{'}_a$ $x_0 \notin B^{'}_a$. Hence $A^{'}_a \nsubseteq B^{'}_a$, which demonstrates that $A^{'}_a \subseteq B^{'}_a$ is not satisfied for all $a \in [0, 1]$.

2. Left as exercise.

□

The above theorem establishes that the properties of fuzzy set inclusion and equality are both cutworthy and strong cutworthy.

Theorem 12.1.26. For any fuzzy set A in the universal set X, the following properties hold

1. $A_a = \bigcap\limits_{b<a} A_b = \bigcap\limits_{b<a} A'_b$;

2. $A'_a = \bigcup\limits_{a<b} A_b = \bigcup\limits_{a<b} A'_b$.

Proof. 1. For any $b < a$, we clearly have $A_a \subseteq A_b$. Hence

$$A_a \subseteq \bigcup_{b<a} A_b.$$

Now, for all $x \bigcap\limits_{b<a} A_b$ and for any $\epsilon > 0$, we have $x \in A_{a-\epsilon}$ (since $a - \epsilon < a$), which means that $\mu_A(x) \geq a - \epsilon$. Since ϵ is an arbitrary number, let $\epsilon \to 0$. This results in $\mu_A(x) \geq a$ (that is, $x \in A_a$). Hence,

$$\bigcap_{b<a} A_b \subseteq A_a,$$

which concludes the proof of the first equation. The proof of the second equation is analogous.

2. Left as exercise.

□

The α-cuts and strong α-cuts have significant role in representing fuzzy sets. In fact, each fuzzy set can be uniquely represented by its α-cuts and strong α-cuts. This helps in extending the properties of crisp sets to fuzzy sets. For this, a special kind of crisp set is defined ${}_aA$ in the next unit.

Along with the α-cuts, convexity also plays a role in fuzzy set theory. By contrast to classical set theory, however, convexity conditions are defined with reference to the membership function rather than the support of the fuzzy set.

Definition 12.1.27. A fuzzy set $\tilde{A}$ is convex iff

$$\mu_{\tilde{A}}(cx + (1-c)y) \geq \min\{\mu_{\tilde{A}}(x), \mu_{\tilde{A}}(y)\}, \;\; x, y \in X, \;\; c \in [0,1].$$

Alternatively, a fuzzy set is convex if all α -level sets are convex. In the figure given below, the set on the right is convex and that on the left is not.

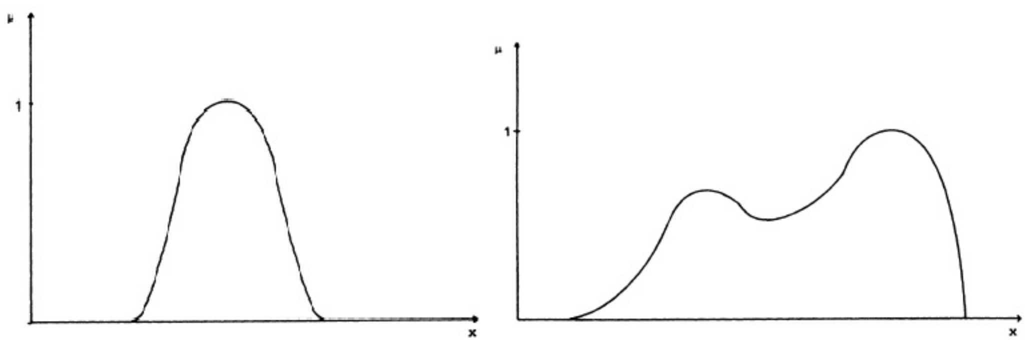


Figure 12.1.3: Convex and Non-convex set

Definition 12.1.28. For a fuzzy set $\tilde{A}$, the cardinality $|\tilde{A}|$ is defined as

$$|\tilde{A}| = \sum_{x \in X} \mu_{\tilde{A}}(x),$$

and

$$\|\tilde{A}\| = \frac{|\tilde{A}|}{|x|}$$

is called the relative cardinality of $\tilde{A}$.

Obviously, the relative cardinality of a fuzzy set depends on the cardinality of the universe . So you have to choose the same universe if you want to compare fuzzy sets by their relative cardinality.

Example 12.1.29. For the fuzzy set "comfortable type of house for a four-person family" from (12.1.2), the cardinality is

$$|\tilde{A}| = 0.2 + 0.5 + 0.8 + 1 + 0.7 + 0.3 = 3.5.$$

Its relative cardinality is

$$\|\tilde{A}\| = \frac{3.5}{10} = 0.35$$

The relative cardinality can be interpreted as the fraction of elements of X being in $\tilde{A}$, weighted by their degrees of membership in $\tilde{A}$. For infinite X, the cardinality is defined by $|\tilde{A}| = \int\limits_{x} \mu_{\tilde{A}}(x)dx$. Of course, $|\tilde{A}|$ does not always exist.

Exercise 12.1.30. 1. Model the following expressions as fuzzy sets :

(a) Very small numbers.

(b) Numbers approximately between 10 and 20.

2. Determine all a-level sets and all strong a-level sets for the following fuzzy set

$$\tilde{A} = \{(x, \mu_{\tilde{C}}(x) : \; x \in R\}$$

$$\begin{aligned} \text{where } \; \mu_{\tilde{C}}(x) &= 0 \; \text{ for } \; x \leq 10 \\ &= \frac{1}{1 + (x-10)^{-2}}, \; \text{ for } \; x > 10. \end{aligned}$$

3. Let A be a fuzzy set defined by

$$A = \frac{0.5}{x_1} + \frac{0.4}{x_2} + \frac{0.7}{x_3} + \frac{0.8}{x_4} + \frac{1}{x_3}.$$

List all the α-cuts and strong α-cuts of A.

4. Let $X = \{1, \ldots, 10\}$. Determine the cardinalities and relative cardinalities of the following fuzzy sets:

(a) $\tilde{B} = \{(2, 0.4), (3, 0.6), (4, 0.8), (5, 1), (6, 0.8), (7, 0.6), (8, 0.4)\}$.

(b) $\tilde{C} = \{(2, 0.4), (4, 0.8), (5, 1), (7, 0.6)\}$.

Few Probable Questions

1. Define the α-cut of a fuzzy set. Prove that for all $\alpha \in [0,1]$, $A'_{\alpha} \subseteq A_{\alpha}$.

2. Define strong α-cut of a fuzzy set. Show that $(A \cap B)'_{\alpha} = A'_{\alpha} \cap B'_{\alpha}$ for every $\alpha \in [0,1]$.

3. Define the union of two fuzzy sets. Show that $(A \cup B)'_{\alpha} = A'_{\alpha} \cup B'_{\alpha}$ for every $\alpha \in [0,1]$.

4. Show that for any collection of fuzzy sets A^i over a universal set X, where i belongs to the index set I, we have
$$\bigcup_{i \in I} A^i_{\alpha} \subseteq \left(\bigcup_{i \in I} A^i \right)_{\alpha}.$$
Can the inequality be replaced by equality? Justify.

Unit 13

Course Structure

- Decomposition theorems
- Zadeh's extension principle.

13.1 Introduction

We saw in the previous unit that α-cuts and strong α-cuts are some sort of bridge between the crisp and fuzzy sets. The first part of this unit deals with the decomposition theorems, which play a very significant role in the later sections as we shall see. The second section deals with the extension principle.

Objectives

After reading this unit, you will be able to

- deduce the decomposition theorems of fuzzy sets
- get an idea of the extension principle and discuss its few implications

13.2 Decomposition Theorems of Fuzzy sets

Let us try to understand the decomposition with an example. Let $X = \{a, b, c, d, e, f\}$ and let the fuzzy set $A = \dfrac{0.2}{a} + \dfrac{0.4}{b} + \dfrac{0.6}{c} + \dfrac{0.8}{d} + \dfrac{1}{e}$. The different α-cuts of A defined by the characteristic functions are as follows.

$$\begin{aligned}
&\text{For } \alpha = 0.2,\ \ A_{0.2} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} + \frac{1}{e},\\
&\text{For } \alpha = 0.4,\ \ A_{0.4} = \frac{0}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} + \frac{1}{e},\\
&\text{For } \alpha = 0.6,\ \ A_{0.6} = \frac{0}{a} + \frac{0}{b} + \frac{1}{c} + \frac{1}{d} + \frac{1}{e},\\
&\text{For } \alpha = 0.8,\ \ A_{0.8} = \frac{0}{a} + \frac{0}{b} + \frac{0}{c} + \frac{1}{d} + \frac{1}{e},\\
&\text{For } \alpha = 1,\ \ A_{1} = \frac{0}{a} + \frac{0}{b} + \frac{0}{c} + \frac{0}{d} + \frac{1}{e}.
\end{aligned}$$

We now convert each of the a-cuts into a special fuzzy set ${}_aA$, defined for $x \in X$ as

$$\mu_{{}_aA}(x) = a.\mu_{A_a}(x).$$

Then, we get the following sets.

$$\begin{aligned}
&\text{For } \alpha = 0.2,\ \ {}_{0.2}A = \frac{0.2}{a} + \frac{0.2}{b} + \frac{0.2}{c} + \frac{0.2}{d} + \frac{0.2}{e},\\
&\text{For } \alpha = 0.4,\ \ {}_{0.4}A = \frac{0}{a} + \frac{0.4}{b} + \frac{0.4}{c} + \frac{0.4}{d} + \frac{0.4}{e},\\
&\text{For } \alpha = 0.6,\ \ {}_{0.6}A = \frac{0}{a} + \frac{0}{b} + \frac{0.6}{c} + \frac{0.6}{d} + \frac{0.6}{e},\\
&\text{For } \alpha = 0.8,\ \ {}_{0.8}A = \frac{0}{a} + \frac{0}{b} + \frac{0}{c} + \frac{0.6}{d} + \frac{0.6}{e},\\
&\text{For } \alpha = 1,\ \ {}_{1}A = \frac{0}{a} + \frac{0}{b} + \frac{0}{c} + \frac{0}{d} + \frac{1}{e}.
\end{aligned}$$

The standard fuzzy union of all these special fuzzy sets yield the original fuzzy set, that is,

$$A =_{0.2} A \cup_{0.4} A \cup_{0.6} A \cup_{0.6} A \cup_{1} A.$$

Theorem 13.2.1. (First Decomposition Theorem). For every fuzzy set A in the universal set X,

$$A = \bigcup_{a\in[0,1]} {}_aA,$$

where the symbols have their usual meaning.

Proof. For each particular $x \in X$, let $a = \mu_A(x)$. Then,

$$\begin{aligned}
\mu_{\bigcup\limits_{a\in[0,1]} {}_aA}(x) &= \sup_{a\in[0,1]} \mu_{{}_aA}(x)\\
&= \max\{\sup_{a\in[0,\alpha]} \mu_{{}_aA}(x),\ \sup_{a\in(\alpha,1]} \mu_{{}_aA}(x)\}.
\end{aligned}$$

Foe each $a \in (\alpha, 1]$, we have $\mu_A(x) = \alpha < a$ and hence, $\mu_{{}_aA}(x) = 0$. On the other hand, for each $a \in [0, \alpha]$, we have $\mu_A(x) = \alpha \geq a$, therefore, $\mu_{{}_aA}(x) = a$. Hence

$$\mu_{\bigcup\limits_{a\in[0,1]} {}_aA}(x) = \sup_{a\in[0,\alpha]} a = \alpha = \mu_A(x).$$

Since the same argument is valid for each $x \in X$, the validity of the theorem is established. □

Theorem 13.2.2. (Second Decomposition Theorem). For any fuzzy set A in X, we have

$$A = \bigcup_{a \in [0,1]} {}_aA',$$

where ${}_aA'$ denotes a special fuzzy set defined by

$$\mu_{{}_aA'}(x) = a.\mu_{A'_a}(x)$$

where, $\bigcup$ denotes the standard fuzzy union.

Proof. Since the proof is analogous to the proof of the First Decomposition theorem, we express it in a more concise form. For each particular $x \in X$, let $\alpha = \mu_A(x)$. Then,

$$\begin{aligned} \mu_{\bigcup\limits_{a\in[0,1]} {}_aA'}(x) &= \sup_{a\in[0,1]} \mu_{{}_aA'}(x) \\ &= \max\{\sup_{a\in[0,\alpha)} \mu_{{}_aA'}(x), \sup_{a\in[\alpha,1]} \mu_{{}_aA'}(x)\} \\ &= \sup_{a\in[0,\alpha)} a = \alpha = \mu_A(x). \end{aligned}$$

□

Definition 13.2.3. The set of all levels $a \in [0, 1]$ that represent distinct a-cuts of a given fuzzy set A is called a level set of A. Formally,

$$\Lambda(A) = \{a : \ \mu_A(x) = a \ \text{ for some } \ x \in X\},$$

where Λ denotes the level set of fuzzy set A defined on X.

Theorem 13.2.4. (Third Decomposition Theorem). For every fuzzy set A in the universal set X,

$$A = \bigcup_{a \in \Lambda(A)} {}_aA,$$

where $\Lambda(A)$ is the level set of A.

Proof. Analogous to the proofs of the other decomposition theorems. □

Exercise 13.2.5. Let A be the fuzzy set on $X = \{a, b, c, d, e\}$ defined as

$$A = \frac{0.2}{a} + \frac{0.4}{b} + \frac{0.6}{c} + \frac{0.8}{d} + \frac{1}{e}.$$

Verify the second decomposition theorem for this fuzzy set.

13.3 Zadeh's Extension Principle

The Extension principle is a basic concept that transforms a given fuzzy set of one universal set to another universal set, provided we have a point-to-point mapping of a function $f(\cdot)$ known.

Suppose $f : X \to Y$ is a crisp function and $A \in \mathscr{F}(X)$ defined as

$$A = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \ldots + \frac{\mu_A(x_n)}{x_n}.$$

Then the extension principle states that the image of fuzzy set A under the mapping $f(\cdot)$ can be expressed as a fuzzy set B given as

$$B = \frac{\mu_A(x_1)}{y_1} + \frac{\mu_A(x_2)}{y_2} + \ldots + \frac{\mu_A(x_n)}{y_n},$$

where $y_i = f(x_i)$ for $i = 1(1)n$.

If f us a many-to-one function, then there exists $x_1, x_2 \in X$ such that $x_1 \neq x_2$ and $f(x_1) = f(x_2) = y^*$, $y^* \in Y$. In this case, the membership value of the fuzzy set B at $y = Y^*$ will be given by

$$\mu_B(y^*) = \max\{\mu_A(x_1), \mu_A(x_2)\}.$$

In general, we have,

$$\mu_B(y) = \max_{x \in f^{-1}(y)} \mu_A(x),$$

where $f^{-1}(y)$ denotes the set of all points x in the universe of discourse X such that $f(x) = y$. For infinite case, we have,

$$\mu_B(y) = \sup_{x \in f^{-1}(y)} \mu_A(x).$$

This is called the extension principle.

