MA 105 : Calculus
Division 1, Lecture 01

Prof. Sudhir R. Ghorpade
IIT Bombay

Generalities about the Course

- Instructor: Prof. Sudhir R. Ghorpade, 106B, Maths Dept.
- Lecture Hours: Mon, Thu 2.00 – 3.25 PM, in LA 001.
- Tutorial: Wed, 2 – 2.55 pm, in LT 001 – 006.
- Attendance: Compulsory! (Also, it will be good for you!)
- Office Hours: Mondays 11.30 am – 12.30 pm.
- Evaluation Plan: Short Quizzes in Tuts (10 %), Common quizzes (10% $\times$ 2), Mid-Sem (30 %), End-Sem (40%).
- More Information:
    - The Booklet
    - Moodle page of the course
    - Instructor's web page, and especially, the course page: http://www.math.iitb.ac.in/~srg/autumn2019.html

Text, References, and Acknowledgements

The treatment of calculus in these lectures will be based on the following two books by S. R. Ghorpade and B. V. Limaye, which are published by Springer, New York.

[GL-1] A Course in Calculus and Real Analysis, 2nd Ed., 2018.

[GL-2] A Course in Multivariable Calculus and Analysis, 2010.

Besides these, the other references listed in the booklet, especially the book of Thomas and Finney, may be consulted. For later parts of the course, it is also useful to see the book Basic Multivariable Calculus by J. E. Marsden, A. J. Tromba and A. Weinstein (Springer, New York, 1993).

Acknowledgement: I shall mainly use the slides of Calculus lectures prepared recently by Prof. B. V. Limaye. These slides acknowledged the use of the lecture notes of similar courses given by myself and by Prof. Prachi Mahajan in the past.

Notation

- $\mathbb{N} := \{1, 2, 3, \ldots\}$ [the set of positive integers]
- $\mathbb{Z} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ [the set of integers]
- $\mathbb{Q} :=$ the set of all rational numbers $= \{m/n : m, n \in \mathbb{Z},\ n \neq 0\}$

There is no rational number whose square is 2.

Proof: Suppose not! Then $(p/q)^2 = 2$, that is, $p^2 = 2q^2$ for some $p, q \in \mathbb{Z}$ such that $q \neq 0$, and p and q have no common factor. Now p^2 is even, and so p is even. Hence there is an integer r such that $p = 2r$. Then $2q^2 = p^2 = (2r)^2 = 4r^2$, and so $q^2 = 2r^2$. Thus q^2 is even, and so q is also even. Thus 2 is a common factor of p and q, which is a contradiction.

Optional Exercise: If $d \in \mathbb{N}$ is not the square of an integer, then show that there is no rational number whose square is d.

Let $\mathbb{R}$ denote the set of all real numbers. We will assume the following things about the set $\mathbb{R}$.

The set $\mathbb{Q}$ of all rational numbers is contained in $\mathbb{R}$, and the set $\mathbb{R}$ of all real numbers satisfies

- Algebraic Properties regarding addition and multiplication.
- Order Properties: There is a subset $\mathbb{R}^+$ of $\mathbb{R}$ such that
  (i) Given $a \in \mathbb{R}$, exactly one of the following holds:
  $a \in \mathbb{R}^+$ or $a = 0$ or $-a \in \mathbb{R}^+$
  (ii) $a, b \in \mathbb{R}^+ \Longrightarrow a + b \in \mathbb{R}^+$ and $ab \in \mathbb{R}^+$.

  Define $a < b$ if $b - a \in \mathbb{R}^+$. Thus $\mathbb{R}^+ = \{x \in \mathbb{R} : 0 < x\}$.
- Completeness Property, which we shall state later.

Elements of the set $\mathbb{R} \setminus \mathbb{Q}$, that is, those real numbers which are not rational numbers, are called **irrational numbers**.

Notation: We write $a \leq b$ if $a < b$ or $a = b$.
Also, we write $a > b$ if $b < a$, and $a \geq b$ if $a > b$ or $a = b$.

Boundedness of a subset of $\mathbb{R}$

Let E be a subset of $\mathbb{R}$, that is, $E \subset \mathbb{R}$.

- E is called **bounded above** if there is $\alpha \in \mathbb{R}$ such that $x \leq \alpha$ for all $x \in E$.
  Any such α is an **upper bound** of E.
- E is called **bounded below** if there is $\beta \in \mathbb{R}$ such that $x \geq \beta$ for all $x \in E$.
  Any such β is a **lower bound** of E.
- E is **bounded** if it is bounded above and bounded below.

We say that M is the **maximum** of E if M is an upper bound of E and $M \in E$, and we say that m is the **minimum** of E if m is a lower bound of E and $m \in E$.

Supremum (sup or lub) and Infimum (inf or glb)

Let $E \subset \mathbb{R}$.

- A real number α is called a **supremum** or a **least upper bound** of E if α is an upper bound of E (that is, $x \le \alpha$ for all $x \in E$), and $\alpha \le u$ for every upper bound u of E.

- A real number β is called an **infimum** or a **greatest lower bound** of E if β is a lower bound of E (that is, $\beta \le x$ for all $x \in E$), and $v \le \beta$ for every lower bound v of E.

- If E has a supremum, then it is unique, and it is denoted by sup E or lub E. Similarily, if E has an infimum, then it is unique and is denoted by inf E or glb E.

Example: Let $E := \{x \in \mathbb{R} : 0 < x \le 1\}$. Then $\sup E = 1$ and $\inf E = 0$. Also, $\max E = 1$, but E has no minimum.

Completeness Property of $\mathbb{R}$:

A nonempty subset of $\mathbb{R}$ that is bounded above has a supremum, that is, a least upper bound.

Consequences of the Completeness Property:

- A nonempty subset E of $\mathbb{R}$ that is bounded below has an infimum, that is, a greatest lower bound.
  In fact, the set $F := \{-x : x \in E\}$ is bounded above, and if α is the lub of F, then $\beta := -\alpha$ is the glb of E.
- Archimedean Property: Given $x \in \mathbb{R}$, there is $n \in \mathbb{N}$ such that $n > x$.
  Proof: Suppose not! Then $n \leq x$ for all $n \in \mathbb{N}$, that is, x is an upper bound of the set $\mathbb{N}$. Let $\alpha := \sup \mathbb{N}$. Then $\alpha - 1$ is not an upper bound of $\mathbb{N}$, that is, there is $n_0 \in \mathbb{N}$ such that $\alpha - 1 < n_0$. But then $\alpha < n_0 + 1 \leq \alpha$, since $(n_0 + 1) \in \mathbb{N}$ and α is an upper bound of $\mathbb{N}$. Thus we obtain $\alpha < \alpha$, which is a contradiction.

- Let $x \in \mathbb{R}$. Applying the Archimedean property to x and $-x$, we see that there are $\ell, n \in \mathbb{N}$ such that $-\ell < x < n$. The largest among finitely many integers k satisfying $-\ell \le k \le n$ and also $k \le x$ is called the **integer part** of x, and is denoted by $[x]$ or by $\lfloor x \rfloor$. Note that $[x]$ is the largest integer $\le x$ and it is characterized by the following two properties: (i) $[x] \in \mathbb{Z}$ and (ii) $x - 1 \le [x] \le x$.

- Let $a \in \mathbb{R}^+$ and $n \in \mathbb{N}$. Then there is a unique $b \in \mathbb{R}^+$ such that $b^n = a$. This real number b is called the **positive** n**th root of** a, and we denote it by $a^{1/n}$.

  Example (the positive square root of 2):
  Let $S := \{x \in \mathbb{R} : x^2 \le 2\}$. Then S is nonempty since $1 \in S$ and S is bounded above by 2. By the completeness property of $\mathbb{R}$, let $b := \sup S$. Then $b \ge 1 > 0$. Also, we obtain $b^2 = 2$ by showing that both $b^2 < 2$ and $b^2 > 2$ lead to contradictions. (Verify!) Thus $b := \sqrt{2}$. Since $b \in S$, we see that $b = \max S$.

- Let $a < b$ in $\mathbb{R}$. Then there is a rational number r such that $a < r < b$. In fact, we can consider $r := m/n$, where $n > 1/(b-a)$ and $m := [na] + 1$.
- Let $a < b$ in $\mathbb{R}$. Then there is an irrational number s such that $a < s < b$. In fact, since $a + \sqrt{2} < b + \sqrt{2}$, let $r \in \mathbb{Q}$ be such that $a + \sqrt{2} < r < b + \sqrt{2}$. Then $a < r - \sqrt{2} < b$, where $s := r - \sqrt{2}$ is an irrational number.

Thus we obtain the following important result.

Between any two real numbers, there is a rational number as well as an irrational number.

Optional Exercise: Write down more detailed versions of the "proofs" sketched above. Consult [GL-1], if desired.

Intervals

Given any $a, b \in \mathbb{R}$, we define

- $(a, b) := \{x \in \mathbb{R} : a < x < b\}$ [open interval]
- $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$ [closed interval]
- Semi-open intervals $(a, b]$ and $[a, b)$ are defined similarly.
- It is also useful to consider symbols ∞ and $-\infty$ and define the infinite intervals

$$(a, \infty) := \{x \in \mathbb{R} : x > a\}, \quad [a, \infty) := \{x \in \mathbb{R} : x \geq a\},$$
$$(-\infty, a) := \{x \in \mathbb{R} : x < a\}, \quad (-\infty, a] := \{x \in \mathbb{R} : x \leq a\}.$$

  Also, one writes $\mathbb{R} = (-\infty, \infty)$ and refers to this as an infinite interval, or sometimes, a doubly infinite interval.
- In general, a subset I of $\mathbb{R}$ is an interval if

$$x, y \in I, \ x < y \Longrightarrow [x, y] \subseteq I.$$

- One can show that every interval in $\mathbb{R}$ is open, closed, semi-open, or infinite interval.

Absolute Value

- For $x \in \mathbb{R}$, the absolute value or the modulus of x is

$$|x| := \begin{cases} x & \text{if } x \geq 0, \\ -x & \text{if } x < 0. \end{cases}$$

- Basic Properties: For any $x, y \in \mathbb{R}$,
  - $|x + y| \leq |x| + |y|$ [Triangle Inequality]
  - $||x| - |y|| \leq |x - y|$.
- Optional Exercises: (i) Show that for any $a, b \in \mathbb{R}$ with $a \geq 0$, $b \geq 0$, and $n \in \mathbb{N}$,

$$\left| \sqrt[n]{a} - \sqrt[n]{b} \right| \leq \sqrt[n]{|a - b|}.$$

  (ii) For any $n \in \mathbb{N}$ and any nonnegative real numbers $a_1, \ldots, a_n$, prove the AM-GM inequality:

$$\frac{a_1 + \cdots + a_n}{n} \geq \sqrt[n]{a_1 \cdots a_n}.$$

Functions

- Given sets D, E, a function $f : D \to E$ assigns to each $x \in D$, a unique element of E, denoted $f(x)$. We refer to D as the domain of f and E as the co-domain of f. The set $\{f(x) : x \in D\}$ of all values taken by the function is called the range of f. [For a formal definition, see [GL-1].]
- A function $f : D \to E$ is said to be:
  - one-one (or injective) if for any $x_1, x_2 \in D$,
    $$f(x_1) = f(x_2) \Longrightarrow x_1 = x_2.$$
  - onto (or surjective) if its range is E.
- If $f : D \to E$ is bijective, then it has an inverse $g : E \to D$ with the property that the composites $g \circ f$ and $f \circ g$ are identity functions, i.e.,

$$g(f(x)) = x \text{ for all } x \in D \quad \text{and} \quad f(g(y)) = y \text{ for all } y \in E.$$

Examples and Types of Functions

- Example/Exercise: Consider $f_1, f_2, f_3 : \mathbb{R} \to \mathbb{R}$ and $f_4 : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ defined by

  $$f_1(x) = 2x+1, \quad f_2(x) = x^2, \quad f_3(x) = x^3, \quad \text{and} \quad f_4(x) = \frac{1}{x}.$$

  Are these one-one/onto/bijective?
- Functions that mainly arise in Calculus are usually of following types.
  - Polynomial functions
  - Rational functions
  - Algebraic functions
  - Transcendental functions (this includes logarithmic, exponential and trigonometric functions).
- Besides these, we can construct functions by pieceing together known functions (such as those belonging to the above classes). A good example is the absolute value function $x \longmapsto |x|$.

Sequences

Definition

Let X be any set. A sequence in X is a function from the set $\mathbb{N}$ of natural numbers to the set X.

The value of this function at $n \in \mathbb{N}$ is denoted by $a_n \in X$, and a_n is called the nth **term** of the sequence.

We shall use the notation (a_n) to denote a sequence.

Note: $\{a_n : n \in \mathbb{N}\}$ is the set of all terms of the sequence (a_n). Thus if $X := \mathbb{R}$ and $a_n := (-1)^n$ for $n \in \mathbb{N}$, then the sequence (a_n) is given by $-1, 1, -1, 1, \ldots$, but $\{a_n : n \in \mathbb{N}\} = \{-1, 1\}$.

Initially, we let $X := \mathbb{R}$, that is, we consider sequences in $\mathbb{R}$. Later, we shall consider sequences in $\mathbb{R}^2$ and in $\mathbb{R}^3$.

Examples

Examples:

1. $a_n := n$ for $n \in \mathbb{N}$: $1, 2, 3, 4, \ldots$
2. $a_n := 1/n$ for $n \in \mathbb{N}$: $1,\ 1/2,\ 1/3,\ 1/4, \ldots$
3. $a_n := n^2$ for $n \in \mathbb{N}$: $1, 4, 9, 16, \ldots$
4. $a_n := \sqrt{2}$ for $n \in \mathbb{N}$: $\sqrt{2}, \sqrt{2}, \ldots$ This is an example of a constant sequence.
5. $a_n := 2^n$ for $n \in \mathbb{N}$: $2, 4, 8, 16, \ldots$
6. $a_n := (-1)^n$ for $n \in \mathbb{N}$: $-1, 1, -1, 1, \ldots$
7. $a_1 := 1, a_2 := 1$ and $a_n := a_{n-1} + a_{n-2}$ for $n \geq 3$: $1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \ldots$ This sequence is known as the Fibonacci sequence.
8. $a_n := 1/2 + \cdots + 1/2^n$ for $n \in \mathbb{N}$. Check: $a_n = 1 - (1/2^n)$.

Bounded sequences

A sequence (a_n) of real numbers is said to be **bounded above** if the set $\{a_n : n \in \mathbb{N}\}$ is bounded above, that is, if there is a real number α such that $a_n \leq \alpha$ for every $n \in \mathbb{N}$.

A sequence (a_n) of real numbers is said to be **bounded below** if the set $\{a_n : n \in \mathbb{N}\}$ is bounded below, that is, if there is a real number β such that $\beta \leq a_n$ for every $n \in \mathbb{N}$.

A sequence (a_n) of real numbers is said to be **bounded** if it is bounded above as well as bounded below, that is, if there are real numbers α, β such that $\beta \leq a_n \leq \alpha$ for every $n \in \mathbb{N}$.

If a sequence is not bounded, it is said to be **unbounded**.

Let us check which of the sequences mentioned earlier are bounded above and/or bounded below.

Examples of bounded and unbounded sequences:

- $a_n := n$ for $n \in \mathbb{N}$: $\beta = 1$.
- $a_n := 1/n$ for $n \in \mathbb{N}$: $\beta = 0,\ \alpha = 1$.
- $a_n := n^2$ for $n \in \mathbb{N}$: $\beta = 1$.
- $a_n := \sqrt{2}$ for $n \in \mathbb{N}$: $\beta = \sqrt{2} = \alpha$.
- $a_n := 2^n$ for $n \in \mathbb{N}$: $\beta = 2$.
- $a_n := (-1)^n$ for $n \in \mathbb{N}$: $\beta = -1,\ \alpha = 1$.
- $a_1 := 1, a_2 := 1$ and $a_n := a_{n-1} + a_{n-2}$ for $n \geq 3$: $\beta = 1$.
- $a_n := 1/2 + \cdots + 1/2^n$ for $n \in \mathbb{N}$: $\beta = 1/2,\ \alpha = 1$.

Note: The sequence given by $a_n := n$ for $n \in \mathbb{N}$, is not bounded above by the Archimedean property of $\mathbb{R}$.

Toward convergence of a sequence

Let (a_n) be a sequence in $\mathbb{R}$, and let a be a real number.

- Convergence of a sequence (a_n) to a real number a should mean that the term a_n is as close to a as we like for all sufficiently large n.
- Fix any positive number. Construct the sequence

  $$|a_1 - a|, |a_2 - a|, \ldots, |a_n - a|, \ldots$$

  After a certain stage, all the entries from this sequence should be smaller than the fixed positive number.
- The fixed positive number is often denoted by ϵ (epsilon).
- Let n_0 indicate how far one needs to go in the sequence to ensure that the entries from $|a_{n_0} - a|$ <u>onward</u> are smaller than ϵ, that is, $a_{n_0}, a_{n_0+1}, a_{n_0+2}, \ldots$ all belong to $(a - \epsilon, a + \epsilon)$.

Definition of convergence of a sequence

Definition

Let (a_n) be a sequence of real numbers. We say that (a_n) is convergent if there is $a \in \mathbb{R}$ such that the following condition holds. For every $\epsilon > 0$, there is $n_0 \in \mathbb{N}$ such that $|a_n - a| < \epsilon$ for all $n \geq n_0$.

This is known as the ϵ-n_0 definition of convergence of a sequence.