Example 13.3.1. Let us consider a fuzzy set A with the universal set $X = [-10, 10]$ given as

$$A = \sum_{x \in X} \frac{\mu_A(x)}{x} = \frac{0.1}{(-2)} + \frac{0.4}{(-1)} + \frac{0.8}{0} + \frac{0.9}{1} + \frac{0.3}{2}.$$

We find a fuzzy set B with the universal set $Y = [-10, 10]$ using the extension principle for mapping function $y = f(x) = x^2 + x - 3$. The images of the points $-2, -1, 0, 1, 2$ under the mapping f are given below:

$$f(-2) = -1, \quad f(-1) = -3, \quad f(0) = -3, \quad f(1) = -1, \quad f(2) = 3.$$

The images of $-2, -1, 0, 1, 2$ are $-1, -3$ and 3. We are left only to find their membership values in the universe of discourse Y. f is clearly a many-to-one function where -2 and 1 have the same image and $-1, 0$ have the same image. Saying the other way round, the point -1 has two distinct preimages under the mapping f. So, the membership value for the point -1 is given by

$$\mu_B(-1) = \max\{\mu_A(-2), \mu_A(1)\} = \max\{0.1, 0.9\} = 0.9.$$

Similarly, -3 has two preimages -1 and 0. Thus the membership value for the point -3 can be similarly found out as

$$\mu_B(-3) = \max\{\mu_A(-1), \mu_A(0)\} = 0.8.$$

Thus, by the extension principle, the required fuzzy set is given by

$$B = \frac{0.9}{(-1)} + \frac{0.8}{(-3)} + \frac{0.3}{3}.$$

Example 13.3.2. Let $A \in \mathscr{F}(X)$, where $X = [-50, 50]$ is given by

$$A = \frac{0.2}{0} + \frac{0.7}{1} + \frac{0.5}{2} + \frac{0.6}{3} + \frac{0.1}{4}.$$

We find a fuzzy set with the universal set $Y = [-50, 50]$ using the extension principle for $y = f(x) = -3x^2 + x$. As in the previous example, we find the images of the points $0, 1, 2, 3, 4$ under the mapping f. They are written as follows:

$$f(0) = 0, \quad f(1) = -2, \quad f(2) = -10, \quad f(3) = -24, \quad f(4) = -44.$$

This function f is one-one. So, on simple application of the extension principle, the resulting fuzzy set is given by

$$\begin{aligned} B &= \frac{0.2}{f(0)} + \frac{0.7}{f(1)} + \frac{0.5}{f(2)} + \frac{0.6}{f(3)} + \frac{0.1}{f(4)} \\ &= \frac{0.2}{0} + \frac{0.7}{(-2)} + \frac{0.5}{(-10)} + \frac{0.6}{(-24)} + \frac{0.1}{(-44)}. \end{aligned}$$

Extension Principle (Generalized)

Suppose f is a function from n-dimensional Cartesian product space $X_1 \times X_2 \times \cdots \times X_n$ to a one dimensional universe of discourse Y such that $y = f(x_1, x_2, \ldots, x_n)$ and suppose that $A_1, A_2, \ldots, A_n$ are fuzzy sets on the sets $X_1, X_2, \ldots, X_n$ respectively. Then the extension principle asserts that the membership function for the fuzzy set B, induced by the mapping f is defined by

$$\begin{aligned} \mu_B(y) &= \max_{(x_{i1},\ldots,x_{in})=f^{-1}(y)} \min_i \mu_{A_i}(x_i) && \text{if } f^{-1}(y) \neq \emptyset \\ &= 0 && \text{if } f^{-1}(y) = \emptyset. \end{aligned}$$

Exercise 13.3.3. Let $A \in \mathscr{F}(X)$, where $X = [-10, 10]$ where

$$A = \frac{0.5}{(-1)} + \frac{0.8}{0} + \frac{1}{1} + \frac{0.4}{2}.$$

Find the fuzzy set $B = f(A)$, with the universal set $Y = [-10, 10]$ where $y = f(x) = x^2$.

Few Probable Questions

1. State and prove the first decomposition theorem.
2. State and prove the second decomposition theorem.
3. Define the level set for a fuzzy set A. State and prove the third decomposition theorem.

Unit 14

Course Structure

- Fuzzy Relations: Crisp versus fuzzy relations, fuzzy matrices and fuzzy graphs, composition of fuzzy relations, relational join, binary fuzzy relations.

14.1 Introduction

We are familiar with the idea of relations in the classical set theory. Basically, if X and Y are two crisp sets, then a relation on them is a crisp subset of $X \times Y$. One should has immediately guessed the definition of fuzzy relations on the basis of the definitions of fuzzy sets. I fact, in a similar manner, fuzzy relations are fuzzy subsets of $X \times Y$, that is, mappings from $X \to Y$. They have been studied by a number of authors, in particular by Zadeh [1965, 1971], Kaufmann [1975], and Rosenfeld [1975]. Applications of fuzzy relations are widespread and important. We shall consider some of them and point to more possible uses at the end of this unit. We shall exemplarily consider only binary relations . A generalization to n-ary relations is straightforward.

Objectives

After reading this unit, you will be able to

- capture the essence of crisp relations from a new perspective
- define fuzzy relations and compare them with the previously defined crisp relations
- gather knowledge on the various operations on fuzzy relations
- define certain compositions of fuzzy relations and compare them with their crisp counterparts

14.2 Fuzzy Relations

A crisp relation represents the presence or absence of association, interaction or interconnection between the elements of two or more sets. This concept can be generalised to allow for various degrees or strengths of association between elements. Degrees of association can be represented by membership grades in a fuzzy relation in the same way as degrees of membership are represented in the fuzzy set. In fact, a crisp relation can be viewed as a particular case of fuzzy relation.

14.2.1 Crisp Relation

A subset of the Cartesian product $X_1 \times X_2 \times \cdots \times X_r$ is called an r-ary relation over $X_1, X_2, \ldots, X_r$. Again, the most common case is for $r = 2$; in this situation, the relation is a subset of the Cartesian product $X_1 \times X_2$. This subset is called a binary relation from X_1 into X_2. If three, four, or five sets are involved, the relations are called ternary, quarternary and quinary respectively.

Every crisp relation R can be defined by a characteristic function which assigns the value 1 to every tuple of the universal set belonging to the relation and 0 to every tuple not belonging to it. It can also be put this way: the "strength" or "degree" of this relationship between the ordered pairs of elements in each universe is measured by the characteristic function, denoted by χ_R, where the value 1 denotes complete relationship and the value 0 denotes no relationship, that is,

$$\begin{aligned} \chi_R(x, y) &= 1 \quad \text{if } (x, y) \in R \\ &= 0 \quad \text{if } (x, y) \notin R. \end{aligned}$$

One can think of this strength of relation as a mapping from ordered pairs of sets defined on the universal sets to the real numbers. When the universes, or sets are finite, the relation can be conveniently represented by a matrix, called a *relation matrix*. An r-ary relation can be represented by the r-dimensional relation matrix. Hence, binary relations can be represented by two-dimensional matrices.

Example 14.2.1. Let

$$R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

be a relation between two sets $X = \{a, b, c\}$ and $Y = \{1, 2, 3\}$. This relation matrix shows that every element of X is completely related to every element of Y. This relation can also be represented by the *Saggital diagram* given in figure 14.2.1.

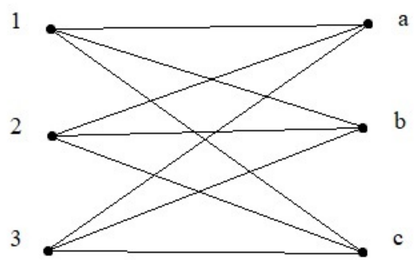


Figure 14.2.1

Operations on Crisp sets

Define R and S as two separate relations on the Cartesian universe $X \times Y$ and define null relation and the complete relation as the relation matrices O and E respectively. An example of a 4×4 form of the matrices

O and E are given as

$$O = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

The following operations for two crisp relations R and S can now be defined as

1. **Union:** $R \cup S$ with $\chi_{R \cup S}(x, y) = \max\{\chi_R(x, y), \chi_S(x, y)\}$;
2. **Intersection:** $R \cap S$ with $\chi_{R \cap S}(x, y) = \min\{\chi_R(x, y), \chi_S(x, y)\}$;
3. **Complement:** $\overline{R}$ with $\chi_{\overline{R}}(x, y) = 1 - \chi_R(x, y)$;
4. **Containment:** $R \subset S$ if $\chi_R(x, y) \leq \chi_S(x, y)$.

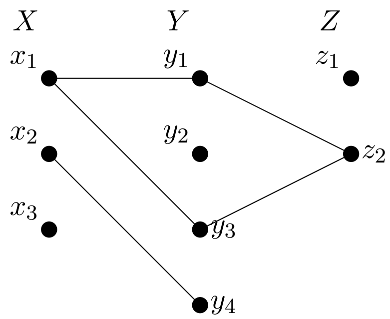
The properties of commutativity, associativity, idempotence, De-Morgan's principles all hold for crisp relations. The null relation O and the complete relation E are analogous to the null set $\emptyset$ and the whole set X, respectively.

Composition of Relations

Let $R \subset X \times Y$ and $S \subset Y \times Z$. Is it possible to find a relation between X and Z via the relations R and S? The answer is affirmative and is done using a relation called composition. Let us illustrate this with an example. Let R and S be given by

$$R = \{(x_1, y_1), (x_1, y_3), (x_2, y_4)\} \quad \text{and} \quad S = \{(y_1, z_2), (y_3, z_2)\}.$$

The sagittal diagram for these two is given in figure below. From the above diagram, it is clear that the only



"path" between the relation R and S is the two routes that start at x_1 and end at z_2 (that is, $x_1 - y_1 - z_2$ and $x_1 - y_3 - z_2$). There are two common forms of composition operation : one is called the *max-min composition* and the other is called the *max-product composition*. These two compositions along with the corresponding characteristic functions are given below.

Max-min Composition

$$\begin{aligned} T &= R \circ S \\ \chi_T(x, z) &= \max_{y \in Y} \{\min\{\chi_R(x, y), \chi_S(y, z)\}\}. \end{aligned}$$

Max-product Composition (or Max-dot Composition)

$$\begin{aligned} T &= R \circ S \\ \chi_T(x,z) &= \max_{y \in Y} \{\chi_R(x,y) \cdot \chi_S(y,z)\} . \end{aligned}$$

It is worthwhile to note that for crisp case, the max-min as well as the max-product compositions are identical. Let us see the above example.

Example 14.2.2. The matrix representations for the crisp relations R and S discussed above are given as

$$R = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{matrix} y_1 & y_2 & y_3 & y_4 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

and

$$S = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{array} \begin{array}{c} \begin{matrix} z_1 & z_2 \end{matrix} \\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \end{array}$$

The resulting relation T would then be determined by max-min composition.

$$\begin{aligned} \mu_T(x_1,z_1) &= \max\{\min(1,0), \min(0,0), \min(1,0), \min(0,0)\} \\ &= 0 \\ \mu_T(x_1,z_2) &= \max\{\min(1,1), \min(0,0), \min(0,1), \min(0,0)\} \\ &= 1 \\ \mu_T(x_2,z_1) &= \max\{\min(0,0), \min(0,0), \min(0,0), \min(1,0)\} \\ &= 0 \\ \mu_T(x_2,z_2) &= \max\{\min(0,1), \min(0,0), \min(0,1), \min(1,0)\} \\ &= 0 \\ \mu_T(x_3,z_1) &= \max\{\min(0,0), \min(0,0), \min(0,0), \min(0,0)\} \\ &= 0 \\ \mu_T(x_3,z_2) &= \max\{\min(0,1), \min(0,0), \min(0,1), \min(0,0)\} \\ &= 0 \end{aligned}$$

Thus, the relation matrix T is given by

$$T = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{matrix} z_1 & z_2 \end{matrix} \\ \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \end{array}$$

Check that the max-dot product is identical for this T.

14.2.2 Fuzzy Relations

Fuzzy relation generalises the crisp relation into one that allows "partial" relationship. For example, a fuzzy relation "friend" describes the degree of friendship between two persons (in contrast to either being friend or not in crisp relation).

Definition 14.2.3. A fuzzy relation R is a subset of two universal sets X and Y where, the degree of association between any two elements of X and Y is given by the membership function $\mu_R : X \times Y \to [0, 1]$.

The "strength" of the relation between ordered pairs of X and Y is measured with the function μ_R. This definition can also be generalised for any n-dimensional Cartesian product of the universal sets.

Example 14.2.4. Let R be a fuzzy relation between two sets $X = \{\text{New York City, Paris}\}$ and $Y = \{\text{Beijing, New York City, London}\}$, which represents the relational concept "very far". This concept can be written in notation as

$$R = \frac{1}{\text{(NYC, Beijing)}} + \frac{0}{\text{(NYC, NYC)}} + \frac{0.6}{\text{(NYC, London)}} + \frac{0.9}{\text{(Paris, Beijing)}} + \frac{0.7}{\text{(Paris, NYC)}} + \frac{0.3}{\text{(Paris, London)}}.$$

This relation can also be represented by the following matrix.

$$R = \begin{array}{c} \\ \text{Beijing} \\ \text{NYC} \\ \text{London} \end{array} \begin{array}{c} \begin{array}{cc} \text{NYC} & \text{Paris} \end{array} \\ \begin{pmatrix} 1 & 0.9 \\ 0 & 0.7 \\ 0.6 & 0.3 \end{pmatrix} \end{array}$$

Note 14.2.5. Such matrices whose entries lie in the set $[0, 1]$ are called *fuzzy matrices*. All the relation matrices we have so far encountered are fuzzy matrices.

14.2.3 Binary Fuzzy Relations

Given a fuzzy relation $R \subset X \times Y$, its domain is a fuzzy set on X, denoted by dom$R$, whose membership function is defined by

$$\mu_{\text{dom}R}(x) = \max_{y \in Y} \mu_R(x, y), \text{ for } x \in X.$$

That is, each element of the set X belongs to the domain of R to the degree equal yo the strength of its strongest relation to any member of Y whose membership function is defined as

$$\mu_{\text{ran}R}(x) = \max_{x \in X} \mu_R(x, y), \text{ for } y \in Y,$$

ran$R$ being the range of R. This means that the strength of the strongest relation that each element of Y has to an element of X is equal to the degree of that element's membership in the range of R. The height of a fuzzy relation R is a number $h(R)$ defined by

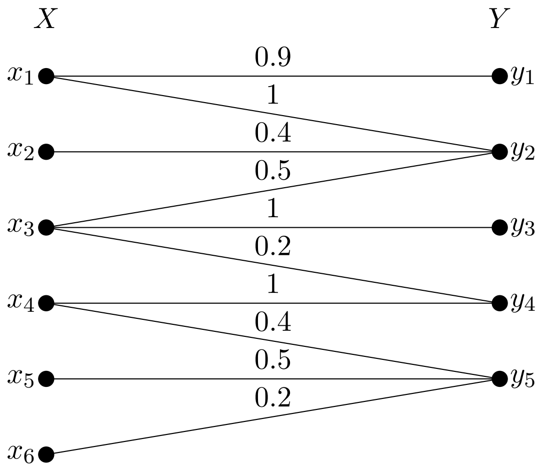
$$h(R) = \max_{y \in Y} \max_{x \in X} \mu_R(x, y).$$

A convenient representation of R are the membership matrices $R = [\mu_R(x, y)]$. Another useful representation of binary relations is the saggital diagram. The only difference with the saggital diagrams of the crisp relations is that the lines are labelled with the values of the membership grades.

Example 14.2.6. Let R be a fuzzy relation on $X \times Y$ given by

$$R = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{array} \begin{array}{c} \begin{array}{ccccc} y_1 & y_2 & y_3 & y_4 & y_5 \end{array} \\ \begin{pmatrix} 0.9 & 1 & 0 & 0 & 0 \\ 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0.5 & 1 & 0.2 & 0 \\ 0 & 0 & 0 & 1 & 0.4 \\ 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0.2 \end{pmatrix} \end{array}$$

The corresponding saggital diagram is given below.



14.2.4 Operations on Fuzzy Relations

Let R and S be fuzzy relations on the Cartesian space $X \times Y$. Then the following operations apply for the membership values for various set operations.

1. **Union:** $R \cup S$ with $\mu_{R\cup S}(x, y) = \max\{\mu_R(x, y), \mu_S(x, y)\}$.
2. **Intersection:** $R \cap S$ with $\mu_{R\cap S}(x, y) = \min\{\mu_R(x, y), \mu_S(x, y)\}$.
3. **Complement:** $\overline{R}$ with $\mu_{\overline{R}}(x, y) = 1 - \mu_R(x, y)$.
4. **Containment:** $R \subset S \Rightarrow \mu_R(x, y) \leq \mu_S(x, y)$.

Just as for crisp relations, the properties of commutativity, associativity, distributivity, idempotency all hold for fuzzy relations.

14.2.5 Fuzzy Cartesian Product and Composition

Because fuzzy relations are in general fuzzy sets, we can define the Cartesian product to be a relation between two or more fuzzy sets. Let $A \in \mathscr{F}(X)$ and $B \in \mathscr{F}(Y)$. Then the Cartesian product between A and B will result in a fuzzy relation R, or

$$A \times B = R \subset X \times Y$$

with the membership function

$$\mu_R(x, y) = \mu_{A\times B}(x, y) = \min\{\mu_A(x), \mu_B(y)\}$$

Example 14.2.7. Let $A \in \mathscr{F}(X)$ and $B \in \mathscr{F}(Y)$ where $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2\}$. The sets A and B are given below.

$$A = \frac{0.2}{x_1} + \frac{0.5}{x_2} + \frac{1}{x_3} \text{ and } B = \frac{0.3}{y_1} + \frac{0.9}{y_2}.$$

Note that A can be represented as a row vector of size 3×1 and B can be represented as a column vector of size 1×2. The Fuzzy Cartesian product results in a fuzzy relation R of size 3×2 given below.

$$A \times B = R = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{pmatrix} 0.2 & 0.2 \\ 0.3 & 0.5 \\ 0.3 & 0.9 \end{pmatrix} \end{array}$$

14.2.6 Fuzzy Composition

Fuzzy composition can be defined just as it was defined for crisp sets. Suppose $R \subset X \times Y$ and $S \subset Y \times Z$ are two fuzzy relations. $T \subset X \times Z$ is a fuzzy relation whose membership function is determined by max-min composition and max-dot composition in the following manner.

Max-min Composition

$$\begin{aligned} T &= R \circ S \\ \mu_T(x, z) &= \max_{y \in Y} \{\min\{\mu_R(x, y), \mu_S(y, z)\}\}. \end{aligned}$$

Max-dot Composition

$$\begin{aligned} T &= R \circ S \\ \mu_T(x, z) &= \max_{y \in Y} \{\mu_R(x, y) \cdot \mu_S(y, z)\}. \end{aligned}$$

Note 14.2.8. In the case of fuzzy relations, the membership functions for the max-min and max-dot compositions need not be identical. Also, it should be noted that neither crisp nor fuzzy compositions are commutative in general, that is,

$$R \circ S \neq S \circ R.$$

Example 14.2.9. Let $X = \{x_1, x_2\}$, $Y = \{y_1, y_2\}$ and $Z = \{z_1, z_2, z_3\}$. Consider the fuzzy relations $R \subset X \times Y$ and $S \subset Y \times Z$ as follows.

$$R = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{pmatrix} 0.7 & 0.5 \\ 0.8 & 0.4 \end{pmatrix} \end{array}$$

and

$$S = \begin{array}{c} \\ y_1 \\ y_2 \end{array} \begin{array}{c} \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{pmatrix} 0.9 & 0.6 & 0.2 \\ 0.1 & 0.7 & 0.5 \end{pmatrix} \end{array}$$

We find the fuzzy relation $T = R \circ S$ on $X \times Z$ using the max-min and max-dot compositions as follows.

1. Using Max-min Composition:

$$T = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{pmatrix} 0.7 & 0.6 & 0.5 \\ 0.8 & 0.6 & 0.4 \end{pmatrix} \end{array}$$

2. Using Max-dot Composition:

$$T = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{ccc} z_1 & z_2 & z_3 \end{array} \\ \begin{pmatrix} 0.63 & 0.42 & 0.25 \\ 0.72 & 0.48 & 0.20 \end{pmatrix} \end{array}$$

[The calculations are left as exercise]

14.2.7 Equivalence Relations

A crisp relation R on a universe X can also be thought of as a relation from X to X. The relation R is an equivalence relation if it is reflexive, symmetric and transitive. These can be defined using the characteristic functions as given below.

1. **Reflexive:** $(x, x) \in R$ or $\chi_R(x, x) = 1$, for all $x \in X$;
2. **Symmetric:** $(x, y) \in R \Rightarrow (y, x) \in R$ or $\chi_R(x, y) = 1 \Rightarrow \chi_R(y, x) = 1$ for all $x, y \in X$;
3. **Transitivity:** $(x, y), (y, z) \in R \Rightarrow (x, z) \in R$ or $\chi_R(x, y) = 1, \chi_R(y, z) = 1 \Rightarrow \chi_R(x, z) = 1$ for all $x, y, z \in X$.

The reflexive, symmetry and transitivity properties can be extended to fuzzy relations by defining them in terms of their membership functions. Thus, a fuzzy relation defined on a universal set X is reflexive if

$$\mu_R(x, x) = 1 \text{ for all } x \in X.$$

If this is not the case for some $x \in X$, then the relation is called *irreflexive*; if it is not satisfied for all $x \in X$, the relation is called *anti-reflexive*. A weaker form of reflexivity, known as ϵ-reflexivity, is sometimes defined by the condition

$$\mu_R(x, x) \geq \epsilon, \text{ where } 0 < \epsilon < 1.$$

A fuzzy relation is symmetric if and only if

$$\mu_R(x, y) = \mu_R(y, x), \text{ for all } x, y \in X,$$

Whenever this equality is not satisfied for some $x, y \in X$, the relation is called asymmetric. Furthermore, when $\mu_R(x, y) > 0$ and $\mu_R(y, x) > 0$ implies $x = y$ for all $x, y \in X$, the relation is called antisymmetric.

A fuzzy relation R is transitive (or max-min transitive) if

$$\mu_R(x, z) \geq \max_{y \in Y} \{\min\{\mu_R(x, y), \mu_R(y, z)\}\}$$

for each pair $(x, z) \in X \times X$. A relation failing to satisfy this inequality for some members of X is called non-transitive, and if

$$\mu_R(x, z) > \max_{y \in Y} \{\min\{\mu_R(x, y), \mu_R(y, z)\}\}$$

holds for all $(x, z) \in X \times X$, then the relation is called anti-transitive.

Example 14.2.10. Let R be the fuzzy relation defined on the set of cities and representing the concept "very near". We may assume that a city is certainly (that is, to a degree of 1) very near to itself. The relation is therefore reflexive. Furthermore, if city A is very near to city B, then city B is also very near to city A and to the same degree. Hence, the relation is also symmetric. Finally, if A is very near to city B to some degree, say 0.7 and city B is very near to city C to some degree, say 0.8, it is possible, though not necessary that city A is very near to city C to a smaller degree, say 0.5. Hence, this relation is non-transitive.

Observe that the definition of max-min transitivity is based upon the max-min composition. Hence, alternative definitions of fuzzy transitivity are also possible.

14.2.8 Fuzzy Equivalence Relation

A fuzzy binary relation that is reflexive, symmetric and transitive, is known as fuzzy equivalence relation, or similarity relation.

Example 14.2.11. The fuzzy relation R given by the membership matrix

$$R = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \end{array} \begin{array}{c} \begin{array}{ccccccc} a & b & c & d & e & f & g \end{array} \\ \begin{pmatrix} 1 & 0.8 & 0 & 0.4 & 0 & 0 & 0 \\ 0.8 & 1 & 0 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0.9 & 0.5 \\ 0.4 & 0.4 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0.9 & 0.5 \\ 0 & 0 & 0.9 & 0 & 0.9 & 1 & 0.5 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0.5 & 1 \end{pmatrix} \end{array}$$

is a similarity relation. Verify!

Few Probable Questions

1. The fuzzy binary relation R is defined on sets $X = \{1, 2, \ldots, 100\}$ and $Y = \{50, 51, \ldots, 100\}$ and represents the relation "x is much smaller than y". It is defined by membership function

$$\begin{aligned} R(x, y) &= 1 - \frac{x}{y}, \ \text{for } x \le y \\ &= 0, \ \text{otherwise,} \end{aligned}$$

   where $x \in X$ and $y \in Y$.

   (a) What is the domain of R?

   (b) What is the range of R?

   (c) What is the height of R?

Unit 15

Course Structure

- Fuzzy Arithmetic: Fuzzy numbers, arithmetic operations on fuzzy numbers (multiplication and division on $\mathbb{R}^+$ only), fuzzy equations.

15.1 Introduction

Fuzzy sets were introduced by Zadeh in order to deal with uncertainties in demarcating the boundaries of sets in real life. However, often we come across various situations in which quantities are uncertain. Expressions like "about three", "approximately three", "around three", "roughly three", "almost three", "more or less three" are rampant. This makes it necessary to introduce numbers that take these uncertainties into account. So, the fuzzy numbers were introduced to indicate a real number in a hazy but practical way. Just as the real numbers reinforce the classical set theory, we have the fuzzy numbers reinforcing the fuzzy set theory. Van Nauta Lemke tried to formalise the fuzzy numbers for the first time though it was Zadeh who came out with a precise definition in the year 1975. We will start with the definition of fuzzy numbers and then go on to define the operations on them.

Objectives

After reading this unit, you will be able to

- define fuzzy numbers and various operations on fuzzy numbers
- learn about fuzzy equations and some ways of solving them

15.2 Fuzzy Numbers

As we started in the introduction, a fuzzy number is some sort of approximate number. So, intuitively it has something to do with the real number set $\mathbb{R}$. If we take the universal set to be $\mathbb{R}$, then the membership function is of the form

$$\mu_A : \mathbb{R} \to [0, 1].$$

In such a way, we are essentially defining a fuzzy set, where every real number is assigned some sort of “uncertainty" owing to their corresponding membership degrees. These sets form the basis of fuzzy numbers under certain conditions as we shall see now.

Definition 15.2.1. A fuzzy set A on $\mathbb{R}$ is called a fuzzy number if it satisfies the following properties.

1. A must be a normal fuzzy set;
2. A_α must be a closed interval for every $\alpha \in (0, 1]$;
3. The support of A must be bounded.

As opposed to real numbers, which are single numbers, the fuzzy numbers is a fuzzy set, which can be hard to grasp at an early stage. However, if we consider an exact number, then it has to be a fuzzy singleton having membership value 1 (all others being zero); if one considers the numbers “close to" 1.2, then it is a fuzzy set which contains real numbers near to 1.2 with varying membership grades on the basis of its distance from 1.2. We can similarly express the phrase “approximately" within the interval $[1, 2]$ by a fuzzy interval.

The three conditions can somewhat be interpreted by intuition. The normal fuzzy set essentially captures the approximate nature of the fuzzy numbers. For example, numbers “close to" 3 will have the membership value exactly 1. The bounded support and all α-cuts being closed intervals help in defining meaningful arithmetic operations on them. Further, since the α-cuts are required to be closed intervals for all $\alpha \in (0, 1]$, this means that every fuzzy number is a convex fuzzy set.

The following is a necessary and sufficient condition for a fuzzy set on $\mathbb{R}$ to be a fuzzy number.

Theorem 15.2.2. Let $A \in \mathscr{F}(\mathbb{R})$. Then A is a fuzzy number if and only if there exists a closed interval $[a, b] \neq \emptyset$ such that

$$\begin{aligned} \mu_A(x) &= 1, \quad \text{for } x \in [a, b] \\ &= l(x), \quad \text{for } x \in (-\infty, a) \\ &= r(x), \quad \text{for } x \in (b, \infty), \end{aligned}$$

where

1. l is a monotonically increasing function from $(-\infty, a)$ to $[0, 1]$, continuous from the right such that $l(x) = 0$ for $x \in (-\infty, w_1)$;
2. r is a monotonically decreasing function from (b, ∞) to $[0, 1]$, continuous from the left such that $r(x) = 0$ for $x \in (w_2, \infty)$.

Proof. Let A be a fuzzy number. Then A_α is a closed interval for every $\alpha \in (0, 1]$. For $\alpha = 1$, A_1 is a non-empty closed interval since A is normal. Hence, there exists a pair $a, b \in \mathbb{R}$ such that $A_1 = [a, b]$, where $a \leq b$. That is, $\mu_A(x) = 1$, for all $x \in [a, b]$ and $\mu_A(x) < 1$, for all $x \notin [a, b]$. Now, let us define a new function $l(x) = \mu_A(x)$ for any $x \in (-\infty, a)$. Then $0 \leq l(x) < 1$ since $0 \leq \mu_A(x) < 1$ for every $x \in (-\infty, a)$. Let $x \leq y < a$; then by definition 12.1.27,

$$\mu_A(x) \geq \min\{\mu_A(x), \mu_A(a)\} = \mu_A(x)$$

and $\mu_A(a) = 1$. Hence, $l(y) \geq l(x)$; that is, l is monotonic increasing.