In this case, we say that (a_n) **converges** to a, or that a is a **limit** of (a_n), and we write

$$\lim_{n\to\infty} a_n = a \quad \text{or} \quad a_n \to a \ \ (\text{as } n \to \infty).$$

If a sequence does not converge, we say that the sequence **diverges** or it is **divergent**.

The convergence of a sequence is unaltered if a finite number of its terms are replaced by some other terms.

Examples:

(i) Let $a \in \mathbb{R}$ and $a_n := a$ for all $n \in \mathbb{N}$. Then $a_n \to a$. We can let $n_0 := 1$.

(ii) $a_n := 1/n$ for all $n \in \mathbb{N}$. Then $a_n \to 0$.

Let $\epsilon > 0$ be given. We want to find $n_0 \in \mathbb{N}$ such that $|(1/n) - 0| < \epsilon$ for all $n \geq n_0$.

Choose any $n_0 \in \mathbb{N}$ which is greater than $1/\epsilon$. This is possible because of the Archimedean property of $\mathbb{R}$. For example, we can let $n_0 := [1/\epsilon] + 1$.

MA 105 : Calculus
Division 1, Lecture 02

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Generalities about the course
- Text and References
- Basic notations: $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$.
- The set $\mathbb{R}$ of real numbers as a set containing $\mathbb{Q}$ and satisfying the
  - Algebraic Properties
  - Order Properties
  - Completeness Property
- Upper bound, lower bound, supremum,and infimum
- Consequences of the Completeness Property:
  - Existence of infimum of a nonempty set bounded below;
  - Archimedean Property;
  - Existence and uniqueness of positive nth roots of positive real numbers;
  - Existence of a rational as well as an irrational between any two real numbers $a < b$.

Recap of the previous lecture Contd.

- Intervals: open, closed, semi-open, infinite
- Functions. Notions of one-one, onto, and bijective functions; Inverse functions
- Examples of functions. Mention of:
  - Polynomials functions
  - Rational functions
  - Algebraic functions
  - Transcendental functions.
  - Functions obtained by piecing together known functions
- Sequences: Basic definitions and examples
- Boundedness. Examples of bounded and unbounded sequences
- Notion of convergence of a sequence. A few examples

Definition of convergence of a sequence

Definition

Let (a_n) be a sequence of real numbers. We say that (a_n) is convergent if there is $a \in \mathbb{R}$ such that the following condition holds. For every $\epsilon > 0$, there is $n_0 \in \mathbb{N}$ such that

$$|a_n - a| < \epsilon \text{ for all } n \geq n_0.$$

In this case, we say that (a_n) **converges** to a, or that a is a **limit** of (a_n), and we write $\lim_{n\to\infty} a_n = a$ or $a_n \to a$ as $n \to \infty$).

The part in red may be stated as: $|a_n - a| < \epsilon$ for all large n. In general, when we say that a statement holds for all large n, it means that there is $n_0 \in \mathbb{N}$ such that it holds for all $n \geq n_0$.

If a sequence is not convergent, then it is said to be **divergent**, *or we say that the sequence* **diverges**.

The convergence of a sequence is unaltered if a finite number of its terms are replaced by some other terms.

Examples:

(i) Let $a \in \mathbb{R}$ and $a_n := a$ for all $n \in \mathbb{N}$. Then $a_n \to a$. We can let $n_0 := 1$.

(ii) $a_n := 1/n$ for all $n \in \mathbb{N}$. Then $a_n \to 0$.

Let $\epsilon > 0$ be given. We want to find $n_0 \in \mathbb{N}$ such that $|(1/n) - 0| < \epsilon$ for all $n \geq n_0$.

Choose any $n_0 \in \mathbb{N}$ which is greater than $1/\epsilon$. This is possible because of the Archimedean property of $\mathbb{R}$. For example, we can let $n_0 := [1/\epsilon] + 1$.

(iii) $a_n := 2/(n^2+1)$ for $n \in \mathbb{N}$. Then $a_n \to 0$. Now

$$\left|\frac{2}{n^2+1} - 0\right| = \frac{2}{n^2+1} < \frac{2}{n^2} \quad \text{for all } n \in \mathbb{N}.$$

Choose $n_0 \in \mathbb{N}$ such that $n_0 > \sqrt{2}/\sqrt{\epsilon}$. For example, let $n_0 := [\sqrt{2}/\sqrt{\epsilon}] + 1$. Then $|a_n - 0| < \epsilon$ for all $n \geq n_0$.

(iv) $a_n := 5/(3n+1)$ for $n \in \mathbb{N}$. Then $a_n \to 0$. Now

$$\frac{5}{3n+1} < \frac{5}{3n} \quad \text{for all } n \in \mathbb{N}.$$

Choose $n_0 \in \mathbb{N}$ such that $n_0 > 5/3\epsilon$. For example, let $n_0 := [5/3\epsilon] + 1$. Then $|a_n - 0| < \epsilon$ for all $n \geq n_0$.

(v) $a_n := (-1)^n$ for all $n \in \mathbb{N}$. Then the sequence (a_n) is divergent, that is, it is not convergent.

Suppose (a_n) is convergent. Then there is a real number a such that $a_n \to a$. Let $\epsilon := 1/2$. Find $n_0 \in \mathbb{N}$ such that

$$|(-1)^n - a| < \frac{1}{2} \text{ for all } n \geq n_0.$$

Since $(-1)^{2n_0} = 1$ and $(-1)^{2n_0+1} = -1$,

$$\begin{aligned} 2 &= |(-1)^{2n_0} - (-1)^{2n_0+1}| \\ &\leq |(-1)^{2n_0} - a| + |a - (-1)^{2n_0+1}| \\ &< \frac{1}{2} + \frac{1}{2} = 1, \end{aligned}$$

which is a contradiction.

Uniqueness of limit

Theorem

A convergent sequence has a unique limit.

Proof: Let (a_n) be a sequence. Assume for a moment that $a_n \to a$ and $a_n \to b$, where $a \neq b$. Let $\epsilon := |a-b|/2 > 0$.

Let $n_0 \in \mathbb{N}$ be such that $n \geq n_0 \implies |a_n - a| < \epsilon$, and let $m_0 \in \mathbb{N}$ be such that $n \geq m_0 \implies |a_n - b| < \epsilon$. Consider $n := \max\{n_0, m_0\}$. Then

$$|a-b| \leq |a-a_n| + |a_n - b| < \epsilon + \epsilon = |a-b|,$$

which is a contradiction. Hence $a = b$. $\square$

A useful result

Theorem

Every convergent sequence is bounded.

Proof: Suppose $a_n \to a$. Let $\epsilon := 1$. There is $n_0 \in \mathbb{N}$ such that

$$|a_n - a| < 1 \text{ for all } n > n_0.$$

Hence

$$|a_n| \leq |a_n - a| + |a| < 1 + |a| \text{ for all } n > n_0.$$

Define $\alpha := \max\left\{|a_1|, \ldots, |a_{n_0}|, |a| + 1\right\}$. Then $|a_n| \leq \alpha$ for all $n \in \mathbb{N}$. Hence (a_n) is bounded. $\square$

- A bounded sequence need not be convergent.
  Example: $a_n := (-1)^n$ for $n \in \mathbb{N}$.
- If $a_n := (-1)^n n$ for $n \in \mathbb{N}$, then (a_n) is divergent since it is not bounded.

Limit theorems for sequences

Suppose (a_n) and (b_n) are convergent sequences. Then

- $\lim_{n\to\infty}(a_n \pm b_n) = (\lim_{n\to\infty} a_n) \pm (\lim_{n\to\infty} b_n)$ resp.,
- $\lim_{n\to\infty}(a_n b_n) = (\lim_{n\to\infty} a_n) \cdot (\lim_{n\to\infty} b_n)$.

  In particular, $\lim_{n\to\infty}(r\, a_n) = r(\lim_{n\to\infty} a_n)$ for $r \in \mathbb{R}$.
- If $\lim_{n\to\infty} b_n \neq 0$, then there is $\delta > 0$ such that $|b_n| \geq \delta > 0$ for all large n and

$$\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{\lim_{n\to\infty} a_n}{\lim_{n\to\infty} b_n}.$$

- If $a_n \leq b_n$ for all large $n \in \mathbb{N}$, then $\lim_{n\to\infty} a_n \leq \lim_{n\to\infty} b_n$.
- (Sandwich theorem) If $a_n \leq c_n \leq b_n$ for all large $n \in \mathbb{N}$, and $\lim_{n\to\infty} a_n = c = \lim_{n\to\infty} b_n$, then $\lim_{n\to\infty} c_n = c$.

Examples

(i) Let $a \in \mathbb{R}$, and $a_n := a^n$ for $n \in \mathbb{N}$.
Then (a_n) is convergent $\iff -1 < a \leq 1$.

Clearly, if $a := 0$, then $a_n \to 0$, and if $a := 1$, then $a_n \to 1$.
Also, if $a := -1$, then we have seen that (a_n) is divergent.

Let $0 < |a| < 1$, and $r := 1/|a|$. Then $r > 1$, and so $r = 1 + h$ with $h > 0$. By the binomial theorem,

$$r^n = (1+h)^n = 1 + nh + \cdots + h^n > nh \quad \text{for all } n \in \mathbb{N}.$$

Hence $0 \leq |a_n| = |a|^n = (1/r^n) \leq (1/nh) \to 0$. Thus $a_n \to 0$.

Let $s := |a| > 1$. Then $s = 1 + h$ with $h > 0$, and for all $n \in \mathbb{N}$, $|a_n| = s^n > nh$. Hence (a_n) is unbounded, and so it is divergent.

Examples (continued)

(ii) Let $a_n := \dfrac{n^3 + 3n^2 + 1}{n^4 + 8n^2 + 2}$ for $n \in \mathbb{N}$. Then $a_n \to 0$,

since $0 \le a_n \le \dfrac{1}{n} + \dfrac{3}{8n^2} + \dfrac{1}{n^4} \to 0$.

(iii) Let $a_n := \dfrac{1}{n} \sin\left(\dfrac{1}{n}\right)$ for $n \in \mathbb{N}$. Then $a_n \to 0$,

since $|a_n| \le \dfrac{1}{n} \to 0$.

Monotonic sequences

A sequence (a_n) is said to be **increasing** if $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$, that is, $a_1 \leq a_2 \leq a_3 \leq \cdots$.

A sequence (a_n) is said to be **decreasing** if $a_n \geq a_{n+1}$ for all $n \in \mathbb{N}$, that is, $a_1 \geq a_2 \geq a_3 \geq \cdots$.

A sequence is **monotonic** if it is either increasing or decreasing.

Examples:

- The sequences (n) and $(-1/n)$ are increasing.
- The sequences $(-n)$ and $(1/n)$ are decreasing.
- The sequences $((-1)^n)$ and $((-1)^n n)$ are neither increasing nor decreasing. Thus they are not monotonic.

Convergence theorems

- An increasing sequence (a_n) that is bounded above is convergent, and $\lim_{n\to\infty} a_n = \sup\{a_n : n \in \mathbb{N}\}$.
- A decreasing sequence (a_n) that is bounded below is convergent, and $\lim_{n\to\infty} a_n = \inf\{a_n : n \in \mathbb{N}\}$.

Thus, a monotonic bounded sequence is convergent.

Proof: Suppose (a_n) is increasing and bounded above. Let $a := \sup\{a_n : n \in \mathbb{N}\}$. Let $\epsilon > 0$. Then $a - \epsilon$ is not an upper bound of $\{a_n : n \in \mathbb{N}\}$, and so there is $n_0 \in \mathbb{N}$ such that $a - \epsilon < a_{n_0}$. But then

$$a - \epsilon < a_{n_0} \le a_n \le a \quad \text{for all } n \ge n_0.$$

Thus $|a_n - a| < \epsilon$ for all $n \ge n_0$. Hence $a_n \to a$. A similar proof holds for a decreasing sequence that is bounded below.

Examples: 1. The sequence (a_n) defined by $a_n := 1 + (1/2^2) + \cdots + (1/n^2)$ for $n \in \mathbb{N}$ is clearly increasing. Further, since

$$\begin{aligned} a_n &\leq 1 + \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{(n-1) \cdot n} \\ &= 1 + \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n}\right) \end{aligned}$$

we see that $a_n \leq 2 - (1/n) < 2$. Thus (a_n) is bounded. Hence by the result in the previous slide, (a_n) is convergent.

2. Consider the sequence (a_n) defined recursively by

$$a_1 := \frac{3}{2} \quad \text{and} \quad a_{n+1} := \frac{1}{2}\left(a_n + \frac{2}{a_n}\right) \quad \text{for } n \in \mathbb{N}.$$

Note that if $1 \leq a_n \leq 2$, then it is easily deduced from the recurrence relation that $1 \leq a_{n+1} \leq 2$. Since $1 \leq a_1 \leq 2$, we see by induction that $1 \leq a_n \leq 2$ for all $n \in \mathbb{N}$. In particular, (a_n) is bounded. Also, $a_2 = 17/12 < a_1$.

Let us check whether the sequence (a_n) is decreasing. Since

$$a_n - a_{n+1} = a_n - \frac{a_n^2+2}{2a_n} = \frac{a_n^2-2}{2\,a_n} \quad \text{for all } n \in \mathbb{N},$$

(a_n) is decreasing if and only if $a_n^2 - 2 \geq 0$ for all $n \in \mathbb{N}$. But

$$a_1^2 \geq 2 \quad \text{and} \quad a_{n+1}^2 - 2 = \frac{\left(a_n^2-2\right)^2}{4a_n^2} \geq 0 \text{ for all } n \in \mathbb{N}.$$

Hence the sequence (a_n) is decreasing. It follows that (a_n) is convergent. Let $a_n \to a$. Then $a_{n+1} \to a$ as well. Moreover, since $1 \leq a_n \leq 2$ for all $n \in \mathbb{N}$, we see that $1 \leq a \leq 2$. Now

$$a_{n+1} = \frac{1}{2}\left(a_n + \frac{2}{a_n}\right) \to \frac{1}{2}\left(a + \frac{2}{a}\right).$$

Since the limit of a sequence is unique, we see that

$$\frac{1}{2}\left(a + \frac{2}{a}\right) = a, \quad \text{that is,} \quad a^2 = 2.$$

Thus $a = \pm\sqrt{2}$. But since $1 \leq a \leq 2$, it follows that $a = \sqrt{2}$.

Let D be a subset of $\mathbb{R}$, and f be a real-valued function defined on D. Then we write $f : D \to \mathbb{R}$. The subset $f(D) := \{f(x) : x \in D\}$ of $\mathbb{R}$ is called the **range** of f.

Definition

Let $D \subset \mathbb{R}$, $f : D \to \mathbb{R}$ and $c \in D$. We say that f is **continuous** at c if
(x_n) is a sequence in D, $x_n \to c \implies f(x_n) \to f(c)$.

We say that f is **continuous on** D if f is continuous at each point of D.

We say that f is **discontinuous** on D if it is not continuous on D, that is, if there is $c \in D$ at which f is not continuous.

Examples

(i) Let $D := \mathbb{R}$ and $f(x) := x$ for $x \in \mathbb{R}$. Clearly, f is continuous on $\mathbb{R}$.

(ii) Let $D := [0, \infty)$ and $f(x) := \sqrt{x}$ for $x \in [0, \infty)$. First we show that f is continuous at 0. Let $x_n \to 0$, where $x_n \geq 0$ for each $n \in \mathbb{N}$. Consider $\epsilon > 0$. Find $n_0 \in \mathbb{N}$ such that $x_n < \epsilon^2$ for all $n \geq n_0$. Then $0 \leq f(x_n) = \sqrt{x_n} < \epsilon$ for all $n \geq n_0$, and so $f(x_n) \to 0 = f(0)$. Thus f is continuous at 0.

Next, let $c \in (0, \infty)$, and $x_n \to c$, where $x_n \geq 0$ for each $n \in \mathbb{N}$. It is easy to see that $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x - y|}$ for all $x, y \in [0, \infty)$ by considering $x \geq y$ and $x < y$. Since $|x_n - c| \to 0$, we obtain $|\sqrt{x_n} - \sqrt{c}| \leq \sqrt{|x_n - c|} \to 0$, that is, $f(x_n) \to f(c)$. Thus f is continuous at c.

More generally, if $k \in \mathbb{N}$, and $f(x) := x^{1/k}$ for $x \in [0, \infty)$, then f is continuous on $[0, \infty)$. We shall later give a proof of this result.

Basic Properties of Continuous functions

- Let $D \subset \mathbb{R}$, and let $f, g : D \to \mathbb{R}$ be functions. Let $c \in D$, and let f and g be continuous at c. Then $f + g$ and $f \cdot g$ are continuous at c. Further, f/g is continuous at c if $g(c) \neq 0$. These results follow from the limit theorems for sequences.

  Consequences:

  (i) A polynomial is continuous on $\mathbb{R}$.

  (ii) A **rational function** $r(x) := p(x)/q(x)$, where p and q are polynomials, is continuous at $c \in \mathbb{R}$ if $q(c) \neq 0$.

- Composite of continuous functions: Let $D \subset \mathbb{R}$, $E \subset \mathbb{R}$. If $f : D \to \mathbb{R}$ and $g : E \to \mathbb{R}$ are functions such that $f(D) \subset E$, f is continuous at $c \in D$, and g is continuous at $f(c) \in E$, then $g \circ f : D \to \mathbb{R}$ is continuous at c.

MA 105 : Calculus
Division 1, Lecture 03

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Example of a divergent sequence
- Basic results: Uniqueness of limit, Convergent sequences are bounded; Limit theorems for sequences
- Sandwich theorem. Application to the sequence of powers of a fixed real number
- Monotonic sequences and their basic properties. Examples
- Oral mention of supplementary topics: Notion of subsequences and Cauchy sequences. Bolzano-Weierstrass Theorem and Cauchy Criterion
- Continuity of a function at a point. Continuous and discontinuous functions. Examples.
- Basic properties:
    - Sums, products and quotients of continuous functions are continuous (wherever defined)
    - Composites of continuous functions are continuous.