Assume now that $l(x)$ is not continuous from the right. This means that for some $x_0 \in (-\infty, a)$ there exists a sequence of numbers $\{x_n\}$ such that $x_n \geq x_0$ for any n and

$$\lim_{n\to\infty} x_n = x_0,$$

but

$$\lim_{n\to\infty} l(x_n) = \lim_{n\to\infty} \mu_A(x_n) = \alpha > l(x_0) = \mu_A(x_0).$$

Now, $x_n \in A_\alpha$ for any n since A_α is a closed interval and hence also $x_0 \in A_\alpha$. Therefore, $l(x_0) = \mu_A(x_0) \geq \alpha$, which is a contradiction. Hence, $l(x)$ is continuous from the right. We can similarly define r and show that r is monotonically decreasing from the left.

Since A is a fuzzy number A'_0 is bounded. Hence, there exists a pair $w_1, w_2 \in \mathbb{R}$ of finite numbers such that $\mu_A(x) = 0$ for $x \in (-\infty, w_1) \cup (w_2, \infty)$.

Conversely, suppose the given condition holds. We have to show that A is a fuzzy number. Clearly, A is normal and its support, that is, A'_0 is bounded since it is a subset of $[w_1, w_2]$. We only have to show that A_α is a closed interval for every $\alpha \in (0, 1]$. Let

$$\begin{aligned} x_\alpha &= \inf\{x : l(x) \geq \alpha, x < a\}, \\ y_\alpha &= \sup\{x : r(x) \geq \alpha, x > b\}, \end{aligned}$$

for each $\alpha \in (0, 1]$. We claim that $A_\alpha = [x_\alpha, y_\alpha]$ for all $\alpha \in (0, 1]$.

For any $x_0 \in A_\alpha$, if $x_0 < \alpha$, then $l(x_0) = \mu_A(x_0) \geq \alpha$. That is, $x_0 \in \{x : l(x) \geq \alpha, x < \alpha\}$ and consequently, $x_0 \geq \inf\{x : l(x) \geq \alpha, x < \alpha\} = x_\alpha$. If $x_0 > b$, then $r(x_0) = \mu_A(x_0) \geq \alpha$. This implies that, $x_0 \in \{x : r(x) \geq \alpha, x > b\}$ and so $x_0 \leq \sup\{x : r(x) \geq \alpha, x > b\} = y_\alpha$. Obviously, $x_\alpha \leq a$ and $y_\alpha \geq b$; that is, $[a, b] \subseteq [x_\alpha, y_\alpha]$. Therefore, $x_0 \in [x_\alpha, y_\alpha]$ and hence, $A_\alpha \subseteq [x_\alpha, y_\alpha]$. It remains to prove that $x_\alpha, y_\alpha \in A_\alpha$.

By the definition of x_α, we find a sequence $\{x_n\}$ in $\{x : l(x) \geq \alpha, x < a\}$ such that $\lim_{n\to\infty} x_n = x_\alpha$, where $x_n \geq x_\alpha$ for any n. Since l is continuous from the right, we have

$$l(x_\alpha) = l\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} l(x_n) \geq \alpha.$$

Hence, $x_\alpha \in A_\alpha$. We can similarly prove that $y_\alpha \in A_\alpha$. □

Using fuzzy numbers, one can define the concept of fuzzy cardinality for fuzzy sets that are defined on finite universal sets.

Definition 15.2.3. Let A be a fuzzy set defined on a finite universal set X. Then its fuzzy cardinality, $|\tilde{A}|$ is a fuzzy number defined on $\mathbb{N}$ by the following.

$$|\tilde{A}|(|A_\alpha|) = \alpha$$

for all $\alpha \in \Lambda(A)$.

Exercise 15.2.4. Determine which fuzzy sets defined by the following functions are a fuzzy number.

1. $\mu(x) = \begin{cases} \sin(x) & \text{for } 0 \leq x \leq \pi \\ 0, & \text{otherwise;} \end{cases}$

2. $\mu(x) = \begin{cases} x & \text{for } 0 \leq x \leq 1 \\ 0, & \text{otherwise;} \end{cases}$

3. $\mu(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 10 \\ 0, & \text{otherwise;} \end{cases}$

4. $\mu(x) = \begin{cases} \min\{1, x\} & \text{for } x \geq 0 \\ 0, & \text{for } x < 0. \end{cases}$

15.2.1 Arithmetic Operations on Fuzzy Numbers

Fuzzy arithmetic is based on two properties:

1. Each fuzzy set and thus each fuzzy number is uniquely represented by its α-cuts;

2. α-cuts of fuzzy number are closed intervals of real numbers for all $\alpha \in (0, 1]$.

We are familiar with the arithmetic operations on the closed intervals from unit 11. Using them, we will now define the arithmetic operations on the fuzzy numbers in terms of arithmetic operations on their α-cuts.

Definition 15.2.5. Let A and B be two fuzzy numbers. Let $*$ denote any of the operations addition, subtraction, multiplication and division. Then we define a fuzzy set on $\mathbb{R}$ by defining its α-cut, $(A * B)_\alpha$ as

$$(A * B)_\alpha = A_\alpha * B_\alpha$$

for any $\alpha \in (0, 1]$. When $*$ is division, then it is required that $0 \notin B_\alpha$ for all $\alpha \in (0, 1]$. Due to the first decomposition theorem, $A * B$ can be represented as

$$A * B = \bigcup_{\alpha \in (0,1]} {}_\alpha(A * B).$$

Since $(A * B)_\alpha$ is a closed interval for each $\alpha \in (0, 1]$, and A and B are fuzzy numbers, $A * B$ is also a fuzzy number. Let us check the following illustration.

Example 15.2.6. Let A and B be two fuzzy numbers whose membership functions are as follows:

$$\begin{aligned}
\mu_A(x) &= 0, \text{ for } x \leq -1 \text{ and } x > 3 \\
&= \frac{x+1}{2}, \text{ for } -1 < x \leq 1 \\
&= \frac{3-x}{2}, \text{ for } 1 < x \leq 3,
\end{aligned}$$

$$\begin{aligned}
\mu_B(x) &= 0, \text{ for } x \leq 1 \text{ and } x > 5 \\
&= \frac{x-1}{2}, \text{ for } 1 < x \leq 3 \\
&= \frac{5-x}{2}, \text{ for } 3 < x \leq 5.
\end{aligned}$$

The membership functions are given in figure 15.2.1. Their α-cuts are

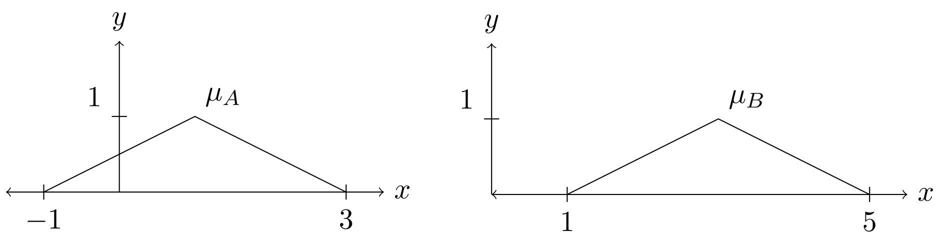


Figure 15.2.1

$$A_\alpha = [2\alpha - 1, 3 - 2\alpha], \text{ and } B_\alpha = [2\alpha + 1, 5 - 2\alpha].$$

Note that for every $\alpha > 0$, B_α does not contain 0. So the division of the α-cuts are possible. Now, we shall find the addition, subtraction, multiplication and division using interval arithmetic. For $\alpha \in (0, 1]$,

$$\begin{aligned}
(A + B)_\alpha &= A_\alpha + B_\alpha \\
&= [2\alpha - 1, 3 - 2\alpha] + [2\alpha + 1, 5 - 2\alpha] \\
&= [(2\alpha - 1) + (2\alpha - 1), (3 - 2\alpha) + (5 - 2\alpha)] \\
&= [4\alpha, 8 - 4\alpha].
\end{aligned}$$

Similarly, we find

$$\begin{aligned}
(A - B)_\alpha &= [4\alpha - 6, 2 - 4\alpha] \\
(A \cdot B)_\alpha &= [-4\alpha^2 + 12\alpha - 5, 4\alpha^2 - 16\alpha + 15], \text{ for } \alpha \in (0, 0.5] \\
&= [4\alpha^2 - 1, 4\alpha^2 - 16\alpha + 15], \text{ for } \alpha \in (0.5, 1] \\
(A/B)_\alpha &= \left[\frac{2\alpha - 1}{2\alpha + 1}, \frac{3 - 2\alpha}{2\alpha + 1}\right], \text{ for } \alpha \in (0, 0.5] \\
&= \left[\frac{2\alpha - 1}{5 - 2\alpha}, \frac{3 - 2\alpha}{2\alpha + 1}\right], \text{ for } \alpha \in (0.5, 1].
\end{aligned}$$

The resulting fuzzy numbers have their respective membership functions as follows

$$\begin{aligned}
\mu_{A+B}(x) &= 0, \text{ for } x \leq 0 \text{ and } x > 8 \\
&= \frac{x}{4}, \text{ for } 0 < x \leq 4 \\
&= \frac{8 - x}{4}, \text{ for } 4 < x \leq 8, \\
\mu_{A-B}(x) &= 0, \text{ for } x \leq -6 \text{ and } x > 2 \\
&= \frac{x + 6}{4}, \text{ for } -6 < x \leq -2 \\
&= \frac{2 - x}{4}, \text{ for } -2 < x \leq 2, \\
\mu_{A \cdot B}(x) &= 0, \text{ for } x < -5 \text{ and } x \geq 15 \\
&= \frac{\sqrt{3 - (4 - x)}}{2}, \text{ for } -5 \leq x < 0 \\
&= \frac{\sqrt{1 + x}}{2}, \text{ for } 0 \leq x < 3 \\
&= \frac{4 - \sqrt{1 + x}}{2}, \text{ for } -3 \leq x < 15, \\
\mu_{A/B}(x) &= 0, \text{ for } x < -1 \text{ and } x \geq 3 \\
&= \frac{x + 1}{2 - 2x}, \text{ for } -1 \leq x < 0 \\
&= \frac{5x + 1}{2x + 2}, \text{ for } 0 \leq x < \frac{1}{3} \\
&= \frac{3 - x}{2x + 2}, \text{ for } \frac{1}{3} \leq x < 3.
\end{aligned}$$

We can also define arithmetic operations on fuzzy numbers using Extension principle.

Definition 15.2.7. Let A and B denote fuzzy numbers. Then we define a fuzzy set $A * B$ on $\mathbb{R}$ by the equation

$$\mu_{(A*B)}(z) = \sup_{z = x*y} \min\{\mu_A(x), \mu_B(y)\}.$$

More specifically, we can write

$$\begin{aligned}
\mu_{(A+B)}(z) &= \sup_{z=x+y} \min\{\mu_A(x), \mu_B(y)\}, \\
\mu_{(A-B)}(z) &= \sup_{z=x-y} \min\{\mu_A(x), \mu_B(y)\}, \\
\mu_{(A\cdot B)}(z) &= \sup_{z=x\cdot y} \min\{\mu_A(x), \mu_B(y)\}, \\
\mu_{(A/B)}(z) &= \sup_{z=x/y} \min\{\mu_A(x), \mu_B(y)\}.
\end{aligned}$$

Exercise 15.2.8. Let A and B be two fuzzy numbers whose membership functions are given by

$$\begin{aligned}
\mu_A(x) &= \frac{x+2}{2}, \text{ for } -2 < x \leq 0 \\
&= \frac{2-x}{2}, \text{ for } 0 < x < 2 \\
&= 0, \text{ otherwise,}
\end{aligned}$$

$$\begin{aligned}
\mu_B(x) &= \frac{x-2}{2}, \text{ for } 2 < x \leq 4 \\
&= \frac{6-x}{2}, \text{ for } 4 < x \leq 6 \\
&= 0, \text{ otherwise.}
\end{aligned}$$

Calculate the fuzzy numbers $A+B,\ A-B,\ A\cdot B$, and A/B.

15.3 Fuzzy Equations

One area of fuzzy set theory in which fuzzy numbers and arithmetic operations on fuzzy numbers play a fundamental role are fuzzy equations. These are equations in which coefficients and unknowns are fuxzy numbers, and formulae are constructed by operations of fuzzy arithmetic. Such equations have a great potential applicability. Unfortunately, their theory has not been sufficiently developed as yet; moreover, some of the published work in this area is rather controversial. Due to the lack of a well-established theory of fuzzy equations, we only intend to characterize some properties of fizzy equations by discussing equations of two very simple types: $A + X = B$ and $A \cdot X = B$, where A and B are fuzzy numbers, and X is an unknown fuzzy number for which either of the equations is to be satisfied.

15.3.1 Equations of type $A + X = B$

We know that linear equations of these kind have solution $X = B - A$. However, fuzzy numbers do not exhibit such simplicity. For example, if we take two intervals $A = [a_1, a_2]$ and $B = [b_1, b_2]$, which are particular fuzzy numbers, then $B - A = [b_1 - a_2, b_2 - a_1]$ and

$$\begin{aligned}
A + (B - A) &= [a_1, a_2] + [b_1 - a_2, b_2 - a_1] \\
&= [a_1 + b_1 - a_2, a_2 + b_2 - a_1] \\
&\neq [b_1, b_2] = B
\end{aligned}$$

whenever $a_1 \neq a_2$. Thus, $X = B - A$ is not a solution of the equation.

Let $X = [x_1, x_2]$. Then for $A = [a_1, a_2]$ and $B = [b_1, b_2]$, the equation $A + X = B$ becomes

$$[a_1 + x_1, a_2 + x_2] = [b_1, b_2]$$

which results in the following system of linear equations.

$$\begin{aligned} a_1 + x_1 &= b_1 \\ a_2 + x_2 &= b_2. \end{aligned}$$

The solutions of these two equations are $[x_1, x_2] = [b_1 - a_1, b_2 - a_2]$. Since X must be an interval, we should have $b_1 - a_1 \leq b_2 - a_2$. If this inequality is true, then the given equation has the solution $X = [b_1 - a_1, b_2 - a_2]$.

Taking motivation from these intervals, we will attempt at solving the given fuzzy equation for any arbitrary fuzzy number. We have already seen that fuzzy arithmetic operations are defined using the α-cuts. So here, in order to find the solution of the equation

$$A + X = B \tag{15.3.1}$$

where $A,\ B$ are known fuzzy numbers and X is an unknown fuzzy number, we shall make use of the α-cuts of A and B to find the α-cuts of X and then find X using decomposition theorem. So let us start with the procedure.

For $\alpha \in (0, 1]$, let $A_\alpha = [a_1^\alpha, a_2^\alpha]$, $B_\alpha = [b_1^\alpha, b_2^\alpha]$ and $X_\alpha = [x_1^\alpha, x_2^\alpha]$ be the α-cuts of A, B and X. Then the equation (15.3.1) has solution if and only if

1. $b_1^\alpha - a_1^\alpha \leq b_2^\alpha - a_2^\alpha$ for every $\alpha \in (0, 1]$, and

2. $\alpha \leq \beta$ implies $b_1^\alpha - a_1^\alpha \leq b_1^\beta - a_1^\beta \leq b_2^\beta - a_2^\beta \leq b_2^\alpha - a_2^\alpha$.

The first condition ensures the existence of a solution $X_\alpha = [b_1^\alpha - a_1^\alpha, b_2^\alpha - a_2^\alpha]$ of the equation

$$A_\alpha + X_\alpha = B_\alpha.$$

The second condition guarantees the nested property of the solutions of the above equation for $\alpha \leq \beta$. Now, if a solution X_α exists for every $\alpha \in (0, 1]$ and the second condition is satisfied, then by first decomposition theorem,

$$X = \bigcup_{\alpha \in (0,1]} {}_\alpha X.$$

Let us have the following example as an illustration.

Example 15.3.1. Let A and B be two fuzzy numbers given as below.

$$\begin{aligned} A &= \frac{0.2}{[0,1)} + \frac{0.6}{[1,2)} + \frac{0.8}{[2,3)} + \frac{0.9}{[3,4)} + \frac{1}{4} + \frac{0.5}{(4,5]} + \frac{0.1}{(5,6]} \\ B &= \frac{0.1}{[0,1)} + \frac{0.2}{[1,2)} + \frac{0.6}{[2,3)} + \frac{0.7}{[3,4)} + \frac{0.8}{[4,5)} + \frac{0.9}{[5,6)} + \frac{1}{6} + \frac{0.5}{(6,7]} + \frac{0.4}{(7,8]} + \frac{0.2}{(8,9]} + \frac{0.1}{(9,10]}. \end{aligned}$$

Then the α-cuts of A and B and the solution of the associated equation $A_\alpha + X_\alpha = B_\alpha$ are given by the following.

α	A_α	B_α	X_α
1.0	$[4,4]$	$[6,6]$	$[2,2]$
0.9	$[3,4]$	$[5,6]$	$[2,2]$
0.8	$[2,4]$	$[4,6]$	$[2,2]$
0.7	$[2,4]$	$[3,6]$	$[1,2]$
0.6	$[1,4]$	$[2,6]$	$[1,2]$
0.5	$[1,5]$	$[2,7]$	$[1,2]$
0.4	$[1,5]$	$[2,8]$	$[1,3]$
0.3	$[1,5]$	$[2,8]$	$[1,3]$
0.2	$[0,5]$	$[1,9]$	$[1,4]$
0.1	$[0,6]$	$[0,10]$	$[0,4]$

Thus, by the first decomposition theorem, the solution of the fuzzy number is given by

$$X = \bigcup_{\alpha\in(0,1]} {}_\alpha X = \frac{0.1}{[0,1)} + \frac{0.7}{[1,2)} + \frac{1}{2} + \frac{0.4}{(2,3]} + \frac{0.2}{(3,4]}.$$

15.3.2 Equations of type $A \cdot X = B$

Let A, B be fuzzy numbers on $\mathbb{R}^+$. It is easy to show that $X = B/A$ is not a solution of the equation. For each $\alpha \in (0,1]$, we obtain the interval equation

$$A_\alpha \cdot X_\alpha = B_\alpha.$$

Similar to the previous equation, our fuzzy equation can be solved by solving these interval equations for all $\alpha \in (0,1]$. Let $A_\alpha = [a_1^\alpha, a_2^\alpha]$, $B_\alpha = [b_1^\alpha, b_2^\alpha]$, and $X_\alpha = [x_1^\alpha, x_2^\alpha]$. Then, the solution of the fuzzy equation exists if and only if the following conditions are satisfied.

1. $\dfrac{b_1^\alpha}{a_1^\alpha} \le \dfrac{b_2^\alpha}{a_2^\alpha}$ for each $\alpha \in (0,1]$, and

2. $\alpha \le \beta$ implies $\dfrac{b_1^\alpha}{a_1^\alpha} \le \dfrac{b_1^\beta}{a_1^\beta} \le \dfrac{b_2^\beta}{a_2^\beta} \le \dfrac{b_2^\alpha}{a_2^\alpha}$.

If the solution exists, it has the form

$$X = \bigcup_{\alpha\in(0,1]} X_\alpha.$$

Let us see the following illustration.

Example 15.3.2. Let A and B be fuzzy numbers with the following membership functions.

$$\begin{aligned} \mu_A(x) &= 0, \text{ for } x \le 3 \text{ and } x > 5 \\ &= x-3, \text{ for } 3 < x \le 4 \\ &= 5-x, \text{ for } 4 < x \le 5. \end{aligned}$$

$$\begin{aligned} \mu_B(x) &= 0, \text{ for } x \le 12 \text{ and } x > 32 \\ &= \frac{x-12}{8}, \text{ for } 12 < x \le 20 \\ &= \frac{32-x}{12}, \text{ for } 20 < x \le 32. \end{aligned}$$

Then, $A_\alpha = [\alpha + 3, 5 - \alpha]$ and $B_\alpha = [8\alpha + 12, 32 - 12\alpha]$. We see that

$$\frac{8\alpha + 12}{\alpha + 3} \leq \frac{32 - 12\alpha}{5 - \alpha};$$

which implies that

$$X_\alpha = \left[\frac{8\alpha + 12}{\alpha + 3}, \frac{32 - 12\alpha}{5 - \alpha}\right]$$

for each $\alpha \in (0, 1]$. Also, it can be checked that for $\alpha \leq \beta$, $X_\beta \subseteq X_\alpha$ for each pair $\alpha, \beta \in (0, 1]$. Hence the solution of our fuzzy equation is a fuzzy number $X = \bigcup\limits_{\alpha \in (0,1]} X_\alpha$ having membership function

$$\begin{aligned}
\mu_X(x) &= 0, \ \text{ for } x \leq 4 \text{ and } x \geq \frac{32}{5} \\
&= \frac{12 - 3x}{x - 8}, \ \text{ for } 4 < x \leq 5 \\
&= \frac{32 - 5x}{12 - x}, \ \text{ for } 5 \leq x \leq \frac{32}{5}.
\end{aligned}$$

Exercise 15.3.3. Let A and B be two fuzzy numbers given in exercise 15.2.8 and C be another fuzzy number having membership function

$$\begin{aligned}
\mu_C(x) &= \frac{x - 6}{2}, \ \text{ for } 6 < x \leq 8 \\
&= \frac{10 - x}{2}, \ \text{ for } 8 < x \leq 10 \\
&= 0, \ \text{ otherwise }.
\end{aligned}$$

For these fuzzy numbers, solve the following equations for X.

1. $A + X = B$;

2. $B \cdot X = C$.

Few Probable Questions

1. Define a fuzzy number. Deduce a necessary and sufficient condition for a fuzzy set on $\mathbb{R}$ to be a fuzzy number.

2. Let A and B be two fuzzy numbers whose membership functions are given by

$$\begin{aligned}
\mu_A(x) &= 0, \ \ x < -3 \\
&= \frac{x + 3}{5}, \ \ -3 \leq x \leq 2 \\
&= \frac{4 - x}{2}, \ \ 2 \leq x \leq 4 \\
&= 0, \ \ x > 4
\end{aligned}$$

$$
\begin{aligned}
\mu_B(x) &= 0, \;\; x < -1 \\
&= \frac{x+1}{1}, \;\; -1 \leq x \leq 0 \\
&= \frac{6-x}{6}, \;\; 0 \leq x \leq 6 \\
&= 0, \;\; x > 6.
\end{aligned}
$$

Find $A + B$ and $A \cdot B$.

Unit 16

Course Structure

- Variational Problems with fixed Boundaries: Variation, Linear functional, Euler-Lagrange equation, Functionals dependent on higher order derivatives, Functionals dependent on functions of several variables.

16.1 Introduction

The calculus of variations is concerned with solving Maximum and Minimum problems for functionals (functions whose domain contains functions and range will be the set of real numbers $\mathbb{R}$). This unit contains methods to obtain extremum of a given functional in one variable, several variables and for functional involving higher order derivatives.

Objective

After reading this unit, you will be able to pursue the maximum and minimum of functionals in one variable, several variables and for functional involving higher order derivatives.

Example 16.1.1. Consider the set of all rectifiable plane curves. The length of any curve between two points (x_0, y_0) and (x_1, y_1) on the curve $y = y(x)$ is given by

$$I[y(x)] = \int_{x_0}^{x_1} \left[1 + \left(\frac{dy}{dx}\right)^2\right]^{\frac{1}{2}} dx.$$

This integral defines a functional. Calculus of variations deals with the problem to find the curve $y = y(x)$ such that the definite integral is maximum or minimum.

16.2 Linear Functional

A functional $I[y(x)]$ is said to be linear if

i. $I[y_1(x) + y_2(x)] = I[y_1(x)] + I[y_2(x)]$

ii. $I[cy(x)] = cI[y(x)]$.

16.3 Extremal

A function $y = y(x)$ which extremizes a functional is called extremal.

16.4 Euler's Equation (Necessary condition for existence of extremal)

The necessary condition for functional $I[y(x)] = \int_{x_0}^{x_1} F(x, y, y')\, dx$ to be maximum or minimum is that

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0$$

where F is twice continuously differentiable function and value of $x_0, x_1, y(x_0), y(x_1)$ are prescribed.

Proof. Consider the functional

$$I[y(x)] = \int_{x_0}^{x_1} F(x, y, y')\, dx. \tag{1}$$

Let $y = y(x)$ be extremal of functional and $\bar{y}(x)$ be neighbourhood of $y(x)$ such that

$$\bar{y}(x) = y(x) + \epsilon\eta(x). \tag{2}$$

where ϵ is a small parameter and $\eta(x)$ is an arbitrary function such that $\eta(x_0) = \eta(x_1) = 0$.

From (1) and (2), we obtain

$$\begin{aligned} I[\bar{y}(x)] &= \int_{x_0}^{x_1} F(x, \bar{y}, \bar{y}')\, dx \\ &= \int_{x_0}^{x_1} F(x, y + \epsilon\eta, y' + \epsilon\eta')\, dx \text{ which is a function of } \epsilon\text{, say, } I(\epsilon). \end{aligned}$$

Now,

$$\begin{aligned} I(\epsilon) &= \int_{x_0}^{x_1} F(x, y + \epsilon\eta, y' + \epsilon\eta')\, dx \\ &= \int_{x_0}^{x_1} \left[F(x, y, y') + \epsilon\eta\frac{\partial F}{\partial y} + \epsilon\eta'\frac{\partial F}{\partial y'} + O(\epsilon^2)\right] dx. \text{ [by Taylor's theorem of function of several variables]} \end{aligned}$$

$\therefore \frac{dI}{d\epsilon} = \int_{x_0}^{x_1} \left[\eta\frac{\partial F}{\partial y} + \eta'\frac{\partial F}{\partial y'} + O(\epsilon)\right] dx.$

The necessary condition for existence of extremal is

$$
\begin{aligned}
&\left(\frac{dI}{d\epsilon}\right)_{\epsilon} = 0 \\
\Rightarrow\ & \int_{x_0}^{x_1} \left[\eta\frac{\partial F}{\partial y} + \eta'\frac{\partial F}{\partial y'}\right] dx = 0 \\
\Rightarrow\ & \int_{x_0}^{x_1} \left[\eta\frac{\partial F}{\partial y}\right] dx + \int_{x_0}^{x_1} \left[\eta'\frac{\partial F}{\partial y'}\right] dx = 0 \\
\Rightarrow\ & \int_{x_0}^{x_1} \left[\eta\frac{\partial F}{\partial y}\right] dx + \left[\eta(x)\frac{\partial F}{\partial y'}\right]_{x_0}^{x_1} - \int_{x_0}^{x_1} \left[\eta(x)\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right] dx = 0 \\
\Rightarrow\ & \int_{x_0}^{x_1} \left[\eta\frac{\partial F}{\partial y}\right] dx + 0 - \int_{x_0}^{x_1} \left[\eta(x)\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right] dx = 0 \ [\because \eta(x_0) = \eta(x_1) = 0] \\
\Rightarrow\ & \int_{x_0}^{x_1} \eta(x) \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right] dx = 0.
\end{aligned}
$$

Since $\eta(x)$ is arbitrary, we have

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0$$

which is the required **Euler's equaltion**. □

Remark 16.4.1. This equation is also called as **Euler's -Lagrange's equation**.

Example 16.4.2. Extremize

$$I[y(x)] = \int_0^1 (x\sin y + \cos y)\, dx$$

where $y(0) = 0, y(1) = \frac{\pi}{4}$.

Solution. Given that $F(x, y, y') = x\sin y + \cos y$.

By Euler's equation

$$
\begin{aligned}
&\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0 \\
\Rightarrow\ & x\cos y - \sin y = 0 \\
\Rightarrow\ & \tan y = x \\
\Rightarrow\ & y = \tan^{-1}(x).
\end{aligned}
$$

This is the required extremal which satisfies the boundary conditions $y(0) = \tan^{-1}(0) = 0$, $y(1) = \tan^{-1}(1) = \frac{\pi}{4}$. ■

16.5 Functional Dependent on Higher Derivatives

16.5.1 Euler-Poisson Equation

Consider the functional

$$I[y(x)] = \int_{x_0}^{x_1} F(x, y, y', y'', ..., y^n)\, dx$$

where values of $x_0,\ y(x_0),\ y(x_1),\ y'(x_0),\ y'(x_1), ..., y^{n-1}(x_0),\ y^{n-1}(x_1)$ are prescribed. The necessary condition for existence of extremal of $I[y(x)]$ is

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) + ... + (-1)^n \frac{d^n}{dx^n}\left(\frac{\partial F}{\partial y^n}\right) = 0.$$

Particular Case

1. If $I[y(x)] = \int_{x_0}^{x_1} F(x, y, y', y'')\, dx$, the necessary condition for existence of extremal is

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) = 0.$$

2. If $I[y(x)] = \int_{x_0}^{x_1} F(x, y, y', y'', y''')\, dx$, the necessary condition for existence of extremal is

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) - \frac{d^3}{dx^3}\left(\frac{\partial F}{\partial y'''}\right) = 0.$$

Example 16.5.1. Find the extremals of the functional

$$I[y(x)] = \int_{x_0}^{x_1} [(y'')^2 - 2(y')^2 + y^2]\, dx.$$

Solution. Given that $F = (y'')^2 - 2(y')^2 + y^2$.