Basic Properties of Continuous functions

- Let $D \subset \mathbb{R}$, and let $f, g : D \to \mathbb{R}$ be functions. Let $c \in D$, and let f and g be continuous at c. Then $f + g$ and $f \cdot g$ are continuous at c. Further, f/g is continuous at c if $g(c) \neq 0$. These results follow from the limit theorems for sequences.

  Consequences:

  (i) A polynomial is continuous on $\mathbb{R}$.

  (ii) A **rational function** $r(x) := p(x)/q(x)$, where p and q are polynomials, is continuous at $c \in \mathbb{R}$ if $q(c) \neq 0$.

- Composite of continuous functions: Let $D \subset \mathbb{R}$, $E \subset \mathbb{R}$. If $f : D \to \mathbb{R}$ and $g : E \to \mathbb{R}$ are functions such that $f(D) \subseteq E$, f is continuous at $c \in D$, and g is continuous at $f(c) \in E$, then $g \circ f : D \to \mathbb{R}$ is continuous at c.

Equivalent condition for continuity of a function

Let $D \subseteq \mathbb{R}$, $f : D \to \mathbb{R}$ and $c \in D$. Then f is continuous at c if and only if the following ϵ-δ condition holds:
For every $\epsilon > 0$, there exists $\delta > 0$ such that

$$x \in D \text{ and } |x - c| < \delta \implies |f(x) - f(c)| < \epsilon.$$

Note: Proof is skipped. (For those interested, deriving the continuity from the ϵ-δ condition is easy. The converse isn't difficult if you write down the negation of the ϵ-δ condition.)

Example: Define $f(x) := \begin{cases} x \sin(1/x) & \text{if } x \in \mathbb{R},\ x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$

Given any $\epsilon > 0$, we can take $\delta := \epsilon$. Then

$$x \in \mathbb{R}, \text{ and } |x - 0| < \delta \implies |f(x) - f(0)| \leq |x| < \delta = \epsilon.$$

Hence f is continuous at 0.

Intermediate value property of a function

Definition

Let I be an interval in $\mathbb{R}$. A function $f : I \to \mathbb{R}$ is said to have the **intermediate value property** (IVP) on I if for any $a, b \in I$ with $a < b$, and $r \in \mathbb{R}$,

$$r \text{ lies between } f(a) \text{ and } f(b) \implies r = f(c) \text{ for some } c \in [a, b].$$

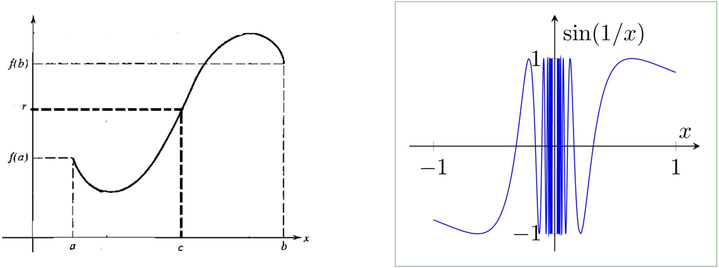


Figure: Illustration of (i) a continuous function having the IVP, and (ii) a discontinuous function having the IVP

Theorem (Intermediate value theorem)

Let I be an interval in $\mathbb{R}$, and let $f : I \to \mathbb{R}$ be a continuous function. Then f has the intermediate value property on I. In particular, the range $J := f(I)$ of f is an interval.

Proof (Sketch): Let $a, b \in I$ with $a < b$. First, suppse $r \in (f(a), f(b))$, Consider $S := \{x \in [a, b] : f(x) < r\}$ and $c := \sup S$. Find sequence (c_n) in S with $c_n \to c$. Use the continuity of f to obtain $f(c) \leq r$ and $a \leq c < b$. Also find (d_n) in (c, b) with $d_n \to c$ to conclude $a < c$ and $f(c) \geq r$.

Examples: (i) Define $f : [0, 1] \to \mathbb{R}$ by $f(0) := 0$ and $f(x) := \sin(1/x)$ if $0 < x \leq 1$. Then f has the IVP. But f is not continuous at 0. So the converse of above theorem is false.

(ii) Define $f : [-1, 2] \to \mathbb{R}$ by $f(x) := 1$ if $x < 0$, $f(0) := 2$ and $f(x) := x$ if $x > 0$. Now $f(-1) = 1 < 1.5 < 2 = f(0)$. But $1.5 \neq f(x)$ for any $x \in [-1, 0]$ (although $f(1.5) = 1.5$). Thus f does not have the IVP. Note: $f([-1, 2]) = (0, 2]$.

An application of the Intermediate Value Theorem

(i) Let $n \in \mathbb{N}$ be odd, and let $p(x) := a_0 + a_1 x + \cdots + a_n x^n$ for $x \in \mathbb{R}$, where $a_0, a_1, \ldots, a_n \in \mathbb{R}$ and $a_n \neq 0$. Then p has at least one (real) root.

Idea behind a proof:

- Suppose $a_n > 0$ without loss of generality.
- If we take $x := b > 0$ large enough, then $p(b) > 0$.
- If we take $x := a < 0$ small enough, then $p(a) < 0$.
- Since p is continuous, it has the IVP on $\mathbb{R}$.
- The number 0 lies between $p(a)$ and $p(b)$.
- There is $c \in \mathbb{R}$ such that $a < c < b$ and $p(c) = 0$.

(ii) The polynomial $p(x) = x^4 + 2x^3 - 2$ has a root in $(0, 1)$: Since p is continuous on [0,1], it has the IVP on $[0, 1]$. Also, $p(0) = -2 < 0$, and $p(1) = 1 > 0$, and so there is $c \in (0, 1)$ such that $p(c) = 0$.

Continuous functions on $[a, b]$

Let $D \subset \mathbb{R}$. A function $f : D \to \mathbb{R}$ is called **bounded** if there is $\alpha \in \mathbb{R}$ such that $|f(x)| \leq \alpha$ for all $x \in D$. We say that a function f **attains its bounds** if there are $x_1, x_2 \in D$ with $f(x_1) = \inf\{f(x) : x \in D\}$ and $f(x_2) = \sup\{f(x) : x \in D\}$.

Theorem (Extreme value property)

Let $[a, b]$ be a closed and bounded interval, and let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is bounded and attains its bounds on $[a, b]$. Thus there are $c_1, c_2 \in [a, b]$ such that

$$\begin{aligned} f(c_1) &= \min\{f(x) : x \in [a, b]\}, \\ f(c_2) &= \max\{f(x) : x \in [a, b]\}. \end{aligned}$$

In fact, $f([a, b]) = [f(c_1), f(c_2)]$, a closed and bounded interval.

Proof: Skipped. See [GL-1] if you are interested.

Examples:

(i) Define $f : (0, 1] \to \mathbb{R}$ by $f(x) := 1/x$ for $x \in (0, 1]$.
Then f is continuous but unbounded.

(ii) Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) := x$ for $x \in \mathbb{R}$.
Then f is continuous but unbounded.

(iii) Define $f : [-1, 1] \to \mathbb{R}$ by $f(x) := 1/x$ if $x \neq 0$ and $f(0) := 0$. Then f is discontinuous and unbounded.

(iv) Define $f : (0, \infty) \to \mathbb{R}$ by $f(x) := 1/(1 + x^2)$.
Then f is continuous and bounded, but it does not attain its bounds. (Note: $\sup\{f(x) : x \in (0, \infty)\} = 1$ and $\inf\{f(x) : x \in (0, \infty)\} = 0$.)

(v) Define $f : [-1, 1] \to \mathbb{R}$ by $f(x) = -1$ if $x < 0$ and $f(x) := 1$ if $x \geq 0$. Then f is bounded and attains its bounds, but f is not continuous on $[-1, 1]$. Thus the converse of the above theorem is false.

Limits of functions of a real variable

Let $D \subset \mathbb{R}$, and let $c \in \mathbb{R}$. Let $f : D \to \mathbb{R}$, and let $\ell \in \mathbb{R}$.

Suppose the values of the function f are as close to the real number ℓ as we like for all x close enough to c (on either side of c), except possibly when $x = c$, then we say that the limit of $f(x)$, as x approaches c, is equal to $\ell \in \mathbb{R}$.

To consider a limit of $f(x)$ as x approaches c, the function f must be defined near the point c (but not necessarily at c).

Definition

Let $D \subset \mathbb{R}$, $f : D \to \mathbb{R}$ and let $c \in \mathbb{R}$ be such that there is $r > 0$ with $(c - r, c) \cup (c, c + r) \subset D$.

We say that $\lim_{x \to c} f(x)$ exists if there is $\ell \in \mathbb{R}$ such that (x_n) is a sequence in D, $x_n \neq c$ and $x_n \to c \implies f(x_n) \to \ell$.

Examples

(i) Consider $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} 3x - 1 & \text{if} \quad x \in \mathbb{R} \text{ and } x \neq 0, \\ 1 & \text{if} \quad x := 0. \end{cases}$$

Let $x_n \to 0$ and $x_n \neq 0$ for $n \in \mathbb{N}$. Then $f(x_n) = 3x_n - 1 \to -1$. Hence $\lim_{x \to 0} f(x) = -1$.

(ii) Let $f(x) := a_0 + a_1 x + \cdots + a_n x^n$ for $x \in \mathbb{R}$, where $n \in \mathbb{N}$ and $a_0, \ldots, a_n \in \mathbb{R}$.

Let $c \in \mathbb{R}$, and let $x_n \to c$, $x_n \neq c$ for all $n \in \mathbb{N}$. Then $f(x_n) \to f(c)$. Hence $\lim_{x \to c} f(x) = \ell$, where $\ell := f(c)$.

Examples (continued)

(iii) Let $f(x) := [x]$ for $x \in \mathbb{R}$.

If $x_n := 1 - (1/n)$ for $n \in \mathbb{N}$, then $x_n \to 1$ and $f(x_n) = 0$ for all $n \in \mathbb{N}$, and so $f(x_n) \to 0$.

If $x_n = 1 + (1/n)$ for $n \in \mathbb{N}$, then $x_n \to 1$ and $f(x_n) = 1$ for all $n \in \mathbb{N}$, and so $f(x_n) \to 1$.

Thus $\lim_{x \to 1} f(x)$ does not exist.

(iv) Let $f(x) := \sin(1/x)$ for $x \in \mathbb{R} \setminus \{0\}$.

If $x_n := 2/(2n+1)\pi$ for $n \in \mathbb{N}$, then $x_n \to 0$, but since $f(x_n) = \sin((2n+1)\pi/2) = (-1)^n$ for all $n \in \mathbb{N}$, we see that $(f(x_n))$ does not converge.

Thus $\lim_{x \to 0} f(x)$ does not exist.

Continuity and Limit

Let us relate the concepts of continuity and limit.

Definition

Let $D \subset \mathbb{R}$. We say that $c \in \mathbb{R}$ is an **interior point** of D if there is $r > 0$ such that $(c - r, c + r) \subset D$.

Theorem

Let c be an interior point of D, and let $f : D \to \mathbb{R}$. Then f is continuous at c $\iff$ $\lim_{x \to c} f(x)$ exists and it is equal to $f(c)$.

Thus for f to be continuous at an interior point c of D,

(i) f must be defined at c,

(ii) $\lim_{x \to c} f(x)$ must exist, and

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

Limit theorems for functions of a real variable

Suppose $\lim_{x\to c} f(x)$ and $\lim_{x\to c} g(x)$ exist. Then

- $\lim_{x\to c}(f \pm g)(x) = \lim_{x\to c} f(x) \pm \lim_{x\to c} g(x)$ resp.,
- $\lim_{x\to c}(f \cdot g)(x) = \left(\lim_{x\to c} f(x)\right) \cdot \left(\lim_{x\to c} g(x)\right)$.
- If $\lim_{x\to c} g(x) \neq 0$, then $\lim_{x\to c} \left(\frac{f}{g}\right)(x) = \dfrac{\lim_{x\to c} f(x)}{\lim_{x\to c} g(x)}$.
- If $f(x) \leq g(x)$ for all $x \neq c$ near c, then $\lim_{x\to c} f(x) \leq \lim_{x\to c} g(x)$.
- (Sandwich theorem)
  If $f(x) \leq h(x) \leq g(x)$ for all x near c and $x \neq c$, and if $\lim_{x\to c} f(x) = \ell = \lim_{x\to c} g(x)$, then $\lim_{x\to c} h(x) = \ell$.

The proofs follow from the limit theorems for sequences.

Examples

(i) Let $0 < |x| < \pi/2$. As we shall see later,

$$-|x| < \sin x < |x| \quad \text{and} \quad -|x| < 1 - \cos x < |x|.$$

Hence $\lim_{x \to 0} \sin x = 0$ and $\lim_{x \to 0} \cos x = 1$ by the sandwich theorem.

(ii) Let $0 < x < \pi/2$. As we shall see later,
$0 < x \cos x < \sin x < x$, and $\sin(-x) = -\sin x$ fo $x \in \mathbb{R}$.
Hence $\displaystyle\lim_{x \to 0} \frac{\sin x}{x} = 1$ by the sandwich theorem.

(iii) Let $f(x) := x \sin\left(\frac{1}{x}\right)$ for $x \in \mathbb{R} \setminus \{0\}$.

Then $-|x| \leq x \sin\left(\frac{1}{x}\right) \leq |x|$ for all $x \in \mathbb{R} \setminus \{0\}$.

Hence $\displaystyle\lim_{x \to 0} x \sin\left(\frac{1}{x}\right) = 0$ by the sandwich theorem.

Right hand and left hand limits

Definition

Suppose $f : D \to \mathbb{R}$ and $c \in \mathbb{R}$ are such that $(c, c+r) \subset D$ for some $r > 0$. We say that $\lim\limits_{x \to c^+} f(x)$ exists if there is $\ell_1 \in \mathbb{R}$ such that

(x_n) a sequence in D, $x_n \to c$ and $x_n > c \implies f(x_n) \to \ell_1$.

Definition

Suppose $f : D \to \mathbb{R}$ and $c \in \mathbb{R}$ are such that $(c-r, c) \subset D$ for some $r > 0$. We say that $\lim\limits_{x \to c^-} f(x)$ exists if there is $\ell_2 \in \mathbb{R}$ such that

(x_n) a sequence in D, $x_n \to c$ and $x_n < c \implies f(x_n) \to \ell_2$.

Example: Let $f(x) := [x]$ for $x \in \mathbb{R}$.

Then $\lim_{x \to 1^-} f(x) = 0$ and $\lim_{x \to 1^+} f(x) = 1$.

Theorem

Let $f : D \to \mathbb{R}$ and let $c \in \mathbb{R}$ such that there is $r > 0$ with

$$(c - r, c) \cup (c, c + r) \subset D.$$

$\lim_{x \to c} f(x)$ exists $\iff$ $\lim_{x \to c^+} f(x)$ and $\lim_{x \to c^-} f(x)$ exist and are equal. In this case,

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

Equivalent condition for the existence of a limit

Recall:

(Continuity)

Let $D \subset \mathbb{R}$, $f : D \to \mathbb{R}$ and $c \in D$. Then f is continuous at c $\iff$ the following ϵ-δ condition holds:
For every $\epsilon > 0$, there exists $\delta > 0$ such that

$$x \in D \text{ and } |x - c| < \delta \implies |f(x) - f(c)| < \epsilon.$$

Similarly,

(Limit)

Let $D \subset \mathbb{R}$, $f : D \to \mathbb{R}$ and $c \in \mathbb{R}$ be such that there is $r > 0$ with $(c - r, c) \cup (c, c + r) \subset D$. Let $\ell \in \mathbb{R}$. Then $\lim\limits_{x \to c} f(x) = \ell \iff$ the following ϵ-δ condition holds:
For every $\epsilon > 0$, there is $\delta > 0$ such that

$$x \in D \text{ and } 0 < |x - c| < \delta \implies |f(x) - \ell| < \epsilon.$$

Example: Let

$$f(x) := \frac{x}{|x|} \quad \text{for } x \in \mathbb{R} \setminus \{0\}.$$

Then there is no $\ell \in \mathbb{R}$ such that $\lim_{x \to 0} f(x) = \ell$.

For $n \in \mathbb{N}$, let $x_n := (-1)^n/n$. Then $x_n \to 0$, but $(f(x_n))$ is divergent, since $f(x_n) = (-1)^n$ for $n \in \mathbb{N}$. Hence the result.

Aliter: Suppose there is $\ell \in \mathbb{R}$ such that $\lim_{x \to 0} f(x) = \ell$. Let $\epsilon := 1/2$. Then there is $\delta > 0$ such that $|f(x) - \ell| < 1/2$ for all $x \in \mathbb{R}$ satisfying $0 < |x| < \delta$.
Let $x_0 := \delta/2$. Then $f(x_0) = 1$ and $f(-x_0) = -1$. Hence

$$\begin{aligned} 2 &= |f(x_0) - f(-x_0)| \le |f(x_0) - \ell| + |\ell - f(-x_0)| \\ &< \frac{1}{2} + \frac{1}{2} = 1, \quad \text{which is a contradiction.} \end{aligned}$$

Differentiability

Definition

Let $D \subset \mathbb{R}$, and let c be an interior point of D. A function $f : D \to \mathbb{R}$ is said to be **differentiable** at c if

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

exists. In this case, this limit is denoted by $f'(c)$, and it is called the **derivative** of f at c.