The necessary condition for existence of extremal is

$$\begin{aligned}
&\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) = 0\\
\Rightarrow\ & 2y - \frac{d}{dx}\left(-4y'\right) + \frac{d^2}{dx^2}\left(2y''\right) = 0\\
\Rightarrow\ & y + 2\frac{d^2y}{dx^2} + \frac{d^4y}{dx^4} = 0\\
\Rightarrow\ & \frac{d^4y}{dx^4} + 2\frac{d^2y}{dx^2} + y = 0.
\end{aligned}$$

Now the auxiliary equation of the above differential equation is

$$\begin{aligned}
& m^4 + 2m^2 + 1 = 0\\
\Rightarrow\ & (m^2+1)^2 = 0\\
\Rightarrow\ & m = \pm i,\ \pm i.
\end{aligned}$$

Hence the solution is

$$y = (c_1 + c_2 x)\cos x + (c_3 + c_4 x)\sin x$$

where $c_1,\ c_2,\ c_3,\ c_4$ are arbitrary constants.

This is the required extremal. ∎

16.6 Functionals Dependent on Functions of Several Variables

Theorem 16.6.1. A necessary condition for

$$I[y_1(x), y_2(x), ..., y_n(x)] = \int_{x_0}^{x_1} F(x, y_1, y_2, ..., y_n, y_1', y_2', ..., y_n')\, dx$$

to be maximum or minimum is that

$$\frac{\partial F}{\partial y_i} - \frac{d}{dx}\left(\frac{\partial F}{\partial y_i'}\right) = 0;\ i = 1, 2, ..., n.$$

Example 16.6.2. Extremize

$$I[y(x), z(x)] = \int_0^{\frac{\pi}{2}} (y'^2 + z'^2 + 2yz)\, dx$$

with $y(0) = 0, y(\frac{\pi}{2}) = 1, z(0) = 0, z(\frac{\pi}{2}) = -1.$

Solution. Given that $F = y'^2 + z'^2 + 2yz.$

$\therefore$ The necessary condition for existence of extremum is

$$\begin{aligned} & \frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0 \\ \text{and} \qquad & \frac{\partial F}{\partial z} - \frac{d}{dx}\left(\frac{\partial F}{\partial z'}\right) = 0 \\ \text{i.e,}\ \ & 2z - \frac{d}{dx}(2y') = 0 \\ \text{and} \qquad & 2y - \frac{d}{dx}(2z') = 0 \\ \text{i.e,}\ \ & \frac{d^2y}{dx^2} = z \qquad (1) \\ \text{and} \qquad & \frac{d^2z}{dx^2} = y \qquad (2) \end{aligned}$$

From (1) and (2), we obtain that

$$\begin{aligned} & \frac{d^2}{dx^2}\left(\frac{d^2y}{dx^2}\right) = y \\ \Rightarrow\ & \frac{d^4y}{dx^4} = y. \end{aligned}$$

Now the auxiliary equation is

$$\begin{aligned} & m^4 - 1 = 0 \\ \Rightarrow\ & (m^2 - 1)(m^2 + 1) = 0 \\ \Rightarrow\ & m = \pm 1,\ \pm i. \end{aligned}$$

$\therefore$ The solution is

$$y = c_1 e^x + c_2 e^{-x} + c_3 \cos x + c_4 \sin x. \qquad (3)$$

Now

$$\frac{dy}{dx} = c_1e^x - c_2e^{-x} - c_3 \sin x + c_4 \cos x$$
$$\frac{d^2y}{dx^2} = c_1e^x + c_2e^{-x} - c_3 \cos x - c_4 \sin x.$$

Hence from (1)

$$z = \frac{d^2y}{dx^2} = c_1e^x + c_2e^{-x} - c_3 \cos x - c_4 \sin x. \tag{4}$$

Using boundary conditions, it follows from (3) and (4) that

$$y(0) = c_1 + c_2 + c_3 + c_4 = 0 \tag{5}$$
$$y\left(\frac{\pi}{2}\right) = c_1e^{\frac{\pi}{2}} + c_2e^{-\frac{\pi}{2}} + c_4 = 1 \tag{6}$$

and

$$z(0) = c_1 + c_2 - c_3 = 0 \tag{7}$$
$$z\left(\frac{\pi}{2}\right) = c_1e^{\frac{\pi}{2}} + c_2e^{-\frac{\pi}{2}} - c_4 = -1. \tag{8}$$

Solving (5), (6), (7) and (8), we get

$$c_1 = c_2 = c_3 = 0 \text{ and } c_4 = 1.$$

Hence the required extremals are

$$y = \sin x,\ z = -\sin x.$$

■

16.7 Exercise

1. Extremize the functional
$$I[y(x)] = \int_0^{2\pi} (y'^2 - y^2)\, dx$$
with $y(0) = 1, y(2\pi) = 1$.

2. Show that the variational problem of extremizing the functional
$$I[y(x)] = \int_1^3 y(3x - y)\, dx$$
with $y(1) = 1, y(3) = 4\frac{1}{2}$ has no solution.

3. Find the extremals of the functional
$$I[y(x)] = \int_0^1 (y''^2 + 1)\, dx$$
with $y(0) = 0, y'(0) = 1, y(1) = 1, y'(1) = 1$.

4. Find the extremals of the functional
$$I[y(x)] = \int_{x_0}^{x_1} (2xy + y'''^2)\, dx.$$

5. Extremize
$$\int_0^{\frac{\pi}{2}} \left[2xy + \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 \right] \, dt$$
with $x(0) = 0, x(\frac{\pi}{2}) = -1$ and $y(0) = 0, y(\frac{\pi}{2}) = 1$.

Unit 17

Course Structure

- Applications of Calculus of variations on the problems of shortest distance, minimum surface of revolution, Brachistochrone problem, geodesic, etc. Isoperimetric problem.

17.1 Introduction

In this unit various problems like shortest distance, minimum surface of revolution, Brachistochrone problem, geodesic and isoperimetric problem will be solved with the help of calculus of variations.

Objectives

After reading this unit, you will be able to solve

- Shortest distance problem.
- Minimum surface of revolution problem.
- Brachistochrone problem.
- Geodesic problem.
- Isoperimetric problem.

17.2 Shortest distance problem

If two points $A = (x_0, y_0), B = (x_1, y_1)$ in the plane are joined by a curve $y = f(x)$, then the Length Functional is given by $L[y(x)] = \int_{x_0}^{x_1} \sqrt{1 + y'^2}\, dx$. The minimum problem for $L[y(x)]$ is the shortest distance problem.

Example 17.2.1. Find the curve of least length joining two points in a plane.

Solution. Consider the functional
$$I[y(x)] = \int_{x_0}^{x_1} \sqrt{1+y'^2}\,dx.$$

Let $F = \sqrt{1+y'^2}$.

$\therefore$ By Euler's equation
$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0$$
$$\Rightarrow \frac{d}{dx}\left(\frac{y'}{\sqrt{1+y'^2}}\right) = 0$$

Integrating, we have
$$\frac{y'}{\sqrt{1+y'^2}} = \text{ constant } = c_1$$
$$\Rightarrow\ y'^2 = c_1^2(1+y'^2)$$
$$\Rightarrow\ y'^2 = \frac{c_1^2}{1-c_1^2}$$
$$\Rightarrow\ y' = \frac{c_1}{\sqrt{1-c_1^2}} = c = \text{ constant.}$$

Again, we obtain by integrating
$$y = cx + d.$$

This is required least length curve which is a straight line. ■

17.3 Brachistochrone problem

If a smooth body is allowed to slide down a smooth curve from one point to another under gravity, then the Brachistochrone problem is to find the curve along which the time taken will be the least.

Example 17.3.1. Find the curve on which a particle will slide from one point to another point in the shortest time under gravity (Friction and resistance of media are ignored).

Solution. Let a particle slide with zero velocity from the origin O on the curve OA and $P(x, y)$ be the position of the particle at any time t. Then velocity v at $P(x, y)$ is given by

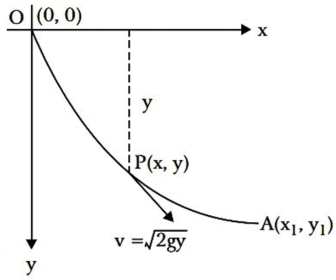


Figure 17.3.1

$$
\begin{aligned}
v^2 &= u^2 + 2gh \\
&= 0 + 2gy \; [\because u = 0, h = y] \\
\Rightarrow v &= \sqrt{2gy}.
\end{aligned}
$$

$\therefore$ Time taken by the particle from O to A is

$$
\begin{aligned}
T &= \int_0^{x_1} \frac{ds}{v} \\
&= \int_0^{x_1} \frac{\sqrt{1+y'^2}}{\sqrt{2gy}}\, dx.
\end{aligned}
$$

Here $F = F(y, y') = \frac{\sqrt{1+y'^2}}{\sqrt{2gy}}$.

Now

$$
\begin{aligned}
\frac{dF}{dx} &= \frac{\partial F}{dy}\frac{dy}{dx} + \frac{\partial F}{dy'}\frac{dy'}{dx} \\
\Rightarrow \; \frac{d}{dx}(F) &= \left(y'\frac{\partial}{dy} + y''\frac{\partial}{dy'} \right) F \\
\Rightarrow \; \frac{d}{dx} &= \left(y'\frac{\partial}{dy} + y''\frac{\partial}{dy'} \right).
\end{aligned}
$$

$\therefore$ Euler's equation becomes

$$
\begin{aligned}
&\frac{\partial F}{dy} - \frac{d}{dx}\left(\frac{\partial F}{dy'}\right) = 0 \\
&\text{or, } F_y - \left(y'\frac{\partial}{dy} + y''\frac{\partial}{dy'} \right) F_{y'} = 0 \\
&\text{or, } F_y - y'F_{yy'} - y''F_{y'y'} = 0 \\
&\text{or, } F_y - \left(y'\frac{\partial}{dy} + y''\frac{\partial}{dy'} \right) F_{y'} = 0 \\
&\text{or, } y'F_y - y'^2F_{yy'} - y'y''F_{y'y'} = 0 \; [\text{ Multiplyinh both sides by } y'] \\
&\text{or, } \frac{d}{dx}(F - y'F_{y'}) = 0.
\end{aligned}
$$

Integrating, we have

$$
F - y'F_{y'} = \text{ constant.}
$$

Thus the necessary condition for existence of extremal is

$$
\begin{aligned}
& F - y'F_{y'} = \text{ constant} \\
\Rightarrow \; & \frac{\sqrt{1+y'^2}}{\sqrt{2gy}} - y'\frac{y'}{\sqrt{1+y'^2}\sqrt{2gy}} = \text{ constant} \\
\Rightarrow \; & \frac{\sqrt{1+y'^2}}{\sqrt{y}} - \frac{y'^2}{\sqrt{1+y'^2}\sqrt{y}} = \text{ constant} \\
\Rightarrow \; & \frac{1}{\sqrt{1+y'^2}\sqrt{y}} = \text{ constant.} \qquad (1)
\end{aligned}
$$

Putting $y' = \cot t$, we obtain from (1)

$$\begin{aligned}
&\frac{1}{\text{cosec } t\sqrt{y}} = c_1 \\
\Rightarrow\ & y = c\sin t \; [c = \frac{1}{c_1^2} = \text{ constant}] \\
\Rightarrow\ & y = \frac{c}{2}(1 - \cos 2t).
\end{aligned} \tag{2}$$

Now

$$\begin{aligned}
\frac{dx}{dt} &= \frac{dx}{dy}\frac{dy}{dt} \\
&= \tan t . \frac{c}{2} 2\sin 2t \\
&= 2c\sin^2 t \\
&= c((1 - \cos 2t)).
\end{aligned}$$

Integrating, we get

$$x = c\left(t - \frac{\sin 2t}{2}\right) = \frac{c}{2}\left(2t - \sin 2t\right). \tag{3}$$

Thus, we get the parametric equation of the required curve from (2) and (3) as

$$\begin{aligned}
x &= \frac{c}{2}\left(2t - \sin 2t\right) \\
y &= \frac{c}{2}(1 - \cos 2t) \\
\text{i.e, } x &= a\left(\theta - \sin\theta\right) \\
y &= a(1 - \cos\theta) \; [\text{Taking } a = \frac{c}{2}, 2t = \theta]
\end{aligned}$$

which is cycloid. ∎

17.4 Minimum Surface of Revolution

In this problem, we have to find a curve $y = y(x)$ passing through two given points (x_0, y_0) and (x_1, y_1) which when rotated about the x -axis gives a minimum surface area. Mathematically, the surface area bounded by such curve is given by

$$S = \int_{x_0}^{x_1} 2\pi y\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

Example 17.4.1. Show that the area of the surface of revolution of a curve $y = y(x)$ is minimum when the curve is catenary.

Solution. The area of the surface of revolution of a curve $y = y(x)$ joining the two points (x_0, y_0) and (x_1, y_1) is

$$S[y(x)] = \int_{x_0}^{x_1} 2\pi y\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

Here $F = F(y, y') = 2\pi y\sqrt{1 + y'^2}$.

Therefore the necessary condition for existence of extremal is

$$\begin{aligned} F - y'F_{y'} &= \text{ constant} \\ \Rightarrow\ 2\pi y\sqrt{1+y'^2} - y'.2\pi y\frac{y'}{\sqrt{1+y'^2}} &= \text{ constant} \\ \Rightarrow\ y\sqrt{1+y'^2} - \frac{yy'^2}{\sqrt{1+y'^2}} &= \text{ constant} \\ \Rightarrow\ \frac{y}{\sqrt{1+y'^2}} &= \text{ constant.} \qquad (1) \end{aligned}$$

Taking $y' = \sinh t$, (1) becomes

$$\begin{aligned} \frac{y}{\cosh t} &= \text{constant} = c(say) \\ \text{or, } y &= c\cosh t. \qquad (2) \end{aligned}$$

Now

$$\begin{aligned} \frac{dx}{dt} &= \frac{dx}{dy}\frac{dy}{dt} \\ &= \text{cosech } t.c\sinh t \\ &= c. \end{aligned}$$

Integrating, we have

$$\begin{aligned} x &= ct + d \\ \Rightarrow\ t &= \frac{x-d}{c}. \qquad (3) \end{aligned}$$

$\therefore$ From (2) and (3), we get

$$y = c\cos\left(\frac{x-d}{c}\right)$$

which is the equation of catenary.

This is the required extremal. ■

17.5 Geodesics

Geodesics on a surface is a curve along which the distance between two points is minimum.

Example 17.5.1. Find the Geodesics on right circular cylinder.