Geometrically speaking, $f'(c)$ is equal to the slope of the tangent to the curve $y = f(x)$ at $x = c$, and so the curve is 'smooth' at c; there is no corner or cusp at c.

Also, $f'(c)$ can be interpreted as the rate of change in f at c.

Examples:

(i) If $f : \mathbb{R} \to \mathbb{R}$ is a **constant function**, then f is differentiable and $f'(c) = 0$ for all $c \in \mathbb{R}$.

(ii) If $f : \mathbb{R} \to \mathbb{R}$ is the **identity function** $f(x) := x$ for $x \in \mathbb{R}$, then f is differentiable and $f'(c) = 1$ for all $c \in \mathbb{R}$.

(iii) Let $f(x) := x^{2/3}$ for $x \in \mathbb{R}$. Then f is not differentiable at 0 since $\dfrac{f(0+h) - f(0)}{h} = \dfrac{1}{h^{1/3}}$ for $h \neq 0$.

(iv) Let $f(x) := \sin x$ for $x \in \mathbb{R}$. Then f is differentiable at 0 since $\displaystyle\lim_{h\to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0} \frac{\sin h - \sin 0}{h - 0} = 1.$

(v) Let $f(0) := 0$ and $f(x) := x \sin(1/x)$ for $x \in \mathbb{R} \setminus \{0\}$. Then f is not differentiable at 0 since

$$\frac{f(0+h) - f(0)}{h} = \sin\left(\frac{1}{h}\right) \quad \text{for } h \neq 0.$$

Right hand and left hand derivatives

Let $D \subset \mathbb{R}$, and let $f : D \to \mathbb{R}$.

- Suppose $c \in D$ is such that $[c, c + r) \subset D$ for some $r > 0$. If the limit
$$\lim_{h \to 0^+} \frac{f(c + h) - f(c)}{h}$$
exists, then it is called the **right hand derivative** of f at c, and we denote it by $f'_+(c)$.
- The **left hand derivative** of f at c is defined similarly, and we denote it by $f'_-(c)$.
- If c is an interior point of D, then f is differentiable at $c \iff f'_+(c)$ and $f'_-(c)$ both exist and are equal.

Example: Let $f(x) := |x|$ for $x \in \mathbb{R}$. Then $f'_-(0) = -1$ and $f'_+(0) = 1$. Hence f is not differentiable at 0.

MA 105 : Calculus
Division 1, Lecture 04

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Equivalent definition of continuity (in terms of - condition)
- Intermediate Value Property (IVP). Examples
- Intermediate Value Theorem. Application
- Extreme Value Property (of continuous functions defined on closed and bounded intervals), Examples
- Limits of functions of a real variable. Examples
- Continuity and Limit
- Limit theorems for functions of a real variable
- Right hand and left hand limits
- Equivalent condition for the existence of a limit
- Differentiability: Definition and Examples
- Right hand and left hand derivatives

Differentiability

Definition

Let $D \subset \mathbb{R}$, and let c be an interior point of D. A function $f : D \to \mathbb{R}$ is said to be **differentiable** at c if

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

exists. In this case, this limit is denoted by $f'(c)$, and it is called the **derivative** of f at c.

Geometrically speaking, $f'(c)$ is equal to the slope of the tangent to the curve $y = f(x)$ at $x = c$, and so the curve is 'smooth' at c; there is no corner or cusp at c.

Also, $f'(c)$ can be interpreted as the rate of change in f at c.

Examples:

(i) If $f : \mathbb{R} \to \mathbb{R}$ is a **constant function**, then f is differentiable and $f'(c) = 0$ for all $c \in \mathbb{R}$.

(ii) If $f : \mathbb{R} \to \mathbb{R}$ is the **identity function** $f(x) := x$ for $x \in \mathbb{R}$, then f is differentiable and $f'(c) = 1$ for all $c \in \mathbb{R}$.

(iii) Let $f(x) := x^{2/3}$ for $x \in \mathbb{R}$. Then f is not differentiable at 0 since $\dfrac{f(0+h) - f(0)}{h} = \dfrac{1}{h^{1/3}}$ for $h \neq 0$.

(iv) Let $f(x) := \sin x$ for $x \in \mathbb{R}$. Then f is differentiable at 0 since $\displaystyle\lim_{h\to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h\to 0} \frac{\sin h - \sin 0}{h - 0} = 1$.

(v) Let $f(0) := 0$ and $f(x) := x \sin(1/x)$ for $x \in \mathbb{R} \setminus \{0\}$. Then f is not differentiable at 0 since

$$\frac{f(0+h) - f(0)}{h} = \sin\left(\frac{1}{h}\right) \quad \text{for } h \neq 0.$$

Right hand and left hand derivatives

Let $D \subset \mathbb{R}$, and let $f : D \to \mathbb{R}$.

- Suppose $c \in D$ is such that $[c, c + r) \subset D$ for some $r > 0$. If the limit

$$\lim_{h \to 0^+} \frac{f(c + h) - f(c)}{h}$$

  exists, then it is called the **right hand derivative** of f at c, and we denote it by $f'_+(c)$.
- The **left hand derivative** of f at c is defined similarly, and we denote it by $f'_-(c)$.
- If c is an interior point of D, then f is differentiable at $c \iff f'_+(c)$ and $f'_-(c)$ both exist and are equal.

Example: Let $f(x) := |x|$ for $x \in \mathbb{R}$. Then $f'_-(0) = -1$ and $f'_+(0) = 1$. Hence f is not differentiable at 0.

- If I is a nonempty open interval in $\mathbb{R}$, that is, if $I := (a, b), (a, \infty), (-\infty, b)$ or $\mathbb{R}$, then we say that a function $f : I \to \mathbb{R}$ is **differentiable on** I if f is differentiable at every $c \in I$.
- If $I := [a, b]$, then we say that $f : I \to \mathbb{R}$ is **differentiable on** I if f is differentiable on (a, b), and if both $f'_+(a)$ and $f'_-(b)$ exist.
- Similarly, we define the differentiability of a function on $I := [a, b), (a, b], [a, \infty)$ or $(-\infty, b]$.

In these cases, there is a new function $f' : I \to \mathbb{R}$, called the **derivative (function)** of f defined on I.
Let $c \in I$. If $g := f'$ is differentiable at c, then we say that f is **twice differentiable** at c, and write $f''(c) := g'(c)$.

$f''(c)$ is called the **second derivative** of f at c. Also, for $n \in \mathbb{N}$, we denote the n**th derivative** of f at c by $f^{(n)}(c)$.

Equivalent condition for differentiability

Let $D \subset \mathbb{R}$, let $f : D \to \mathbb{R}$, and let c be an interior point of D.

Recall that f is said to be differentiable at c if the limit

$$f'(c) := \lim_{h \to 0} \frac{f(c+h) - f(c)}{h} \quad \text{exists.}$$

(Carathéodory Lemma: C-lemma)

Let $f : D \to \mathbb{R}$, and c be an interior point of D. Then f is differentiable at $c \iff$ there is a function $f_1 : D \to \mathbb{R}$ which is continuous at c and satisfies

$$f(x) - f(c) = (x - c) f_1(x) \quad \text{for all } x \in D.$$

In this case, the function f_1 is unique and $f'(c) = f_1(c)$.

The function $f_1 : D \to \mathbb{R}$ is called the **increment function** associated with f and c.

Proof: Suppose f is differentiable at c. Define

$$f_1(x) := \begin{cases} \dfrac{f(x) - f(c)}{x - c} & \text{if } x \in D \setminus \{c\}, \\ f'(c) & \text{if } x = c. \end{cases}$$

Then f_1 is continuous at c since $\lim_{x \to c} f_1(x) = f'(c) = f_1(c)$.

Conversely, suppose there is $f_1 : D \to \mathbb{R}$ as stated. Let $h := x - c$ for $x \in D$, and so $c + h = x$. Then

$$\lim_{h \to 0} \frac{f(c+h) - f(c)}{h} = \lim_{x \to c} \frac{f(x) - f(c)}{x - c} = \lim_{h \to 0} f_1(x) = f_1(c),$$

since f_1 is continuous at c. Hence f is differentiable at c. The uniqueness of the increment function f_1 is obvious. $\square$

If D is an interval and c is an end-point of D, then the C-lemma is valid at c.

The C-lemma relates the concepts of differentiability and continuity. It enables us to prove the following results neatly.

Differentiability $\implies$ Continuity

Theorem

Suppose $f : D \to \mathbb{R}$ is differentiable at an interior point c of D. Then f is continuous at c.

Proof.

Let f_1 be the increment function associated with f and c. Then $f(x) = f(c) + (x - c)f_1(x)$ for $x \in I$. Since $f_1 : I \to \mathbb{R}$ is continuous at c, so is f. $\square$

Remarks:

1. If a function f is not continuous at an interior point c of D, then it cannot be differentiable at c.
Example: Let $f(x) := [x]$ for $x \in \mathbb{R}$, and $c := 1$.

2. The converse of the above theorem is false.
Example: $f(x) = |x|$ for $x \in \mathbb{R}$, and $c := 0$.

Differentiability: Algebraic rules

Let f and g be differentiable at c. Then $f \pm g$ and $f \cdot g$ are differentiable at c, and

$$(f \pm g)'(c) = f'(c) \pm g'(c) \text{ resp.,}$$

$$(f \cdot g)'(c) = f'(c)g(c) + f(c)g'(c).$$

Further, if $g(c) \neq 0$, then f/g is differentiable at c, and

$$\left(\frac{f}{g}\right)'(c) = \frac{f'(c)g(c) - f(c)g'(c)}{g(c)^2}.$$

Note: If f is differentiable at c, then f^n is differentiable at c for each $n \in \mathbb{N}$, and $(f^n)'(c) = n\,f(c)^{n-1}f'(c)$. This also holds for a negative integer n if $f(c) \neq 0$.

Examples: Every polynomial function p is differentiable on $\mathbb{R}$. A rational function p/q is differentiable at c if $q(c) \neq 0$.

Differentiability: Chain Rule

Theorem

Let $D \subset \mathbb{R}$ and $f : D \to \mathbb{R}$. Also, let $E \subset \mathbb{R}$ and $g : E \to \mathbb{R}$ be such that $f(D) \subset E$. Suppose c is an interior point of D and $f(c)$ is an interior point of E. If f is differentiable at c, and if g is differentiable at $f(c)$, then $g \circ f : D \to \mathbb{R}$ is differentiable at c, and $(g \circ f)'(c) = g'(f(c))f'(c)$.

Proof. Let f_1 be the increment functions associated with f and c, and let g_1 be the increment functions associated with g and $d := f(c)$. Then $f(x) - f(c) = (x - c)f_1(x)$ for all $x \in D$, where f_1 is continuous at c, and $g(y) - g(d) = (y - d)g_1(y)$ for all $y \in E$, where g_1 is continuous at d.

Define $h := g \circ f : D \to \mathbb{R}$. Then for $x \in D$,

$$\begin{aligned} h(x) - h(c) &= g(f(x)) - g(f(c)) \\ &= (f(x) - f(c))g_1(f(x)) = (x - c)f_1(x)(g_1 \circ f)(x). \end{aligned}$$

Let $h_1 := f_1 \cdot (g_1 \circ f) : D \to \mathbb{R}$. Then h_1 is continuous at c since f and f_1 are continuous at c and g_1 is continuous at $d = f(c)$. It follows that h_1 is the increment function associated with $g \circ f$ and c, and

$$(g \circ f)'(c) = h_1(c) = g'(f(c))f'(c). \qquad \square$$

<u>Conclusion of the Chain Rule in the Leibnitz notation:</u>

If $y := f(x)$ and $z := g(y)$, then $z = (g \circ f)(x)$ and

$$\left.\frac{dz}{dx}\right|_{x=c} = \left.\frac{dz}{dy}\right|_{y=f(c)} \cdot \left.\frac{dy}{dx}\right|_{x=c}$$

Warning:

Cancelling out dy in the RHS above to obtain the LHS does not yield a proof of the Chain Rule!

Example: Consider

Let

$$f(x) := \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if} \quad x \in \mathbb{R} \setminus \{0\}, \\ 0 & \text{if} \quad x := 0. \end{cases}$$

$$f_1(x) := \begin{cases} x \sin\left(\frac{1}{x}\right) & \text{if} \quad x \in \mathbb{R} \setminus \{0\}, \\ 0 & \text{if} \quad x := 0. \end{cases}$$

Then f_1 is continuous at 0, and $f(x) - f(0) = (x - 0)f_1(x)$ for all $x \in \mathbb{R}$. By the C-lemma, f is differentiable at 0, and $f'(0) = f_1(0) = 0$.
Also, if $c \neq 0$, then f is differentiable at c, and $f'(c) = 2c \sin\frac{1}{c} - \cos\frac{1}{c}$. (Multiplication rule and chain rule.)

Note: The derivative f' of f is not continuous at 0. Why?

Differentiability of the inverse function

Theorem

Let I be an interval, and let c be an interior point of I. If $f : I \to \mathbb{R}$ is a one-one continuous function such that f is differentiable at c and $f'(c) \neq 0$, then $f^{-1} : f(I) \to \mathbb{R}$ is differentiable at $f(c)$ and

$$\left(f^{-1}\right)'(f(c)) = \frac{1}{f'(c)}.$$

Proof: Skipped. You may try, if interested. Or look up [GL-1].

Note: Since $f : I \to \mathbb{R}$ is continuous, it has the IVP, and so $J = f(I)$ is an interval. Hence $f(c)$ is either an interior point of J or an end-point of J. Since f is also one-one, it can be shown to be strictly monotonic (to be defined later), and so $f(c)$ is in fact an interior point of J.

Examples of derivatives of inverse functions

(i) Let $n \in \mathbb{N}$, and let $f(x) := x^n$ for $x \in (0, \infty)$. Then f is one-one and continuous. Consider $c \in (0, \infty)$. Now $f'(c) = n\, c^{n-1} \neq 0$. Further, $f\big((0,\infty)\big) = (0, \infty)$. If $d \in (0, \infty)$ and $f(c) = c^n = d$, then $c = d^{1/n}$ and

$$\left(f^{-1}\right)'(d) = \frac{1}{f'(c)} = \frac{1}{n\, c^{n-1}} = \frac{1}{n\, d^{(n-1)/n}} = \frac{1}{n} d^{(1/n)-1}.$$

(ii) Let $f(x) := \sin x$ for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Then f is one-one and continuous. Consider $c \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Now $f'(c) = \cos c \neq 0$. Further, $f\big((-\frac{\pi}{2}, \frac{\pi}{2})\big) = (-1, 1)$. If $d \in (-1, 1)$ and $f(c) = \sin c = d$, then

$$\left(f^{-1}\right)'(d) = \frac{1}{f'(c)} = \frac{1}{\cos c} = \frac{1}{\sqrt{1 - \sin^2 c}} = \frac{1}{\sqrt{1 - d^2}}.$$

Local Extremum

Let $D \subseteq \mathbb{R}$, and let c be an interior point of D.
We say that a function $f : D \to \mathbb{R}$ has

(i) a **local minimum** at c if there is $\delta > 0$ such that $(c - \delta, c + \delta) \subseteq D$ and

$$f(x) \geq f(c) \quad \text{for all } x \in (c - \delta, c + \delta).$$

(ii) a **local maximum** at c if there is $\delta > 0$ such that $(c - \delta, c + \delta) \subseteq D$ and

$$f(x) \leq f(c) \quad \text{for all } x \in (c - \delta, c + \delta).$$

(iii) a **local extremum** at c if it has a local maximum or a local minimum at c.

Proposition

Let $D \subseteq \mathbb{R}$, and let c be an interior point of D. If $f : D \to \mathbb{R}$ has a local extremum at c, and if f is differentiable at c, then $f'(c) = 0$.

Proof.

Let f be differentiable at c. By the C-lemma, there is $f_1 : D \to \mathbb{R}$ such that f_1 is continuous at c and $f(x) - f(c) = (x - c)f_1(x)$ for all $x \in D$.

Let f have a local minimum at c. Then there is $\delta > 0$ such that $f(x) \geq f(c)$ for all $x \in (c - \delta, c + \delta)$. Hence $f_1(x) \geq 0$ for all $x \in (c, c + \delta)$ and $f_1(x) \leq 0$ for all $x \in (c - \delta, c)$.

Now $0 \leq \lim\limits_{x \to c^+} f_1(x) = f_1(c) = \lim\limits_{x \to c^-} f_1(x) \leq 0$, since f_1 is continuous at c. Hence $f'(c) = f_1(c) = 0$.

A similar argument holds if f has a local maximum at c. $\square$

Examples:

(i) Let $D := [0, 1]$, and $f(x) := x$ for $x \in D$. Then $f(0) = \min\{f(x) : x \in D\}$, $f(1) = \max\{f(x) : x \in D\}$. But $f'_{+}(0) = 1 \neq 0$ and $f'_{-}(1) = 1 \neq 0$. Note: Neither 0 nor 1 is an interior point of D.

(ii) Let $D := [-1, 1]$, and $f(x) = |x|$ for $x \in D$. Then f has a local minimum at the interior point 0 of D. But $f'_{+}(0) = 1 \neq 0$ and $f'_{-}(1) = -1 \neq 0$. Note: f is not differentiable at 0.

(iii) Let $D := [-1, 1]$, and $f(x) := x^2$ for $x \in D$. Then f has a local minimum at the interior point 0 of D, and $f'(0) = 0$.

(iv) Let $D := [-1, 1]$, and $f(x) := x^3$ for $x \in D$. Then 0 is an interior point of D and $f'(0) = 0$. But f does not have a local extremum at 0. Thus the converse of the above proposition is false.

Rolle's Theorem

Theorem

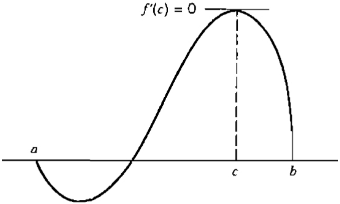
Let $a < b$ and $f : [a, b] \to \mathbb{R}$ be a function such that

(i) f is continuous on $[a, b]$,

(ii) f is differentiable on (a, b), and

(iii) $f(a) = f(b)$.

Then there exists $c \in (a, b)$ such that $f'(c) = 0$.



Thus there exists a point $c \in (a, b)$ such that the tangent to the graph at $(c, f(c))$ is parallel to the x-axis.

MA 105 : Calculus
Division 1, Lecture 05

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Higher order derivatives
- Equivalent condition for Differentiability: C-Lemma
- Consequences of the Carathéodory Lemma (C-Lemma):
    - Differentiability implies Continuity
    - Algebraic Rules (concerning sums, products, and quotients of differentiable functions)
    - Chain Rule.
    - Differentiability of the inverse function. Examples
- Local extremum. Basic properties.
- Rolle's theorem.

Rolle's Theorem

Theorem

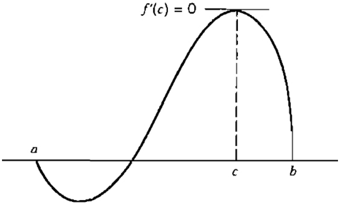
Let $a < b$ and $f : [a, b] \to \mathbb{R}$ be a function such that

(i) f is continuous on $[a, b]$,

(ii) f is differentiable on (a, b), and

(iii) $f(a) = f(b)$.

Then there exists $c \in (a, b)$ such that $f'(c) = 0$.



Thus there exists a point $c \in (a, b)$ such that the tangent to the graph at $(c, f(c))$ is parallel to the x-axis.

Proof. Since f is continuous on $[a, b]$, it is bounded and attains its bounds on $[a, b]$. Let $c_1, c_2 \in [a, b]$ be such that

$$f(c_1) = \min\{f(x) : x \in [a, b]\} \text{ and } f(c_2) = \max\{f(x) : x \in [a, b]\}.$$

If c_1 or c_2 is an interior point of $[a, b]$, then $f'(c_1) = 0$ or $f'(c_2) = 0$ by the previous proposition, and we are done.

If both c_1 and c_2 are end-points of $[a, b]$, then $f(c_1) = f(a) = f(b) = f(c_2)$. Since the minimum and the maximum values of f on $[a, b]$ are the same, f is a constant function on $[a, b]$, and so $f'(c) = 0$ for every $c \in (a, b)$. $\square$

Example:
Let $f(x) := |x|$ for $x \in [-1, 1]$. Then f is continuous on $[-1, 1]$ and $f(-1) = 1 = f(1)$. But there is no $c \in (-1, 1)$ such that $f'(c) = 0$. Note: f is not differentiable at 0.

Application of the Rolle theorem

Let $f(x) := x^4 + 2x^3 - 2$ for $x \in \mathbb{R}$. We show that f has a unique root in $(0, \infty)$, and it is in $(0, 1)$.

- Using the IVP, we have seen that f has a root in (0,1).
- Suppose there are two roots in $(0, \infty)$. Then there are $a, b \in \mathbb{R}$ with $0 < a < b$ such that $f(a) = 0 = f(b)$.
- By Rolle's theorem, there is $c \in (a, b) \subseteq (0, \infty)$ such that $f'(c) = 0$.
- But $f'(x) = 4x^3 + 6x^2 = 2x^2(2x + 3) \neq 0$ for $x \in (0, \infty)$.
- Hence f has exactly one (real) root in $(0, \infty)$.

Lagrange's Mean Value Theorem

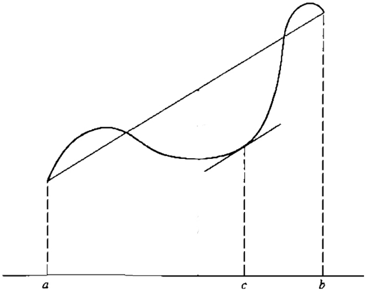
Theorem (MVT)

Let $a < b$ and $f : [a, b] \to \mathbb{R}$ be a function such that

(i) f is continuous on $[a, b]$, and

(ii) f is differentiable on (a, b).

Then there exists $c \in (a, b)$ such that $f'(c) = \dfrac{f(b) - f(a)}{b - a}$.



Thus there exists $c \in (a, b)$ such that the tangent to the graph at $(c, f(c))$ is parallel to the line joining $(a, f(a))$ and $(b, f(b))$.

Proof. For $x \in [a, b]$, define

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

Then F is continuous on $[a, b]$, differentiable on (a, b) and $F(a) = f(a) = F(b)$. By Rolle's theorem, there is $c \in (a, b)$ such that $F'(c) = 0$, that is, $f'(c) = \dfrac{f(b) - f(a)}{b - a}$. $\square$

Note: We have used the Rolle theorem to prove a more general result, namely the Lagrange mean value theorem.

Physical Interpretation of MVT: For $t \in [a, b]$, let $f(t)$ denote the distance travelled by a particle at time t. Then the average speed of the particle is $\dfrac{f(b) - f(a)}{b - a}$, and its instantaneous speed at time $t \in [a, b]$ is $f'(t)$. Thus there is a time $c \in [a, b]$ at which the average speed equals the instantaneous speed.

Consequences of the Mean Value Theorem (MVT)

Proposition

Let I be an interval and $f : I \to \mathbb{R}$. If $f'(x)$ exists and equals 0 for all $x \in I$, then f is constant on I.

Proof. Let x_1, x_2 in I with $x_1 < x_2$. Then $[x_1, x_2] \subseteq I$.
By the MVT,

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) \quad \text{for some } c \in (x_1, x_2).$$

Since $f'(c) = 0$, we obtain $f(x_2) = f(x_1)$. Thus f has the same value at every point of I. $\square$

We have already seen that the converse is also true.

Example: Let $D := \mathbb{R} \setminus \{0\}$, and $f(x) := x/|x|$ for $x \in D$. Then $f' = 0$ on D, but f is not constant on D; this is because D is not an interval.

Consequences of MVT (continued)

Monotonicity

Let I be an interval, and let $f : I \to \mathbb{R}$. We say that f is

- **(monotonically) increasing** on I if
  $x_1, x_2 \in I,\ x_1 < x_2 \implies f(x_1) \leq f(x_2).$
- **(monotonically) decreasing** on I if
  $x_1, x_2 \in I,\ x_1 < x_2 \implies f(x_1) \geq f(x_2).$
- **monotonic** on I if it is increasing on I, or it is decreasing on I.
- **strictly increasing** on I if
  $x_1, x_2 \in I,\ x_1 < x_2 \implies f(x_1) < f(x_2).$
- **strictly decreasing** on I if
  $x_1, x_2 \in I,\ x_1 < x_2 \implies f(x_1) > f(x_2).$
- **strictly monotonic** on I if either it is strictly increasing on I, or it is strictly decreasing on I.

Derivatives and Monotonicity

Proposition

Let I be an interval, and let $f : I \to \mathbb{R}$ be differentiable. Then

(i) $f' \geq 0$ on $I \iff f$ is increasing on I.

(ii) $f' \leq 0$ on $I \iff f$ is decreasing on I.

(iii) $f' > 0$ on $I \implies f$ is strictly increasing on I.

(iv) $f' < 0$ on $I \implies f$ is strictly decreasing on I.

All forward ($\implies$) implications follow from the MVT.

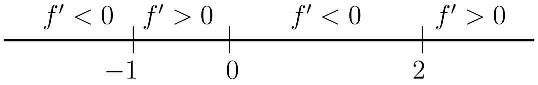
The backward ($\impliedby$) implication in (i) follows since $\big(f(x+h) - f(x)\big)/h \geq 0$ for all $h \neq 0$ with $x + h \in I$.
Similarly, the backward ($\impliedby$) implication in (ii) follows.

The converse of (iii) is false: Let $f(x) := x^3$ for $x \in \mathbb{R}$. Then f is strictly increasing on $\mathbb{R}$, but $f'(0) = 0$.
Similarly, the converse of (iv) is false.

Examples

(i) Let $f(x) := 3x^4 - 4x^3 - 12x^2 + 5$ for $x \in \mathbb{R}$. Let us find intervals on which f is (strictly) increasing, or (strictly) decreasing. For $x \in \mathbb{R}$,

$$\begin{aligned} f'(x) &= 12x^3 - 12x^2 - 24x = 12x(x^2 - x - 2) \\ &= 12(x+1)x(x-2). \end{aligned}$$



- f is strictly decreasing on $(-\infty, -1)$.
- f is strictly increasing on $(-1, 0)$.
- f is strictly decreasing on $(0, 2)$.
- f is strictly increasing on $(2, \infty)$.

(ii) Let $f(x) := x - \sin x$ for $x \in [0, \pi/2]$. Then $f'(x) = 1 - \cos x \geq 0$ for all $x \in [0, \pi/2]$. Hence f is increasing on $[0, \pi/2]$, and so $0 = f(0) \leq f(x)$ for $x \in [0, \pi/2]$. Thus $x \in [0, \pi/2] \implies \sin x \leq x$. In fact, $\sin x < x$ for $x \in (0, \pi/2]$ since $f'(x) > 0$ for $x \in (0, \pi/2]$.

(iii) Let $g(x) := x - 1 + \cos x$ for $x \in [0, \pi/2]$. Then $g'(x) = 1 - \sin x \geq 0$ for all $x \in [0, \pi/2]$. Hence g is increasing on $[0, \pi/2]$, and so $0 = g(0) \leq g(x)$ for $x \in [0, \pi/2]$. Thus $x \in [0, \pi/2] \implies 1 - \cos x \leq x$. In fact, since $g'(x) > 0$ for $x \in [0, \pi/2)$, we see that $g(0) < g(x)$ for $x \in (0, \pi/2)$, that is, $1 - \cos x < x$ for $x \in (0, \pi/2)$.

(iv) Let $h(x) := \sin x - x \cos x$ for $x \in [0, \pi/2]$. Then $h'(x) = x \sin x \geq 0$ for all $x \in [0, \pi/2]$. Hence h is increasing on $[0, \pi/2]$, and so $0 = h(0) \leq h(x)$ for $x \in [0, \pi/2]$. Thus $x \in [0, \pi/2] \implies x \cos x \leq \sin x$. In fact, $x \cos x < \sin x$ for $x \in (0, \pi/2]$ since $h'(x) > 0$ for $x \in (0, \pi/2]$.

The above 3 results yield the inequalities quoted in Lecture 3.

Convexity and Concavity

Let I be an interval. A function $f : I \to \mathbb{R}$ is called **convex** on I if the line segment joining any two points on the graph of f lies on or above the graph of f. If $x_1, x_2 \in I$, then the equation of the line passing through $(x_1, f(x_1))$ and $(x_2, f(x_2))$ is given by $y = f(x_1) + \dfrac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1)$. Thus f is **convex** on I if

$$x_1, x_2, \ x \in I, \ x_1 < x < x_2 \Longrightarrow f(x) \leq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1).$$

Similarly, a function $f : I \to \mathbb{R}$ is called **concave** on I if

$$x_1, x_2, \ x \in I, \ x_1 < x < x_2 \Longrightarrow f(x) \geq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1).$$

A function $f : I \to \mathbb{R}$ is called **strictly convex** on I if the inequality $\leq$ in the definition of a convex function can be replaced by the strict inequality $<$.

A function $f : I \to \mathbb{R}$ is called **strictly concave** on I if the inequality $\geq$ in the definition of a concave function can be replaced by the strict inequality $>$.

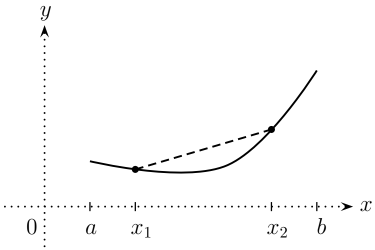
Examples:

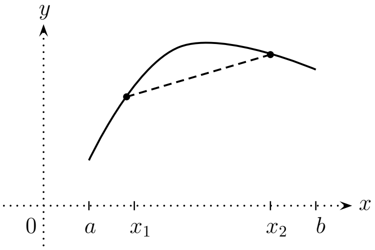
(i) Let $\alpha, \beta \in \mathbb{R}$, and let $f(x) := \alpha + \beta\, x$ for $x \in \mathbb{R}$. Then f is convex as well as concave on $\mathbb{R}$ (but not strictly).

(ii) Let $\alpha \in \mathbb{R}$, and let $f(x) := \alpha\, |x|$ for $x \in \mathbb{R}$. Then f is convex on $\mathbb{R}$ if $\alpha > 0$, and f is concave on $\mathbb{R}$ if $\alpha < 0$ (but not strictly). (Prove this from the first principles.)

(iii) Let $\alpha \in \mathbb{R}$, and let $f(x) := \alpha\, x^2$ for $x \in \mathbb{R}$. Then f is strictly convex on $\mathbb{R}$ if $\alpha > 0$, and f is strictly concave on $\mathbb{R}$ if $\alpha < 0$. (Prove this from the first principles.)

Typical graphs of convex and concave functions on $I := [a, b]$ look as follows:





Let I be an interval, and let $f : I \to \mathbb{R}$ be a function. Suppose the curve $y = f(x)$ has a (unique) nonvertical tangent at each point on it.

It is apparent that if f is convex, then the slopes of the tangents increase as we move from left to right, and if f is concave, then then the slopes of the tangents decrease as we move from left to right. This is made precise in the following.

Proposition

Let I be an interval and $f : I \to \mathbb{R}$ be differentiable. Then

(i) f' is increasing on $I \iff f$ is convex on I.

(ii) f' is decreasing on $I \iff f$ is concave on I.

(iii) f' is strictly increasing on $I \iff f$ is strictly convex on I.

(iv) f' is strictly decreasing on $I \iff f$ is strictly concave on I.

We omit proofs of the above results as they are a bit involved.

Corollary

Let I be an interval and $f : I \to \mathbb{R}$ be twice differentiable.

(i) $f'' \geq 0$ on $I \iff f$ is convex on I.

(ii) $f'' \leq 0$ on $I \iff f$ is concave on I.

(iii) $f'' > 0$ on $I \implies f$ is strictly convex on I.

(iv) $f'' < 0$ on $I \implies f$ is strictly concave on I.

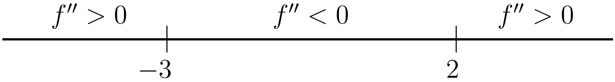
The converse of (iii) in the above corollary is false. To see this, consider $f(x) := x^4$ for $x \in \mathbb{R}$. Then f is strictly convex on $\mathbb{R}$, but $f''(0) = 0$. Similarly, the converse of (iv) is false.

Example:

Let $f(x) := x^4 + 2x^3 - 36x^2 + 62x + 5$ for $x \in \mathbb{R}$. Let us find intervals on which f is (strictly) convex or (strictly) concave. For $x \in \mathbb{R}$,

$$\begin{aligned} f'(x) &= 4x^3 + 6x^2 - 72x + 62, \text{ and} \\ f''(x) &= 12x^2 + 12x - 72 = 12(x+3)(x-2). \end{aligned}$$

Note:



- f is strictly convex on $(-\infty, -3)$.
- f is strictly concave on $(-3, 2)$.
- f is strictly convex on $(2, \infty)$.

Critical Points and Global Extrema

Let $D \subset \mathbb{R}$, and let $f : D \to \mathbb{R}$. An interior point c of D is called a **critical point** of f if either f is not differentiable at c, or if f is differentiable at c and $f'(c) = 0$.

Proposition

Let $f : [a, b] \to \mathbb{R}$ be continuous. Then the global minimum $m := \min\{f(x) : x \in [a, b]\}$ as well as the global maximum $M := \max\{f(x) : x \in [a, b]\}$ of f on $[a, b]$ is attained either at a critical point of f or at an end-point of $[a, b]$.

Proof. There is $c_1 \in [a, b]$ such that $f(c_1) = m$. If $c_1 = a$ or $c_1 = b$, then we are done. Next, let $c_1 \in (a, b)$. Then f has a local minimum at c_1. If f is not differentiable at c_1, then c_1 is a critical point of f. If f is differentiable at c_1, then $f'(c_1) = 0$ by an earlier result, and so c_1 is a critical point of f. Similarly, we argue for the global maximum M of f. □

Finding Global Extrema

Example: Consider $f : [-1, 2] \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} -x & \text{if} \quad -1 \le x \le 0, \\ 2x^3 - 4x^2 + 2x & \text{if} \quad 0 < x \le 2. \end{cases}$$

f is continuous on $[-1, 2]$: $\lim\limits_{x \to 0^-} f(x) = f(0) = \lim\limits_{x \to 0^+} f(x)$.
f is not differentiable at 0 : $f'_-(0) = -1$ and $f'_+(0) = 2$.
For $-1 < x < 0$, $f'(x) = -1 \neq 0$.
For $0 < x < 2$, $f'(x) = 6x^2 - 8x + 2 = 2(3x - 1)(x - 1)$.
Critical points of f: 0, $1/3$, 1. End-points of $[-1, 2]$: -1, 2.

x	-1	0	$1/3$	1	2
$f(x)$	1	0	$8/27$	0	4

f attains its global maximum 4 at $x = 2$, and its global minimum 0 at $x = 0$ as well as at $x = 1$.

Moral of the story

While finding global extrema of a continuous function f defined on $[a, b]$, it is of no use to know whether a local extremum of f is in fact a local maximum or a local minimum. Hence do not find the second derivative of f.