Solution. The element of arc on right circular cylinder $r = a$ (radius of cylinder) is

$$\begin{aligned}
ds^2 &= (dr)^2 + (rd\theta)^2 + (dz)^2 \\
\text{or, } ds^2 &= 0 + (ad\theta)^2 + (dz)^2 \\
\text{or, } ds^2 &= \left[a^2 + \left(\frac{dz}{d\theta}\right)^2\right](d\theta)^2 \\
\text{or, } ds &= \sqrt{1+z'^2}\, d\theta \text{ where } z' = \frac{dz}{d\theta}.
\end{aligned}$$

$\therefore$ Length of any line on the cylinder between P_1 and P_2 is

$$\begin{aligned}
L &= \int_{P_1}^{P_2} ds \\
&= \int_{P_1}^{P_2} \sqrt{1+z'^2}\, d\theta.
\end{aligned}$$

Here $F = F(\theta, z, z') = \sqrt{1+z'^2}$.

$\therefore$ Euler's equation becomes

$$\begin{aligned}
\frac{\partial F}{dz} - \frac{d}{d\theta}\left(\frac{\partial F}{dz'}\right) &= 0 \\
\text{or, } 0 - \frac{d}{d\theta}\left(\frac{z'}{\sqrt{1+z'^2}}\right) &= 0 \\
\text{or, } \frac{d}{d\theta}\left(\frac{z'}{\sqrt{1+z'^2}}\right) &= 0.
\end{aligned}$$

Integrating, we get

$$\begin{aligned}
&\frac{z'}{\sqrt{1+z'^2}} = c \\
\Rightarrow\ & z' = \frac{c}{\sqrt{1-c^2}} = \text{constant } = k \text{ (say)} \\
\Rightarrow\ & \frac{dz}{d\theta} = k.
\end{aligned}$$

Again integrating, we have

$$z = k\theta + b$$

which is circular helix.

Hence the required extremal is circular helix. ∎

17.6 Isoperimetric Prpblem

In this problem we find a curve of given perimeter what will enclose the maximum area. This type of problem is solved by Lagrange's multiplier's method.

To find the extremals of the functional

$$I[y(x)] = \int_{x_0}^{x_1} f(x, y, y')\, dx \tag{1}$$

subject to the condition (constraint)

$$J[y(x)] = \int_{x_0}^{x_1} g(x, y, y')\, dx \tag{2}$$

we have to consider $F = f + \lambda g$ where λ is known as Lagrange multiplier.

Then by Euler equation, the necessary condition for existence of extremals is

$$\frac{\partial F}{dy} - \frac{d}{dx}\left(\frac{\partial F}{dy'}\right) = 0.$$

This gives required extremals of (1) under the condition (2).

Example 17.6.1. Find the extremal of the functional $\int_0^2 y'^2\, dx$ under the constraint $\int_0^2 y\, dx = 1$ given $y(0) = 0$ and $y(2) = 1$.

Solution. Let $I = \int_0^2 y'^2\, dx$ and $J = \int_0^2 y\, dx = 1$.

Here $f = y'^2,\ g = y$.

Consider $F = f + \lambda g = y'^2 + \lambda y$.

Then we obtain by Euler's equation that

$$\begin{aligned} &\frac{\partial F}{dy} - \frac{d}{dx}\left(\frac{\partial F}{dy'}\right) = 0 \\ \Rightarrow\ &\lambda - \frac{d}{dx}(2y') = 0 \\ \Rightarrow\ &\frac{d^2y}{dx^2} = \frac{\lambda}{2}. \end{aligned}$$

Integrating, we have

$$y = \frac{\lambda}{4}x^2 + c_1x + c_2 \tag{1}$$

where c_1 and c_2 are arbitrary constants.

Now $y(0) = 0$ and $y(2) = 1$ gives $c_1 = \frac{1-\lambda}{2},\ c_2 = 0$.

Hence (1) beomes

$$y = \frac{\lambda}{4}x^2 + \frac{1-\lambda}{2}x. \tag{2}$$

Now $\int_0^2 y\,dx = 1$ gives

$$\begin{aligned}
&\int_0^2 \left(\frac{\lambda}{4}x^2 + \frac{1-\lambda}{2}x\right)\,dx = 1\\
\Rightarrow\ &\left[\frac{\lambda}{4}.\frac{x^3}{3} + \frac{1-\lambda}{2}.\frac{x^2}{2}\right]_0^2 = 1\\
\Rightarrow\ &\frac{8\lambda}{12} + 1 - \lambda = 1\\
\Rightarrow\ &\lambda = 0.
\end{aligned}$$

Hence from (2), we get

$$y = \frac{1}{2}x$$

which is the reqiured extremal. ■

17.7 Exercise

1. Find the curve along which the time taken is the least when velocity at any point of it is $v = x$.
2. Find the Geodesics on right circular cone.
3. Find the extremal of the functional
$$I = \int_0^\pi (y'^2 - y^2)\,dx$$
under the condition $y(0) = 0, y(\pi) = 1$ and sybject to constraint
$$\int_0^\pi y\,dx = 1.$$
4. Find the shape of the curve of the given perimeter enclosing maximum area.

Unit 18

Course Structure

- Variational Problems with Moving Boundaries: Transversality conditions, Orthogonality conditions, Functional dependent on two functions, One sided variations.

18.1 Introduction

In this unit we consider functionals with one or both moving boundary points along the curve. Such problem is known as variational problem with moving or free boundaries.

Objective

After reading this unit, you will be able to solve variational problems with moving boundaries.

18.2 Transversality Conditions

Consider the functional

$$I[y(x)] = \int_{x_0}^{x_1} F(x, y, y')\, dx. \tag{1}$$

If boundary point (x_0, y_0) moves along the curve $y = \phi(x)$ and boundary point (x_1, y_1) moves along the curve $y = \psi(x)$, then transversality condition is

$$[F + (\phi' - y')F_{y'}]_{x=x_0} = 0 \tag{2}$$

$$[F + (\psi' - y')F_{y'}]_{x=x_1} = 0. \tag{3}$$

This gives the required extremal of the functional (1).

Note 18.2.1. If boundary point (x_0, y_0) is fixed and boundary point (x_1, y_1) moves along the curve $y = \psi(x)$, then transversality condition is

$$[F + (\psi' - y')F_{y'}]_{x=x_1} = 0.$$

18.3 Orthogonality Conditions

If F in (1) is given by

$$F = A(x, y)(1 + y'^2)^{\frac{1}{2}}$$

where $A(x, y)$ does not vanish at the movable point point x_1, then (3) reduces to

$$\begin{aligned}
& A(x, y).\frac{1 + \psi' y'}{\sqrt{1 + y'^2}} = 0 \text{ at } x = x_1 \\
\Rightarrow\ & \frac{1 + \psi' y'}{\sqrt{1 + y'^2}} = 0 \text{ at } x = x_1 \ [\because A(x, y) \neq 0 \text{ at } x = x_1] \\
\Rightarrow\ & y' = \frac{1}{\psi'} \text{ at } x = x_1 \\
\Rightarrow\ & \psi' y' = -1 \text{ at } x = x_1
\end{aligned}$$

which is the orthogonality condition.

Example 18.3.1. Find the minimum distance between circle $x^2 + y^2 = 1$ and straight line $x + y = 4$.

Solution. We have to extremize functional

$$I[y(x)] = \int_{x_0}^{x_1} \sqrt{1 + y'^2}\, dx \tag{1}$$

subject to condition that end points (x_0, y_0) and (x_1, y_1) varies along the given circle and straight line respectively.

Here $F = \sqrt{1 + y'^2}$.

$\therefore$ By Euler's equation, we have

$$\begin{aligned}
& \frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0 \\
\Rightarrow\ & \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + y'^2}}\right) = 0
\end{aligned}$$

Integrating, we get

$$\begin{aligned}
& \frac{y'}{\sqrt{1 + y'^2}} = \text{ constant} = c_1 \\
\Rightarrow\ & y'^2 = c_1^2(1 + y'^2) \\
\Rightarrow\ & y'^2 = \frac{c_1^2}{1 - c_1^2} \\
\Rightarrow\ & y' = \frac{c_1}{\sqrt{1 - c_1^2}} = c = \text{ constant.}
\end{aligned}$$

Again, we obtain by integrating

$$y = cx + d \tag{2}$$

which is straight line along which the required shortest distance is attained.

Now,

$$x^2+y^2=1 \Rightarrow y=\sqrt{1-x^2}$$

and

$$x+y=4 \Rightarrow y=4-x.$$

$\therefore$ Let $\phi(x)=\sqrt{1-x^2}$ and $\psi(x)=4-x$.

Hence by transversality condition

$$[F+(\phi'-y')F_{y'}]_{x=x_0}=0$$

and

$$[F+(\psi'-y')F_{y'}]_{x=x_1}=0$$

$$\Rightarrow \left[\sqrt{1+y'^2}+\left(-\frac{x}{\sqrt{1-x^2}}-y'\right)\frac{y'}{\sqrt{1+y'^2}}\right]_{x=x_0}=0$$

and

$$\left[\sqrt{1+y'^2}+\left(-1-y'\right)\frac{y'}{\sqrt{1+y'^2}}\right]_{x=x_1}=0$$

$$\Rightarrow \qquad \sqrt{1+c'^2}+\left(-\frac{x_0}{\sqrt{1-x_0^2}}-c\right)\frac{c}{\sqrt{1+c^2}}=0$$

and

$$\sqrt{1+c^2}+(-1-c)\frac{c}{\sqrt{1+c^2}}=0$$

$$\Rightarrow 1-\frac{cx_0}{\sqrt{1-x_0^2}}=0$$

and

$$1-c=0$$

$$\Rightarrow c=1 \text{ and } \frac{x_0}{\sqrt{1-x_0^2}}=1$$

$$\Rightarrow c=1 \text{ and } x_0=\frac{1}{\sqrt{2}}.$$

Since both points (x_0,y_0) and (x_1,y_1) lies on the extremal (2), we get

$$y_0=cx_0+d$$

$$\Rightarrow cx_0+d=\sqrt{1-x_0^2}\ [\because (x_0,y_0) \text{ varies on the curve } y=\phi(x)]$$

$$\Rightarrow 1.\frac{1}{\sqrt{2}}+d=\sqrt{1-\frac{1}{2}}$$

$$\Rightarrow d=0$$

amd

$$y_1=cx_1+d$$

$$\Rightarrow cx_1+d=4-x_1\ [\because (x_1,y_1) \text{ varies on the curve } y=\psi(x)]$$

$$\Rightarrow x_1=2.$$

Thus we obtain

$$\begin{aligned} & c = 1, d = 0 \\ \text{amd} \quad & (x_0, y_0) = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \\ & (x_1, y_1) = (2, 2). \end{aligned}$$

Hence the required extremal from (2) is

$$y = x$$

$$\begin{aligned} \text{and required shortest distance} \quad & = \int_{\frac{1}{\sqrt{2}}}^{2} \sqrt{1+1}\, dx \\ & = 2\sqrt{2} - 1. \end{aligned}$$

■

18.4 Functional dependent on two functions

Let us consider the functional

$$I[y(x), z(x)] = \int_{x_0}^{x_1} F(x, y(x), z(x), y'(x), z'(x))\, dx \tag{1}$$

with the lower point (x_0, y_0, z_0) be fixed and upper point (x_1, y_1, z_1) move in an arbitrary manner or along a given curve or surface.

If the boundary point (x_1, y_1, z_1) moves along some curve $y_1 = \phi(x_1), z_1 = \psi(x_1)$, then the transversality condition

$$[F + (\phi' - y')F_{y'} + (\psi' - z')F_{z'}]_{x=x_1} = 0$$

together with $y_1 = \phi(x_1), z_1 = \psi(x_1)$ gives the necessary equations for determining the two arbitrary constants in the general solution of Euler's equation.

Note 18.4.1. If the boundary point (x_1, y_1, z_1) moves along a given surface $z_1 = \phi(x_1, y_1)$, then the two equations

$$\begin{aligned} & [F - y'F_{y'} + (\phi_x - z')F_{z'}]_{x=x_1} = 0 \\ \text{and} \quad & [F_{y'} + \phi_y F_{z'}]_{x=x_1} = 0 \end{aligned}$$

with $z_1 = \phi(x_1, y_1)$ enable to determine two arbitrary constants in the general solution of Euler's equation.

Example 18.4.2. Find the extremum of the functional

$$I = \int_{x_0}^{x_1} (y'^2 + z'^2 + 2yz)\, dx$$

with $y(0) = 0, z(0) = 0$ and the point (x_1, y_1, z_1) moves over the fixed plane $x = x_1$.

Solution. Given $F = y'^2 + z'^2 + 2yz$.

By Euler's equation, we have

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0$$
$$\frac{\partial F}{\partial z} - \frac{d}{dx}\left(\frac{\partial F}{\partial z'}\right) = 0$$
$$\Rightarrow 2z - \frac{d}{dx}\left(2y'\right) = 0$$
$$\text{and} \qquad 2y - \frac{d}{dx}\left(2z'\right) = 0$$
$$\Rightarrow \frac{d^2y}{dx^2} = z \tag{1}$$
$$\text{and} \qquad \frac{d^2z}{dx^2} = y. \tag{2}$$

Now from (1) and (2), we get

$$\frac{d^4y}{dx^4} = y.$$

The auxiliary equation is

$$m^4 - 1 = 0$$
$$\Rightarrow \ (m^2-1)(m^2+1) = 0$$
$$\Rightarrow \ m = \pm 1, \pm i.$$

Hence the solution is

$$y = c_1 \cosh x + c_2 \sinh x + c_3 \cos x + c_4 \sin x. \tag{3}$$

$\therefore$ From (1), we obtain

$$z = \frac{d^2y}{dx^2} = c_1 \cosh x + c_2 \sinh x - c_3 \cos x - c_4 \sin x. \tag{4}$$

Now $y(0) = 0, z(0) = 0$ gives $c_1 = c_3 = 0$

Since x_1 is fixed, it follows by condition of moving boundary point (x_1, y_1, z_1) that

$$[F_{y'}]_{x=x_1} = 0$$
$$[F_{z'}]_{x=x_1} = 0$$
$$\Rightarrow \ y'(x_1) = 0$$
$$z'(x_1) = 0$$
$$\Rightarrow \ c_2 \cosh x_1 + c_4 \cos x_1 = 0 \text{ [By (3)]}$$
$$c_2 \cosh x_1 - c_4 \cos x_1 = 0 \text{ [By (4)]}$$

If $\cosh x_1 \neq= 0$, then $c_2 = c_4 = 0$ and therefore extremum is obtained by $y = 0, z = 0$.

But if $\cos x_1 = 0$, then $c_2 = 0$ and c_4 remains arbitrary and hence extremum is

$$y = c_4 \sin x$$
$$z = -c_4 \sin x.$$

■

18.5 One Sided Variations

Consider the functional

$$I[y(x)] = \int_{x_0}^{x_1} F(x, y, y')\, dx. \tag{1}$$

We have already discussed the case that the extremal curve passes through end points (x_0, y_0) and (x_1, y_1).