To determine the global extrema of f, find the critical points of f, that is, the points at which the derivative of f does not exist and the points in (a, b) at which the derivative of f exists and is equal to 0. Then find the values of f at the critical points of f, and at the end points a and b of $[a, b]$.

Compare these values of f. The smallest among them is the global minimum of f and the largest among them is the global maximum of f.

Local Extrema and Derivatives

Theorem (First derivative test for a local maximum)

Let $D \subset \mathbb{R}$, c an interior point of D, and $f : D \to \mathbb{R}$. Suppose

(i) f is continuous at c, and

(ii) there is $\delta > 0$ such that $f' \geq 0$ on $(c - \delta, c)$ and $f' \leq 0$ on $(c, c + \delta)$.

Then f has a local maximum at c.

Proof: By (ii), f is increasing on $(c - \delta, c)$, and it is decreasing on $(c, c + \delta)$. Also, $\lim_{h \to 0} f(c + h) = f(c)$ by (i). Hence $x \in (c - \delta, c) \implies f(x) \leq \lim_{h \to 0^-} f(c + h) = f(c)$ and $x \in (c, c + \delta) \implies f(x) \leq \lim_{h \to 0^+} f(c + h) = f(c)$.

Hence f has a local maximum at c. $\square$

Theorem (First derivative test for a local minimum)

Let $D \subset \mathbb{R}$, c an interior point of D, and $f : D \to \mathbb{R}$. Suppose

(i) f is continuous at c, and

(ii) there is $\delta > 0$ such that $f' \leq 0$ on $(c - \delta, c)$ and $f' \geq 0$ on $(c, c + \delta)$.

Then f has a local minimum at c.

Proof: Similar to the proof of the previous result.

Examples:

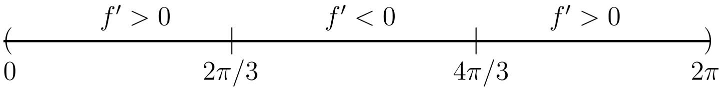
(i) Let $f(x) := |x|$ for $x \in \mathbb{R}$. Then f is continuous at 0, $f' = -1 \leq 0$ on $(-\infty, 0)$ and $f' = 1 \geq 0$ on $(0, \infty)$. Thus f has a local minimum at 0.

(ii) Let $f(x) := -x - 1$ for $x \in (-\infty, 0)$, $f(0) := 0$ and $f(x) := x + 1$ for $x \in (0, \infty)$. Then $f' = -1 \leq 0$ on $(-\infty, 0)$ and $f' = 1 \geq 0$ on $(0, \infty)$. But f does not have a local minimum at 0. Note: f is not continuous at 0.

(iii) Let $f(x) := x + 2\sin x$ for $x \in (0, 2\pi)$.

Then $f'(x) = 1 + 2\cos x = 0 \iff x \in \{2\pi/3, 4\pi/3\}$.

f is continuous at $2\pi/3$ and at $4\pi/3$. We note the following:



f has a local maximum at $2\pi/3$ and a local minimum at $4\pi/3$.

In general, the first derivative test for a local extremum can be informally summarized by the following Thumb Rule.

Suppose f is continuous at c.

f' changes from $+$ to $-$ at $c \implies f$ has a local maximum at c,

f' changes from $-$ to $+$ at $c \implies f$ has a local minimum at c.

Theorem (Second derivative test for a local extremum)

Let $D \subset \mathbb{R}$, and let c be an interior point of D. Suppose $f : D \to \mathbb{R}$ is twice differentiable at c, and $f'(c) = 0$.

(i) If $f''(c) < 0$, then f has a local maximum at c.

(ii) If $f''(c) > 0$, then f has a local minimum at c.

Proof: (i) Let $f''(c) < 0$. Since $f''(c)$ exists, there is $r > 0$ such that f' exits on $(c - r, c + r)$. Now

$$\lim_{x \to c} \frac{f'(x)}{x - c} = \lim_{x \to c} \frac{f'(x) - f'(c)}{x - c} = f''(c) < 0.$$

Hence there is $\delta > 0$ such that $f'(x)/(x - c) < 0$ for all $x \in (c - \delta, c) \cup (c, c + \delta)$. Now $x \in (c - \delta, c) \implies f'(x) > 0$ and $x \in (c, c + \delta) \implies f'(x) < 0$. By the first derivative test, f has a local maximum at c. A similar argument works for (ii). $\square$

Examples:

(i) Let $f(x) := x^4 - 2x^2$ for $x \in \mathbb{R}$. Then

$f'(x) = 4x^3 - 4x = 4(x+1)x(x-1) = 0 \iff x \in \{-1, 0, 1\}$.

f is twice differentiable, and $f''(x) = 4(3x^2 - 1)$ for $x \in \mathbb{R}$.

Since $f''(-1) = 8 > 0$, $f''(0) = -4 < 0$ and $f''(1) = 8 > 0$, f has a local maximum at 0, and has a local minimum at ± 1.

(ii) Let $f(x) := x^4$ for $x \in (-1, 1)$.
Then $f'(x) = 4x^3$ and $f''(x) = 12x^2$ for $x \in (-1, 1)$,
and so $f(0) = f'(0) = f''(0) = 0$.

Hence the second derivative test is not applicable at 0. But since $f' < 0$ on $(-1, 0)$ and $f' > 0$ on $(0, 1)$, the first derivative test shows that f has a local minimum at 0.

Note: The first derivative test is more general than the second.

Point of Inflection

Let I be an interval, and let $f : I \to \mathbb{R}$. An interior point of I at which convexity of f changes to concavity, or the other way round, is called a point of inflection for f. More precisely,

Definition

An interior point c of I is called a **point of inflection** for f if there is $\delta > 0$ such that
either f is convex on $(c - \delta, c)$ and concave on $(c, c + \delta)$,
or f is concave on $(c - \delta, c)$ and convex on $(c, c + \delta)$.

Examples:

Let $f(x) := x^3$ and $g(x) := x^{1/3}$ for $x \in \mathbb{R}$. Then 0 is a point of inflection for f as well as for g.

We give characterizations of a point of inflection of a differentiable and of a twice differentiable function.

Point of Inflection and Derivatives

Theorem (Derivative tests for a point of inflection)

Let c be an interior point of I, and let $f : I \to \mathbb{R}$.

(i) (First derivative test) Suppose there is $\delta > 0$ such that f is differentiable on $(c-\delta, c) \cup (c, c+\delta)$. Then c is point of inflection for $f \iff f'$ is increasing on $(c-\delta, c)$ and f' is decreasing on $(c, c+\delta)$, or the other way round.

(ii) (Second derivative test) Suppose there is $\delta > 0$ such that f is twice differentiable on $(c-\delta, c) \cup (c, c+\delta)$. Then c is point of inflection for $f \iff f'' \geq 0$ on $(c-\delta, c)$ and $f'' \leq 0$ on $(c, c+\delta)$, or the other way round.

Proof: Use the characterizations of convexity and concavity. □

Thumb Rule:

f'' changes sign at $c \iff c$ is a point of inflection for f.

(Necessary condition for a point of inflection)

Let c be an interior point of I, and let $f : I \to \mathbb{R}$. Suppose f is twice differentiable at c. If c is point of inflection for f, then $f''(c) = 0$.

Proof: Since $f''(c)$ exists, there is $r > 0$ such that f is differentiable on $(c - r, c + r)$. We use the first derivative test for a point of inflection.

Convex to concave: There is $\delta > 0$ such that f' is increasing on $(c - \delta, c)$ and f' is decreasing on $(c, c + \delta)$. Let $g := f'$ on $(c - \delta, c + \delta)$. Now g is continuous at c, g is increasing on $(c - \delta, c)$ and g is decreasing on $(c, c + \delta)$. By the first derivative test for g, there is a local maximum of g at c, and so $f''(c) = g'(c) = 0$.

Concave to convex: There is $\delta > 0$ such that f' is decreasing on $(c - \delta, c)$ and f' is increasing on $(c, c + \delta)$. As above, f' hs a local minimum of at c, and so $f''(c) = 0$. □

The condition '$f''(c) = 0$' is not sufficient to conclude that c is a point of inflection for f. Example: Let $f(x) := x^4$ for $x \in \mathbb{R}$. Then f is twice differentiable at 0 and $f''(0) = 0$. But 0 is not a point of inflection for f.

(Sufficient condition for a point of inflection)

Let c be an interior point of I, and let $f : I \to \mathbb{R}$. Suppose f is thrice differentiable at c. If $f''(c) = 0$ and $f'''(c) \neq 0$, then c is point of inflection for f.

Proof: Without loss of generality, suppose $f'''(c) < 0$. Then

$$\lim_{x \to c} \frac{f''(x)}{x - c} = \lim_{x \to c} \frac{f''(x) - f''(c)}{x - c} = f'''(c) < 0.$$

Hence there is $\delta > 0$ such that $f''(x)/(x - c) < 0$ for all $x \in (c - \delta, c) \cup (c, c + \delta)$. Now $f'' > 0$ on $(c - \delta, c)$, and $f'' < 0$ on $(c, c + \delta)$. By the second derivative test for a point of inflection, c is point of inflection for f. $\square$

The condition '$f'''(c) \neq 0$' is not necessary for c to be a point of inflection of f. Example: Let $f(x) := x^5$ for $x \in \mathbb{R}$. Then 0 is a point of inflection for f, but $f'''(0) = 0$.

To conclude the topic of local extrema and points of inflection, let us consider the following example.

Let $f(x) := x^4 - 4x^3$ for $x \in \mathbb{R}$. Then

$f'(x) = 4x^3 - 12x^2 = 4x^2(x-3) = 0 \iff x \in \{0, 3\}$,

$f''(x) = 12x^2 - 24x = 12x(x-2) = 0 \iff x \in \{0, 2\}$.

Also, $f'''(x) = 24(x-1)$ for $x \in \mathbb{R}$.

Since $f''(0) = 0 = f''(2)$ and $f'''(0) = -f'''(2) = -24 \neq 0$, we see that 0 and 2 are points of inflection for f.

Since $f' \leq 0$ on $(-\infty, 3)$ and $f' \neq 0$ on $(-\infty, 0) \cup (0, 3)$, the function f is strictly decreasing on $(-3, 3)$. Hence f does not have a local extremum at 0. Further, since $f'(3) = 0$ and $f''(3) = 36 > 0$, the function f has a local minimum at 3.

MA 105 : Calculus
Division 1, Lecture 06

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Rolle's theorem and its proof (using Extreme Value Property)
- (Lagrange's) Mean Value Theorem and its proof
- Consequences of MVT
- Monotonicity. Relation with derivatives
- Convexity and Concavity.

Convexity and Concavity

Let I be an interval. A function $f : I \to \mathbb{R}$ is called **convex** on I if the line segment joining any two points on the graph of f lies on or above the graph of f. If $x_1, x_2 \in I$, then the equation of the line passing through $(x_1, f(x_1))$ and $(x_2, f(x_2))$ is given by $y = f(x_1) + \dfrac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1)$. Thus f is **convex** on I if

$$x_1, x_2, \ x \in I, \ x_1 < x < x_2 \Longrightarrow f(x) \leq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1).$$

Similarly, a function $f : I \to \mathbb{R}$ is called **concave** on I if

$$x_1, x_2, \ x \in I, \ x_1 < x < x_2 \Longrightarrow f(x) \geq f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1).$$

Remark: It is easy to see that f

$$f((1-\lambda)x_1 + \lambda x_2) \leq (1-\lambda)f(x_1) + \lambda f(x_2) \ \forall \, \lambda \in [0, 1] \text{ and } x_1, x_2 \in I.$$

Likewise for concacity (with $\leq$ changed to $\geq$).

A function $f : I \to \mathbb{R}$ is called **strictly convex** on I if the inequality $\leq$ in the definition of a convex function can be replaced by the strict inequality $<$.

A function $f : I \to \mathbb{R}$ is called **strictly concave** on I if the inequality $\geq$ in the definition of a concave function can be replaced by the strict inequality $>$.

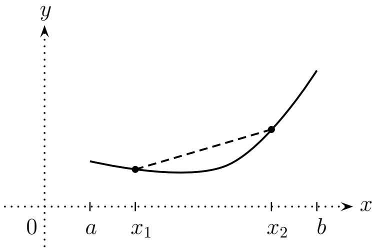
Examples:

(i) Let $\alpha, \beta \in \mathbb{R}$, and let $f(x) := \alpha + \beta\, x$ for $x \in \mathbb{R}$. Then f is convex as well as concave on $\mathbb{R}$ (but not strictly).

(ii) Let $\alpha \in \mathbb{R}$, and let $f(x) := \alpha\, |x|$ for $x \in \mathbb{R}$. Then f is convex on $\mathbb{R}$ if $\alpha > 0$, and f is concave on $\mathbb{R}$ if $\alpha < 0$ (but not strictly). (Prove this from the first principles.)

(iii) Let $\alpha \in \mathbb{R}$, and let $f(x) := \alpha\, x^2$ for $x \in \mathbb{R}$. Then f is strictly convex on $\mathbb{R}$ if $\alpha > 0$, and f is strictly concave on $\mathbb{R}$ if $\alpha < 0$. (Prove this from the first principles.)

Typical graphs of convex and concave functions on $I := [a, b]$ look as follows:



Let I be an interval, and let $f : I \to \mathbb{R}$ be a function. Suppose the curve $y = f(x)$ has a (unique) nonvertical tangent at each point on it.

It is apparent that if f is convex, then the slopes of the tangents increase as we move from left to right, and if f is concave, then then the slopes of the tangents decrease as we move from left to right. This is made precise in the following.

Proposition

Let I be an interval and $f : I \to \mathbb{R}$ be differentiable. Then

(i) f' is increasing on $I \iff f$ is convex on I.

(ii) f' is decreasing on $I \iff f$ is concave on I.

(iii) f' is strictly increasing on $I \iff f$ is strictly convex on I.

(iv) f' is strictly decreasing on $I \iff f$ is strictly concave on I.

We omit proofs of the above results as they are a bit involved.

Corollary

Let I be an interval and $f : I \to \mathbb{R}$ be twice differentiable.

(i) $f'' \geq 0$ on $I \iff f$ is convex on I.

(ii) $f'' \leq 0$ on $I \iff f$ is concave on I.

(iii) $f'' > 0$ on $I \implies f$ is strictly convex on I.

(iv) $f'' < 0$ on $I \implies f$ is strictly concave on I.

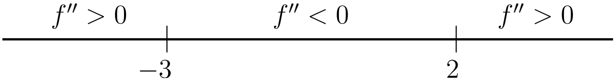
The converse of (iii) in the above corollary is false. To see this, consider $f(x) := x^4$ for $x \in \mathbb{R}$. Then f is strictly convex on $\mathbb{R}$, but $f''(0) = 0$. Similarly, the converse of (iv) is false.

Example:

Let $f(x) := x^4 + 2x^3 - 36x^2 + 62x + 5$ for $x \in \mathbb{R}$. Let us find intervals on which f is (strictly) convex or (strictly) concave. For $x \in \mathbb{R}$,

$$
\begin{aligned}
f'(x) &= 4x^3 + 6x^2 - 72x + 62, \text{ and} \\
f''(x) &= 12x^2 + 12x - 72 = 12(x+3)(x-2).
\end{aligned}
$$

Note:



- f is strictly convex on $(-\infty, -3)$.
- f is strictly concave on $(-3, 2)$.
- f is strictly convex on $(2, \infty)$.

Critical Points and Global Extrema

Let $D \subset \mathbb{R}$, and let $f : D \to \mathbb{R}$. An interior point c of D is called a **critical point** of f if either f is not differentiable at c, or if f is differentiable at c and $f'(c) = 0$.

Proposition

Let $f : [a, b] \to \mathbb{R}$ be continuous. Then the global minimum $m := \min\{f(x) : x \in [a, b]\}$ as well as the global maximum $M := \max\{f(x) : x \in [a, b]\}$ of f on $[a, b]$ is attained either at a critical point of f or at an end-point of $[a, b]$.

Proof. There is $c_1 \in [a, b]$ such that $f(c_1) = m$. If $c_1 = a$ or $c_1 = b$, then we are done. Next, let $c_1 \in (a, b)$. Then f has a local minimum at c_1. If f is not differentiable at c_1, then c_1 is a critical point of f. If f is differentiable at c_1, then $f'(c_1) = 0$ by an earlier result, and so c_1 is a critical point of f. Similarly, we argue for the global maximum M of f. $\square$

Finding Global Extrema

Example: Consider $f : [-1, 2] \to \mathbb{R}$ defined by

$$f(x) := \begin{cases} -x & \text{if} \quad -1 \le x \le 0, \\ 2x^3 - 4x^2 + 2x & \text{if} \quad 0 < x \le 2. \end{cases}$$

f is continuous on $[-1, 2]$: $\lim\limits_{x \to 0^-} f(x) = f(0) = \lim\limits_{x \to 0^+} f(x)$.
f is not differentiable at 0 : $f'_-(0) = -1$ and $f'_+(0) = 2$.
For $-1 < x < 0$, $f'(x) = -1 \neq 0$.
For $0 < x < 2$, $f'(x) = 6x^2 - 8x + 2 = 2(3x - 1)(x - 1)$.
Critical points of f: 0, $1/3$, 1. End-points of $[-1, 2]$: -1, 2.

x	-1	0	$1/3$	1	2
$f(x)$	1	0	$8/27$	0	4

f attains its global maximum 4 at $x = 2$, and its global minimum 0 at $x = 0$ as well as at $x = 1$.

Moral of the story

While finding global extrema of a continuous function f defined on $[a, b]$, it is of no use to know whether a local extremum of f is in fact a local maximum or a local minimum. Hence do not find the second derivative of f.

To determine the global extrema of f, find the critical points of f, that is, the points at which the derivative of f does not exist and the points in (a, b) at which the derivative of f exists and is equal to 0. Then find the values of f at the critical points of f, and at the end points a and b of $[a, b]$.

Compare these values of f. The smallest among them is the global minimum of f and the largest among them is the global maximum of f.

Local Extrema and Derivatives

Intuitively, it is clear that if the slopes of tangents to a curve $y = f(x)$ are positive to the left of $x = c$ and negative to the right of $x = c$, then f has a local maximum at c. A more precise and general formulation of this is the following.

Theorem (First derivative test for a local maximum)

Let $D \subset \mathbb{R}$, c an interior point of D, and $f : D \to \mathbb{R}$. Suppose

(i) f is continuous at c, and

(ii) there is $\delta > 0$ such that $f' \geq 0$ on $(c - \delta, c)$ and $f' \leq 0$ on $(c, c + \delta)$.

Then f has a local maximum at c.

Proof: By (ii), f is increasing on $(c - \delta, c)$, and it is decreasing on $(c, c + \delta)$. Also, $\lim_{h\to 0} f(c + h) = f(c)$ by (i). Hence $x \in (c - \delta, c) \implies f(x) \leq \lim_{h\to 0^-} f(c + h) = f(c)$ and $x \in (c, c + \delta) \implies f(x) \leq \lim_{h\to 0^+} f(c + h) = f(c)$. Hence f has a local maximum at c. $\square$

Theorem (First derivative test for a local minimum)

Let $D \subset \mathbb{R}$, c an interior point of D, and $f : D \to \mathbb{R}$. Suppose

(i) f is continuous at c, and

(ii) there is $\delta > 0$ such that $f' \leq 0$ on $(c - \delta, c)$ and $f' \geq 0$ on $(c, c + \delta)$.

Then f has a local minimum at c.

Proof: Similar to the proof of the previous result.

Examples:

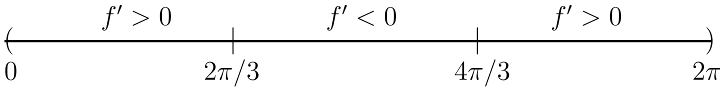
(i) Let $f(x) := |x|$ for $x \in \mathbb{R}$. Then f is continuous at 0, $f' = -1 \leq 0$ on $(-\infty, 0)$ and $f' = 1 \geq 0$ on $(0, \infty)$. Thus f has a local minimum at 0.

(ii) Let $f(x) := -x - 1$ for $x \in (-\infty, 0)$, $f(0) := 0$ and $f(x) := x + 1$ for $x \in (0, \infty)$. Then $f' = -1 \leq 0$ on $(-\infty, 0)$ and $f' = 1 \geq 0$ on $(0, \infty)$. But f does not have a local minimum at 0. Note: f is not continuous at 0.

(iii) Let $f(x) := x + 2\sin x$ for $x \in (0, 2\pi)$.

Then $f'(x) = 1 + 2\cos x = 0 \iff x \in \{2\pi/3, 4\pi/3\}$.

f is continuous at $2\pi/3$ and at $4\pi/3$. We note the following:



f has a local maximum at $2\pi/3$ and a local minimum at $4\pi/3$.

In general, the first derivative test for a local extremum can be informally summarized by the following Thumb Rule.

Suppose f is continuous at c.

f' changes from $+$ to $-$ at $c \implies f$ has a local maximum at c,

f' changes from $-$ to $+$ at $c \implies f$ has a local minimum at c.

Theorem (Second derivative test for a local extremum)

Let $D \subset \mathbb{R}$, and let c be an interior point of D. Suppose $f : D \to \mathbb{R}$ is twice differentiable at c, and $f'(c) = 0$.

(i) If $f''(c) < 0$, then f has a local maximum at c.

(ii) If $f''(c) > 0$, then f has a local minimum at c.

Proof: (i) Let $f''(c) < 0$. Since $f''(c)$ exists, there is $r > 0$ such that f' exits on $(c-r, c+r)$. Now

$$\lim_{x \to c} \frac{f'(x)}{x-c} = \lim_{x \to c} \frac{f'(x) - f'(c)}{x-c} = f''(c) < 0.$$

Hence there is $\delta > 0$ such that $f'(x)/(x-c) < 0$ for all $x \in (c-\delta, c) \cup (c, c+\delta)$. [This follows from the ϵ-δ characterization of limits. Verify!] This implies that $f'(x) > 0$ when $x \in (c-\delta, c)$ and $f'(x) < 0$ when $x \in (c, c+\delta)$. Also f is continuous at c (being differentiable). Hence by the first derivative test, f has a local maximum at c.

A similar argument works for (ii). $\square$

Examples:

(i) Let $f(x) := x^4 - 2x^2$ for $x \in \mathbb{R}$. Then

$f'(x) = 4x^3 - 4x = 4(x+1)x(x-1) = 0 \iff x \in \{-1, 0, 1\}$.

f is twice differentiable, and $f''(x) = 4(3x^2 - 1)$ for $x \in \mathbb{R}$.

Since $f''(-1) = 8 > 0$, $f''(0) = -4 < 0$ and $f''(1) = 8 > 0$, f has a local maximum at 0, and has a local minimum at ± 1.

(ii) Let $f(x) := x^4$ for $x \in (-1, 1)$.
Then $f'(x) = 4x^3$ and $f''(x) = 12x^2$ for $x \in (-1, 1)$,
and so $f(0) = f'(0) = f''(0) = 0$.

Hence the second derivative test is not applicable at 0. But since $f' < 0$ on $(-1, 0)$ and $f' > 0$ on $(0, 1)$, the first derivative test shows that f has a local minimum at 0.

Note: The first derivative test is more general than the second.

Point of Inflection

Let I be an interval, and let $f : I \to \mathbb{R}$. An interior point of I at which convexity of f changes to concavity, or the other way round, is called a point of inflection for f. More precisely,

Definition

An interior point c of I is called a **point of inflection** for f if there is $\delta > 0$ such that
either f is convex on $(c - \delta, c)$ and concave on $(c, c + \delta)$,
or f is concave on $(c - \delta, c)$ and convex on $(c, c + \delta)$.

Examples:

Let $f(x) := x^3$ and $g(x) := x^{1/3}$ for $x \in \mathbb{R}$. Then 0 is a point of inflection for f as well as for g.

We give characterizations of a point of inflection of a differentiable and of a twice differentiable function.

Point of Inflection

Let I be an interval, and let $f : I \to \mathbb{R}$. An interior point of I at which convexity of f changes to concavity, or vice-versa, is called a point of inflection for f. More precisely,

Definition

An interior point c of I is called a **point of inflection** for f if there is $\delta > 0$ such that
either f is convex on $(c - \delta, c)$ and concave on $(c, c + \delta)$,
or f is concave on $(c - \delta, c)$ and convex on $(c, c + \delta)$.

Examples:

Let $f(x) := x^3$ and $g(x) := x^{1/3}$ for $x \in \mathbb{R}$. Then 0 is a point of inflection for f as well as for g.

We give characterizations of a point of inflection of a differentiable and of a twice differentiable function.

Point of Inflection and Derivatives

Theorem (Derivative tests for a point of inflection)

Let c be an interior point of I, and let $f : I \to \mathbb{R}$.

(i) (First derivative test) Suppose f is differentiable on $(c-r, c) \cup (c, c+r)$ for some $r > 0$. Then c is point of inflection for $f \iff$ there is $\delta > 0$ with $\delta < r$ such that f' is increasing on $(c-\delta, c)$ and f' is decreasing on $(c, c+\delta)$, or vice-versa.

(ii) (Second derivative test) Suppose f is twice differentiable on $(c-r, c) \cup (c, c+r)$ for some $r > 0$.. Then c is point of inflection for $f \iff$ there is $\delta > 0$ such that $f'' \geq 0$ on $(c-\delta, c)$ and $f'' \leq 0$ on $(c, c+\delta)$, or vice-versa.

Proof: Use the characterizations of convexity and concavity. $\square$

Thumb Rule:

f'' changes sign at $c \iff c$ is a point of inflection for f.

(Necessary condition for a point of inflection)

Let c be an interior point of I, and let $f : I \to \mathbb{R}$. Suppose f is twice differentiable at c. If c is point of inflection for f, then $f''(c) = 0$.

Proof: Since $f''(c)$ exists, there is $r > 0$ such that f is differentiable on $(c - r, c + r)$. Suppose c is point of inflection for f. By the first derivative test for a point of inflection, there is $\delta > 0$ with $\delta < r$ such that (i) f' is increasing on $(c - \delta, c)$ and f' is decreasing on $(c, c + \delta)$, or (ii) vice-versa. Suppose we are in Case (i). Then for all $x, y \in (c - \delta, c)$ with $x < y$ and for all $z, w \in (c, c + \delta)$ with $z < w$, we must have

$$f'(x) \le f'(y) \quad \text{and} \quad f'(z) \ge f'(w).$$

Since $f''(c)$ exists, f' is continuous at c. Hence taking limit as $y \to c^-$ and as $z \to c^+$, we see that $f'(x) \le f'(c)$ for all $x \in (c - \delta, c + \delta)$. Thus f' has a local maximum at c. Consequently, $f''(c) = 0$. The proof in Case (ii) is similar. $\square$

The condition '$f''(c) = 0$' is not sufficient to conclude that c is a point of inflection for f. Example: Let $f(x) := x^4$ for $x \in \mathbb{R}$. Then f is twice differentiable at 0 and $f''(0) = 0$. But 0 is not a point of inflection for f.

(Sufficient condition for a point of inflection)

Let c be an interior point of I, and let $f : I \to \mathbb{R}$. Suppose f is thrice differentiable at c. If $f''(c) = 0$ and $f'''(c) \neq 0$, then c is point of inflection for f.

Proof: Without loss of generality, suppose $f'''(c) < 0$. Then

$$\lim_{x \to c} \frac{f''(x)}{x - c} = \lim_{x \to c} \frac{f''(x) - f''(c)}{x - c} = f'''(c) < 0.$$

Hence there is $\delta > 0$ such that $f''(x)/(x - c) < 0$ for all $x \in (c - \delta, c) \cup (c, c + \delta)$. Now $f'' > 0$ on $(c - \delta, c)$, and $f'' < 0$ on $(c, c + \delta)$. By the second derivative test for a point of inflection, c is point of inflection for f. $\square$

The condition '$f'''(c) \neq 0$' is not necessary for c to be a point of inflection of f. Example: Let $f(x) := x^5$ for $x \in \mathbb{R}$. Then 0 is a point of inflection for f, but $f'''(0) = 0$.

To conclude the topic of local extrema and points of inflection, let us consider the following example.

Let $f(x) := x^4 - 4x^3$ for $x \in \mathbb{R}$. Then

$f'(x) = 4x^3 - 12x^2 = 4x^2(x-3) = 0 \iff x \in \{0, 3\}$,

$f''(x) = 12x^2 - 24x = 12x(x-2) = 0 \iff x \in \{0, 2\}$.

Also, $f'''(x) = 24(x-1)$ for $x \in \mathbb{R}$.

Since $f''(0) = 0 = f''(2)$ and $f'''(0) = -f'''(2) = -24 \neq 0$, we see that 0 and 2 are points of inflection for f.

Since $f' \leq 0$ on $(-\infty, 3)$ and $f' \neq 0$ on $(-\infty, 0) \cup (0, 3)$, the function f is strictly decreasing on $(-3, 3)$. Hence f does not have a local extremum at 0. Further, since $f'(3) = 0$ and $f''(3) = 36 > 0$, the function f has a local minimum at 3.

Geometric properties of a function

Let I be an interval. We have so far considered the following geometric properties of a function from I to $\mathbb{R}$.

- Boundedness, and the bounds being attained
- Intermediate value property
- Monotonicity
- Convexity/Concavity
- Having global extrema, having local extrema
- Having points of inflection

Further, we have given analytical conditions on the function (such as continuity, nonnegativity of the derivatives, the vanishing of the derivatives etc.) which imply the above-mentioned geometric properties.
Next, we shall now prepare the groundwork to be able to talk about asymptotes.

MA 105 : Calculus
Division 1, Lecture 7

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Convexity and concavity, revisited
- Derivative Tests for Convexity and concavity
- Critical points and global extrema
- Local extrema and derivatives
- First Derivative Test and Second Derivative Test for Local Extrema
- Point of inflection. Derivative Tests
- Necessary condition for a point of inflection
- Sufficient condition for a point of inflection
- Summary of geometric properties of functions
- Brief discussion of asymptotes and "infinite limits"

'Infinite limits' and 'limits at infinity'

Definition

Let (a_n) be a sequence of real numbers. We say that (a_n) **tends to infinity** and write $a_n \to \infty$ if for every $\alpha > 0$, there is $n_0 \in \mathbb{N}$ such that $a_n > \alpha$ for all $n \geq n_0$.

We say that (a_n) **tends to minus infinity** and write $a_n \to -\infty$ if for every $\beta < 0$, there is $n_0 \in \mathbb{N}$ such that $a_n < \beta$ for all $n \geq n_0$.

Note: Since ∞ and $-\infty$ are not numbers, we avoid writing $\lim_{n\to\infty} a_n = \infty$ when $a_n \to \infty$. Likewise when $a_n \to -\infty$.

Examples: (i) Let $a_n := n^2$ for $n \in \mathbb{N}$. To see that $a_n \to \infty$, let $\alpha > 0$. Then $n^2 > \alpha$ if $n > \alpha^{1/2}$. So take $n_0 := [\alpha^{1/2}] + 1$.

(ii) Let $a_n := -n^{1/3}$ for $n \in \mathbb{N}$. To see that $a_n \to -\infty$, let $\beta < 0$. Then $-n^{1/3} < \beta$ if $n > -\beta^3$. So take $n_0 := [-\beta^3] + 1$.

Limits as $x \to \infty$ and as $x \to -\infty$

Definition

Suppose $D \subset \mathbb{R}$ is such that $(a, \infty) \subset D$ for some $a \in \mathbb{R}$. For a function $f : D \to \mathbb{R}$, we say that a **limit** of $f(x)$ exists as $x \to \infty$ if there is $\ell \in \mathbb{R}$ such that

(x_n) is a sequence in D and $x_n \to \infty \implies f(x_n) \to \ell$.

Notation: $f(x) \to \ell$ as $x \to \infty$ or $\lim_{x\to\infty} f(x) = \ell$.

Similarly, we define $f(x) \to \ell$ as $x \to -\infty$ or $\lim_{x\to-\infty} f(x) = \ell$.

Examples:

Let $f(x) := \dfrac{1}{x}$ for $x > 0$. Then $f(x) \to 0$ as $x \to \infty$.

Let $f(x) := \dfrac{x+1}{x-1}$ for $x < 1$. Then $f(x) \to 1$ as $x \to -\infty$.

$f(x) \to \infty$ and $f(x) \to -\infty$

Definition

Let $f : D \to \mathbb{R}$ and $c \in \mathbb{R}$ be such that there is $r > 0$ with $(c - r, c) \cup (c, c + r) \subset D$. We say that $f(x)$ **tends to** ∞ **as** $x \to c$, and write $f(x) \to \infty$ **as** $x \to c$ if

(x_n) is a sequence in D, $x_n \neq c$ and $x_n \to c \implies f(x_n) \to \infty$.

By restricting to sequences (x_n) in D with $x_n < c$ for all $n \in \mathbb{N}$, we can define $f(x) \to \infty$ as $x \to c^-$. Similarly, we can define $f(x) \to \infty$ as $x \to c^+$, and furthermore, we can also define $f(x) \to -\infty$ as $x \to c$ (or as $x \to c^-$ or as $x \to c^+$).

Note: As in the case of sequences, we shall not write $\lim_{x \to c} f(x) = \infty$ when $f(x) \to \infty$ as $x \to c$. Likewise, we shall not write $\lim_{x \to c} f(x) = -\infty$ when $f(x) \to -\infty$ as $x \to c$.

Example: Let $f(x) := 1/x$ for $x \in \mathbb{R} \setminus \{0\}$. Then $f(x) \to \infty$ as $x \to 0^+$ and $f(x) \to -\infty$ as $x \to 0^-$.

Asymptotes

Let $D \subset \mathbb{R}$, and let $f : D \to \mathbb{R}$. There are three possible types of asymptotes: horizontal, oblique and vertical. They all imply nearness of the curve $y = f(x)$ to a straight line.