In this section suppose a restriction is imposed on the classs of permissible curves in such a way that the curve cannot pass through the point of certain region R bounded by the curve $\psi(x, y) = 0$.

In such a problem that extremizing curve C either passes through a region which is completely outside R or C consists of arcs lying outside R and also consists of parts of the boundary of the region R.

Example 18.5.1. Find the shortest path from the point $A(-2, 3)$ to the point $B(2, 3)$ located in the region $y \le x^2$.

Solution. We have to find the extremum of the functional

$$I[y] = \int_{-2}^{2} \sqrt{1 + y'^2}\, dx \tag{1}$$

subject to condition that $y \le x^2, y(-2) = 3, y(2) = 3$. Here $F = \sqrt{1 + y'^2}$.

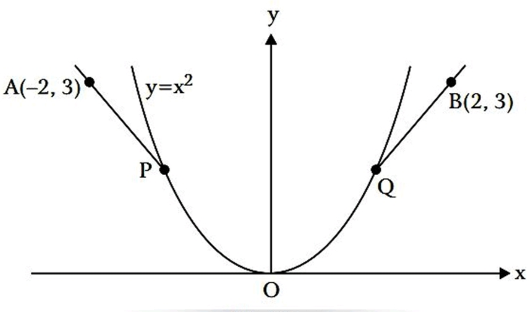


Figure 18.5.1

$\therefore$ By Euler's equation

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0$$
$$\Rightarrow \frac{d}{dx}\left(\frac{y'}{\sqrt{1 + y'^2}}\right) = 0$$

Integrating, we have

$$\begin{aligned}
\frac{y'}{\sqrt{1+y'^2}} &= \text{ constant} = c_1 \\
\Rightarrow\ y'^2 &= c_1^2(1+y'^2) \\
\Rightarrow\ y'^2 &= \frac{c_1^2}{1-c_1^2} \\
\Rightarrow\ y' &= \frac{c_1}{\sqrt{1-c_1^2}} = c = \text{ constant.}
\end{aligned}$$

Integrating, we get

$$y = cx + d \tag{2}$$

This is the required extremal curve where c and d are arbitrary constamts.

Since $F_{y'y'} = [1+y'^2]^{-\frac{3}{2}} \neq 0$, the required extremal will consist of portion of the straight line AP and QB both tangent to the parabola $y = x^2$ and the portion POQ at the parabola.

Let $-\bar{x}$ and $\bar{x}$ be the abscissae of P and Q respectively. Then the condition of tangent of AP and BQ at P and Q are

$$d + c\bar{x} = \bar{x}^2 \tag{3}$$

$$c = 2\bar{x}. \tag{4}$$

Since tangent QB passes through $(2, 3)$, we obtain from (2) that

$$d + 2c = 3. \tag{5}$$

Solving $(3), (4)$ and (5), we get two values of $\bar{x}$ as $\bar{x} = 1$ and $\bar{x} = 3$. The second value is clearly not possible.

$\therefore \qquad \bar{x} = 1.$

Hence from (3) and (4), we have $c = 2, d = -1$.

Hence the required extremal is

$$y = \begin{cases} -2x-1, & \text{if} \quad -2 \le x \le -1 \\ x^2, & \text{if} \quad -1 \le x \le 1 \\ 2x-1, & \text{if} \quad 1 \le x \le 2 \end{cases}$$

which minimize the functional. ■

18.6 Exercise

1. Find the shortest distance between the parabola $y = x^2$ and the straight line $x - y = 5$.
2. Find the shortest distance between the point $(1, 0)$ and the ellipse $4x^2 + 9y^2 = 36$.
3. Find the shortest distance between the point $(-1, 5)$ and the parabola $y^2 = x$.

Unit 19

Course Structure

- Sufficient Conditions for an Extremum: Proper field, Central field, Field of extremals, Embedding in a field of extremals and in a central field.

19.1 Introduction

The aim of this unit is to discuss about some field of extremals and embedding extremals in a field of extremals.

Objective

After reading this unit, you will be able to identify proper field, central field, field of extremals and to solve problems related with embedding extremals in a field of extremals.

19.2 Proper Field

A family of curves $y = y(x, c)$ where c is a parameter is said to form a proper field in a given region D of the xy plane if one and only one curve of the family passes through any point of the region D.

Examples:

1. Family of parallel lines $y = x + c$ (c being an arbitrary constant) inside the circle $x^2 + y^2 = 1$ forms a proper field because through any point at given circle passes only one straight line of the family (Figure 4.2.1).

2. The family of parabola $y = (x + a)^2$ inside the circle $x^2 + y^2 = 1$ does not form a proper field because parabola of this family intersect inside the circle (Figure 4.2.2).

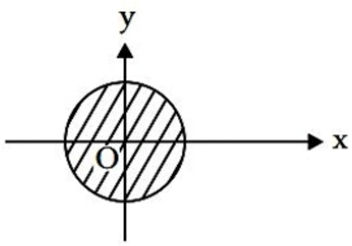


Figure 19.2.1

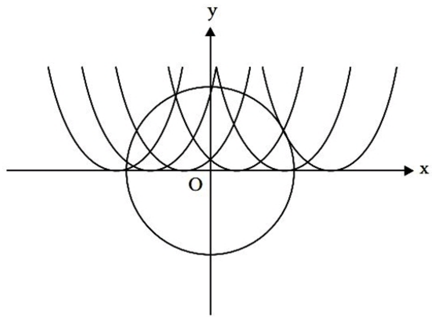


Figure 19.2.2

19.3 Central Field

If all curves of a family $y = y(x, c)$ passes through a single point (x_0, y_0) then such family is said to form a central field over the domain D if these curves cover D without self intersection and the point (x_0, y_0) lies outside D.

Note 19.3.1. If family of curves $y = y(x, c)$ passes through a single point (x_0, y_0) which is not in domain D, then the point (x_0, y_0) is called the centre at pencil of curves.

Example:

Let D be the domain $x > 0$, then $(0, 0) \notin D$. Since family of straight lines $y = cx$ passes through $(0, 0)$, it is the centre at pencil of straight lines. Hence family of straight lines $y = cx$ forms a central field in the domain $x > 0$.

19.4 Field of Extremal

If a proper field or a central field is formed by a family of extremals of given variational problem, then it is called field of extremals.

19.5 Embedding in a Field of Extremals

Let $y = y(x)$ be extremal of functional

$$I[y(x)] = \int_{x_0}^{x_1} F(x, y, y')\, dx$$

with (x_0, y_0) and (x_0, y_0) as two boundary points.

If it is possible to find a family of extremals $y = y(x, c)$ in such a way that the family forms a field and $y = y(x)$ is a member of this family for some value of c and extremal does not lie on the boundary of domain in which family forms a field then $y = y(x)$ is said to be embedded in a extremal field.

19.6 Embedding in a Central Field

If a pencil at extremals originating from the point (x_1, y_1) form a central field which includes the extremal $y = y(x)$, then extremal curve $y = y(x)$ is said to be embedded in central field.

Example 19.6.1. Show that the extremal of the variational problem

$$\int_0^2 (y'^2 + x^2)\, dx$$

with $y(0) = 1, y(2) = 3$ is included in a proper field of extremal of the given functional.

Solution. Here $F = (y'^2 + x^2)$.

$\therefore$ By Euler's equation

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0$$
$$\Rightarrow 0 - \frac{d}{dx}(2y') = 0$$
$$\Rightarrow \frac{d^2 y}{dx^2} = 0.$$

Integrating twice, we have

$$y = cx + d. \tag{1}$$

Using $y(0) = 1, y(2) = 3$, we obtain from (1) that $c = 1, d = 1$.

Hence required extremal is

$$y = x + 1.$$

Now, equation (1) becomes for $c = 1$ as

$$y = x + d$$

which is proper field of extremal in the domain $0 \leq x \leq 2$.

Again for $d = 1$, (1) gives $y = x + 1$ which shows that the extremal $y = x + 1$ is included in the proper field of extremals $y = x + d$. ■

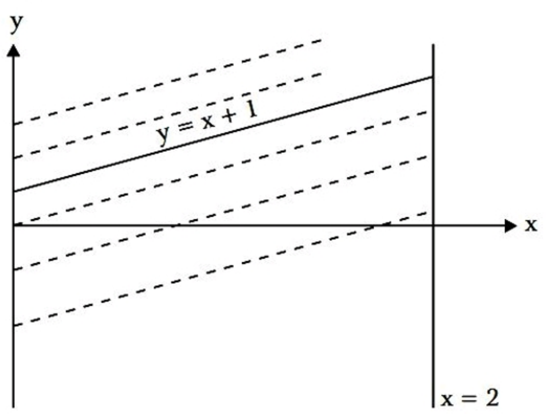


Figure 19.6.1

19.7 Exercise

1. Find the proper and central fields of extremals for the functional
$$\int_0^{\frac{\pi}{4}} (y'^2 - y^2 + 2x^2 + 4).$$

2. Discuss the extremal field for the functional
$$I[y(x)] = \int_0^a (y'^2 - y^2)\, dx$$
with $y(0) = 0, y(a) = 0$.

Unit 20

Course Structure

- Sufficient condition for extremum-Weierstrass condition, Legendre condition. Weak and strong extremum.

20.1 Introduction

This unit contains sufficient condition for extremum namely Legendre condition. Also, weak and strong extremum are discussed here.

Objective

After reading this unit, you will be able to invesitigate extremum for a functional.

20.2 Sufficient Condition for Extremum (Legendre Condition)

Consider the functional

$$I[y(x)] = \int_{x_0}^{x_1} F(x, y, y')\, dx \tag{1}$$

$$\text{with} \qquad y(x_0) = y_0, y(x_1) = y_1.$$

Let C be the extremal curve of the functional (1) and $p = \frac{dy}{dx}$ on C.

If $E(x, y, p, y') = F(x, y, y') - F(x, y, p) - (y' - p)F_p(x, y, p)$ (known as Weirstrass function), then

the extremal is minimum if $E \leq 0$
and the extremal is maximum if $E \geq 0$.

This is the required Legendre condition.

Example 20.2.1. Find Weirstrass function and test the extremal of the functional

$$I[y(x)] = \int_0^a y'^2 \, dx$$

with $y(0)0, y(a) = b$ where $a > 0, b > 0$.

Solution. Here $F = y'^2$.

$\therefore$ By Euler's equation

$$\begin{aligned}
&\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0 \\
&\Rightarrow 0 - \frac{d}{dx}(2y') = 0 \\
&\Rightarrow \frac{d^2y}{dx^2} = 0.
\end{aligned}$$

Integrating twice, we have

$$y = cx + d.$$

Now, $y(0) = 0, y(a) = b$ gives $c = \frac{b}{a}, d = 0$.

Hence required extremal is

$$y = \frac{b}{a}x.$$

Weirstrass Function

The Weirstrass function is

$$\begin{aligned}
E(x, y, p, y') &= F(x, y, y') - F(x, y, p) - (y' - p)F_p(x, y, p) \\
&= y'^2 - p^2 - (y' - p).2p \\
&= (y' - p)^2.
\end{aligned}$$

Since $E(x, y, p, y') = (y' - p)^2 \geq 0$, the extremal is maxima. ■

20.3 Weak and Strong Extremum

Consider the functional

$$I[y(x)] = \int_{x_0}^{x_1} F(x, y, y') \, dx$$
$$\text{with} \quad y(x_0) = y_0, y(x_1) = y_1.$$

Let C be the extremal of the given functional and C is included in a field of extremals.

Then Legendre condition for weak and strong extremum are :

20.3.1 Weak Extremum

1. The curve C is extremal satisfying boundary condition.
2. The extremal C must be embedded in the field of extremals.
3. The Weirstrass function E does not change sign at any point (x, y) close to the curve C and for arbitrary values of y' close to $p(x, y)$ on the extremals.
4. For weak minimum $E \geq 0,\ F_{y'y'} > 0$ on C

   and for weak maximum $E \leq 0,\ F_{y'y'} < 0$ on C.

20.3.2 Strong Extremum

1. The curve C is extremal satisfying boundary condition.
2. The extremal C must be embedded in the field of extremals.
3. At a point (x, y) closed to the curve C and for arbitrary value of y', the Weirstrass function E does not change sign.
4. For strong minimum $E \geq 0$ or $F_{y'y'} > 0$ at any point close to C and also arbitrary value of y'

   and for strong maximum $E \leq 0$ or $F_{y'y'} < 0$ at points closed to the curve C and also arbitrary value of y' .

Example 20.3.1. Test for an extremal of the functional

$$I[y(x)] = \int_0^2 (e^{y'} + 3)\, dx$$

with $y(0) = 0, y(2) = 1$.

Solution. Here $F = e^{y'} + 3$. Hence by Euler's equation

$$\begin{aligned}
&\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0 \\
&\Rightarrow 0 - \frac{d}{dx}(e^{y'}) = 0 \\
&\Rightarrow e^{y'}\frac{d^2y}{dx^2} = 0 \\
&\Rightarrow \frac{d^2y}{dx^2} = 0 \; [\because e^{y'} \neq 0].
\end{aligned}$$

Integrating twice, we obtain

$$y = cx + d.$$

Now, $y(0) = 0, y(2) = 1$ gives $c = \frac{1}{2}, d = 0$.

Hence required extremal is

$$y = \frac{1}{2}x.$$

which clearly satisfies boundary conditions and is included in the central field of extremals $y = cx$.

Also, $\quad F_{y'y'} = e^{y'} > 0$ for any value of y'.

Consequently, the given functional is strong minimum on the extremal $y = \frac{1}{2}x$. ■

20.4 Exercise

1. Test for an extremal of the functional

$$I[y(x)] = \int_0^a (y'^2 - y^2)\, dx$$

   with $y(0), y(a) = 0, a > 0$.

2. Investigate for an extremal of the functional

$$I[y(x)] = \int_0^1 (x + 2y - \frac{1}{2}y'^2)\, dx$$

   with $y(0), y(1) = 0$.

References

1. T. M. Apostol: Mathematical Analysis.
2. M. Spivak: Calculus on Manifolds.
3. W. Rudin: Principles of Mathematical Analysis
4. Fuzzy Sets and Fuzzy Logic Theory and Applications: G.J. Klir and B. Yuan.
5. Introduction to Fuzzy Arithmetic Theory and Applications: A. Kaufmann and M.M. Gupta.
6. Fuzzy Set Theory: R. Lowen.
7. Fuzzy Set Theory and Its Applications: H.-J. Zimmermann.
8. A.S. Gupta: Calculus of Variations with Applications, Prentice –Hall of India.
9. I.M. Gelfand and S.V. Fomin, Calculus of Variations, Prentice Hall Inc.
10. L. Elsgolts: Differential equations and the Calculus of Variations.

NOTES