- A straight line $y = b$ is called a **horizontal asymptote** of the curve $y = f(x)$ if

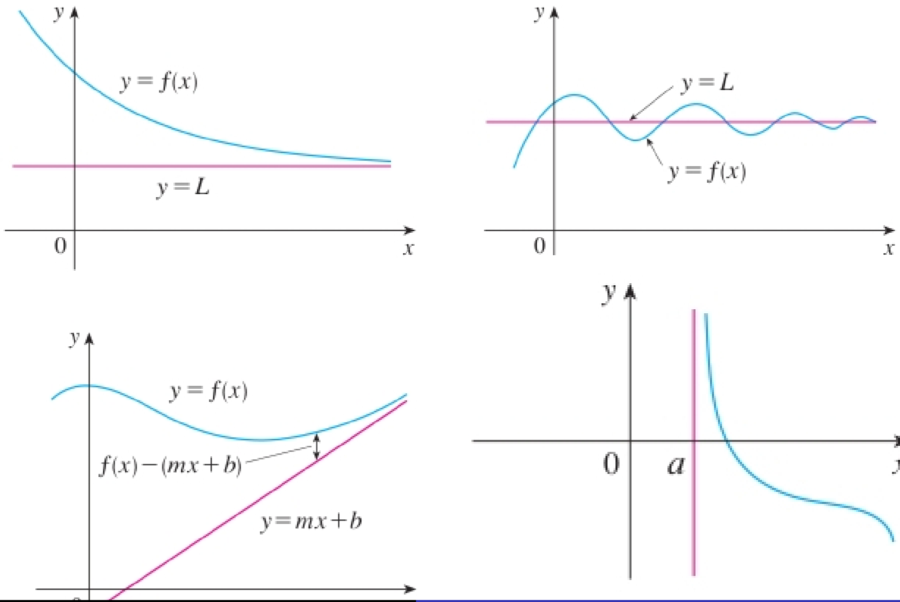
  $$\lim_{x\to\infty} f(x) = b \text{ or } \lim_{x\to-\infty} f(x) = b.$$

- A straight line $y = ax + b$, where $a \neq 0$, is called an **oblique asymptote** of the curve $y = f(x)$ if

  $$\lim_{x\to\infty} (f(x) - ax - b) = 0 \text{ or } \lim_{x\to-\infty} (f(x) - ax - b) = 0.$$

- A line $x = a$ is called a **vertical asymptote** of the curve $y = f(x)$ if
  $f(x) \to \infty$ as $x \to a^-$, or $f(x) \to -\infty$ as $x \to a^-$, or
  $f(x) \to \infty$ as $x \to a^+$, or $f(x) \to -\infty$ as $x \to a^+$.

Illustrations of asymptotes



Examples of asymptotes

Example: Consider $f : (-\infty, 0) \cup (1, \infty) \to \mathbb{R}$ defined by

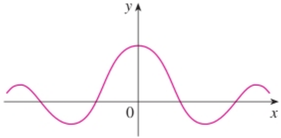
$$f(x) := \begin{cases} (3x^2 + 4x + 1)/x & \text{if} \quad x < 0, \\ (2x-1)/(x-1) & \text{if} \quad x > 1. \end{cases}$$

- For $x > 1$, we have $f(x) = 2 + 1/(x-1)$, and so $\lim_{x \to \infty} f(x) = 2$. Hence the straight line $y = 2$ is a horizontal asymptote of the curve $y = f(x)$.
- For $x < 0$, $f(x) = 3x + 4 + (1/x)$, and so $\lim_{x \to -\infty} (f(x) - (3x+4)) = 0$. Hence the straight line $y = 3x + 4$ is an oblique asymptote of the curve $y = f(x)$.
- $f(x) \to \infty$ as $x \to 1^+$. Hence the straight line $x = 1$ is a vertical asymptote of the curve $y = f(x)$.
- $f(x) \to -\infty$ as $x \to 0^-$. Hence the straight line $x = 0$ is a vertical asymptote of the curve $y = f(x)$.

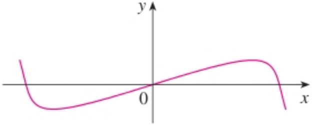
Exercise: Sketch the graph of f.

Even, odd and periodic functions

A function $f : (-a, a) \to \mathbb{R}$ is called **even** if $f(-x) = f(x)$, and it is called **odd** if $f(-x) = -f(x)$ for all $x \in (-a, a)$.

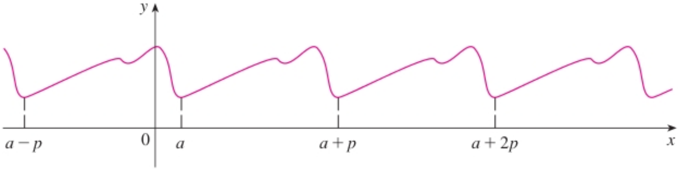


(a) Even function: reflectional symmetry



(b) Odd function: rotational symmetry

If $f : \mathbb{R} \to \mathbb{R}$ and $f(x) = f(x+p)$ for a fixed $p > 0$ and all $x \in \mathbb{R}$, then f is called **periodic** and p is called a **period** of f.



Curve Sketching: Tips

- Look for symmetries of the given function f, that is, find out if f is even/ odd/ periodic?
- Look for intercepts, that is, the intersections of the curve with co-ordinate axes.
- Look for points (if any) where f is not defined. If such a point exists, study the behaviour of f near that point.
- Locate critical points, local maxima, local minima of f.
- Find intervals on which f is increasing or decreasing.
- Find intervals on which f is convex or concave.
- Locate points of inflection of f.
- Determine the behaviour of $f(x)$ as $x \to \infty$ or $x \to -\infty$.
- Find the asymptotes, if any.

Curve Sketching: An Example

Example: Let $f(x) := (x^2 - 1)/(x + 2)$ for $x \in \mathbb{R}$ with $x \neq -2$.

- f is not even, not odd, not periodic.
- Intercepts of f: $(0, -1/2), (1, 0), (-1, 0)$.
- $f'(x) = (x + 2 + \sqrt{3})(x + 2 - \sqrt{3})/(x + 2)^2$ and $f''(x) = 6/(x + 2)^3$ for $x \neq -2$.
- f is increasing on $(-\infty, -2 - \sqrt{3})$ and on $(-2 + \sqrt{3}, \infty)$. f is decreasing on $(-2 - \sqrt{3}, -2)$ and on $(-2, -2 + \sqrt{3})$.
- f is concave on $(-\infty, -2)$ and convex on $(-2, \infty)$.
- Since $f'(-2 - \sqrt{3}) = 0 = f'(-2 + \sqrt{3})$ and $f''(-2 - \sqrt{3}) = -2/\sqrt{3} < 0$, $f''(-2 + \sqrt{3}) = 2/\sqrt{3} > 0$, f has a local maximum at $-2 - \sqrt{3}$ and f has a local minimum at $-2 + \sqrt{3}$. Note: $f(-2 \pm \sqrt{3}) = -4 \pm 2\sqrt{3}$.
- There is no point of inflection for f.
- The straight line $x = -2$ is a vertical asymptote.
- Since $f(x) = x - 2 + 3/(x + 2)$ for $x \neq -2$, $y = x - 2$ is an oblique asymptote.

Mean Value Theorem Revisited

Recall the most important result in differential calculus which we proved in Lecture 7:

Theorem (Mean Value Theorem: MVT)

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

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

Let us now consider an extension of the above result.

Theorem (Extended Mean Value Theorem)

Let $a < b$ and $f : [a, b] \to \mathbb{R}$ be such that f' exists on $[a, b]$. Suppose f' is continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that

$$f(b) = f(a) + f'(a)(b - a) + \frac{f''(c)}{2}(b - a)^2.$$

Proof: For $x \in [a, b]$, let $P(x) := f(a) + f'(a)(x - a)$. Define

$$F(x) := f(x) - P(x) - \frac{f(b) - P(b)}{(b-a)^2}(x-a)^2 \quad \text{for } x \in [a, b].$$

Then $F(a) = 0 = F(b)$. By the Rolle theorem, there exists $c_1 \in (a, b)$ such that $F'(c_1) = 0$. Since $F'(a) = 0$ also, there is $c \in (a, c_1)$ such that $F''(c) = 0$. This gives the desired result.

More generally, we can obtain the following important result.

Theorem (Taylor Theorem)

Let $a < b$ and $n \in \mathbb{N}$. Let $f : [a, b] \to \mathbb{R}$ be such that $f', \ldots, f^{(n)}$ exist on $[a, b]$. Suppose $f^{(n)}$ is continuous on $[a, b]$ and differentiable on (a, b). Then there exists $c \in (a, b)$ such that $f(b) = f(a) + f'(a)(b-a) + \cdots + \dfrac{f^{(n)}(a)}{n!}(b-a)^n + R_n$ where $R_n := f^{(n+1)}(c)(b-a)^{n+1}/(n+1)!$.

Proof (Sketch): For $x \in [a, b]$, let

$$P(x) := f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Consider $F : [a, b] \to \mathbb{R}$ defined by

$$F(x) := f(x) - P(x) - \frac{f(b) - P(b)}{(b-a)^{n+1}}(x-a)^{n+1} \quad \text{for } x \in [a, b].$$

Then $F(a) = 0 = F(b)$. By the Rolle theorem, there exists $c_1 \in (a, b)$ such that $F'(c_1) = 0$. Further,

$$F'(a) = 0 = F'(c_1)$$

and so there is $c_2 \in (a, c_1)$ such that $F''(c_2) = 0$. Now

$$F''(a) = 0 = F''(c_2)$$

and so there is $c_3 \in (a, c_2)$ such that $F'''(c_3) = 0$. Continuing in this way, we see that there is $c = c_{n+1} \in (a, b)$ such that $F^{(n+1)}(c) = 0$, and this yields the desired result.

A result similar to Taylor's theorem holds with a and b interchanged. [This can be seen by applying Taylor's Theorem to $g : [a, b] \to \mathbb{R}$ defined by $g(x) := f(a + b - x)$.] As a consequence, we obtain the following result.

Taylor Formula: Let $n \in \mathbb{N}$. If I is an interval, $a \in I$, and $f : I \to \mathbb{R}$ is such that $f', f'', \ldots, f^{(n)}$ exist on I and $f^{(n+1)}$ exists at every interior point of I, then for $x \in I$, $x \neq a$, there is c_x between a and x such that $f(x) = P_n(x) + R_n(x)$, where

$$\begin{aligned} P_n(x) &:= f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n, \\ R_n(x) &:= \frac{f^{(n+1)}(c_x)}{(n+1)!}(x-a)^{n+1}. \end{aligned}$$

Here the polynomial $P_n(x)$ is of degree $\leq n$, and it can be considered as an 'approximation' of $f(x)$ of order n. It is called the nth **Taylor polynomial** of f around a, and $R_n(x)$ is called the nth **Lagrange remainder** of f around a.

Example:

Let $f(x) := \sin x$ for $x \in \mathbb{R}$. Then for $x \in \mathbb{R}$, $f'(x) = \cos x$, $f''(x) = -\sin x$, $f'''(x) = -\cos x, \ldots$, and in general,

$$f^{(k)}(x) = \begin{cases} (-1)^{k/2} \sin x & \text{if} \quad k \text{ is even,} \\ \\ (-1)^{(k-1)/2} \cos x & \text{if} \quad k \text{ is odd.} \end{cases}$$

Fix $n \in \mathbb{N}$. Then the nth Taylor polynomial of f around 0 is

$$\begin{aligned} P_n(x) &= f(0) + f'(0)(x-0) + \cdots + \frac{f^{(n)}(0)}{n!}(x-0)^n \\ &= \begin{cases} x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \cdots (-1)^{(n-1)/2} \dfrac{x^n}{n!} & \text{if} \quad n \text{ is odd,} \\ \\ x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \cdots (-1)^{(n-2)/2} \dfrac{x^{n-1}}{(n-1)!} & \text{if} \quad n \text{ is even} \end{cases} \end{aligned}$$

for $x \in \mathbb{R}$.

Let $x \in \mathbb{R}$. Then the nth Lagrange remainder around 0 is

$$R_n(x) = \begin{cases} (-1)^{n/2} \dfrac{x^{n+1}}{(n+1)!} \cos c_x & \text{if} \quad n \text{ is even,} \\ \\ (-1)^{(n+1)/2} \dfrac{x^{n+1}}{(n+1)!} \sin c_x & \text{if} \quad n \text{ is odd,} \end{cases}$$

where c_x is between 0 and x. As $|\cos c_x| \le 1$ & $|\sin c_x| \le 1$,

$$|R_n(x)| \le \frac{|x|^{n+1}}{(n+1)!} \to 0 \text{ as } n \to \infty.$$

Hence for $x \in \mathbb{R}$,

$$\sin x = \lim_{n\to\infty} P_n(x) = \lim_{n\to\infty} \sum_{k=1}^{n} (-1)^{(k-1)} \frac{x^{2k-1}}{(2k-1)!}, \text{ that is,}$$

$$\sin x = \sum_{k=1}^{\infty} (-1)^{(k-1)} \frac{x^{2k-1}}{(2k-1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots .$$

The RHS is the **Taylor series** of the function sin around 0.

Riemann Integration

Let $a, b \in \mathbb{R}$ with $a < b$, and let f be a bounded function defined on $[a, b]$. Let us first assume that $f \geq 0$ on $[a, b]$. We are interested in assigning a meaning to the concept of 'area' of the region that lies under the graph of f, between the lines $x = a$, $x = b$, and above the x-axis, that is, of the region

$$R_f := \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\}.$$

A **partition** of $[a, b]$ is a finite ordered set $P := \{x_0, x_1, \ldots, x_n\}$ of points in $[a, b]$ such that $a = x_0 < x_1 < x_2 < \cdots < x_n = b$, where $n \in \mathbb{N}$.

The points in a partition divide the interval $[a, b]$ into nonoverlapping subintervals, $[x_0, x_1]$, ..., $[x_{n-1}, x_n]$. We call $[x_{i-1}, x_i]$ the ith **subinterval** of the partition P.

Example: For $n \in \mathbb{N}$, the set

$$\mathrm{P}_n := \left\{a,\ a + \frac{(b-a)}{n}, \ldots, a + \frac{(n-1)(b-a)}{n},\ b\right\}$$

is a partition of $[a, b]$ into n equal parts, each of length $(b-a)/n$. If $n := 1$, then $\mathrm{P}_1 := \{a, b\}$ is the trivial partition.

Let $f : [a, b] \to \mathbb{R}$ be any bounded function, and define

$m(f) := \inf\{f(x) : x \in [a, b]\}$ and $M(f) := \sup\{f(x) : x \in [a, b]\}$.

Given a partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$, let us define

$$m_i(f) := \inf\{f(x) : x \in [x_{i-1}, x_i]\}$$

and

$$M_i(f) := \sup\{f(x) : x \in [x_{i-1}, x_i]\}$$

for $i = 1, \ldots, n$.

Lower sum and upper sum

Clearly, $m(f) \le m_i(f) \le M_i(f) \le M(f)$ for $i = 1, \ldots, n$. Let

$$L(P, f) := \sum_{i=1}^{n} m_i(f)(x_i - x_{i-1}),$$

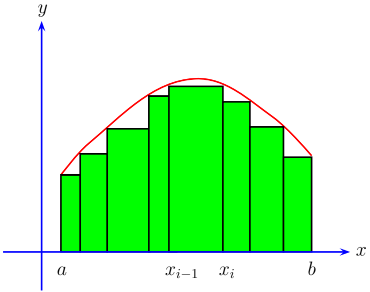
$$U(P, f) := \sum_{i=1}^{n} M_i(f)(x_i - x_{i-1}).$$

$L(P, f)$ is called the **lower sum** of f with respect to P and $U(P, f)$ is called the **upper sum** of f with respect to P.

Since $\sum_{i=1}^{n}(x_i - x_{i-1}) = b - a$, we obtain

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

Underlying assumption: The area of a rectangle $[x_{i-1}, x_i] \times [y_{i-1}, y_i]$ is equal to $(x_i - x_{i-1})(y_i - y_{i-1})$.



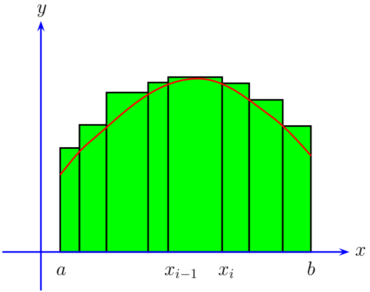


Figure: Approximating the 'area' under a curve by a lower sum which is sum of the areas of the inscribed rectangles, and by an upper sum which is the sum of the areas of the circumscribing rectangles.

MA 105 : Calculus
Division 1, Lecture 8

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Infinite limits of sequences
- Limits of functions at infinity
- Infinite limits of functions
- Asymptotes: Horizontal, Vertical and Oblique
- Even, odd and periodic functions
- Curve Sketching: Tips and examples
- Mean Value Theorem, revisited
- Extended Mean Value Theorem, and its proof
- Taylor theorem, and its proof
- Taylor formula, featuring Taylor polynomial and Lagrange (form of) remainder
- Example: $f(x) = \sin x$. Taylor series for this function around 0.
- Brief discussion of area of planar regions and Riemann integration

Riemann Integration

Let $a, b \in \mathbb{R}$ with $a < b$, and let f be a bounded function defined on $[a, b]$. Let us first assume that $f \geq 0$ on $[a, b]$. We are interested in assigning a meaning to the concept of 'area' of the region that lies under the graph of f, between the lines $x = a$, $x = b$, and above the x-axis, that is, of the region

$$R_f := \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\}.$$

A **partition** of $[a, b]$ is a finite ordered set $P := \{x_0, x_1, \ldots, x_n\}$ of points in $[a, b]$ such that $a = x_0 < x_1 < x_2 < \cdots < x_n = b$, where $n \in \mathbb{N}$.

The points in a partition divide the interval $[a, b]$ into nonoverlapping subintervals, $[x_0, x_1], \ldots, [x_{n-1}, x_n]$. We call $[x_{i-1}, x_i]$ the ith **subinterval** of the partition P.

Example: For $n \in \mathbb{N}$, the set

$$\mathrm{P}_n := \left\{a,\ a + \frac{(b-a)}{n}, \ldots, a + \frac{(n-1)(b-a)}{n},\ b\right\}$$

is a partition of $[a, b]$ into n equal parts, each of length $(b-a)/n$. If $n := 1$, then $\mathrm{P}_1 := \{a, b\}$ is the trivial partition.

Let $f : [a, b] \to \mathbb{R}$ be any bounded function, and define

$m(f) := \inf\{f(x) : x \in [a, b]\}$ and $M(f) := \sup\{f(x) : x \in [a, b]\}$.

Given a partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$, let us define

$$m_i(f) := \inf\{f(x) : x \in [x_{i-1}, x_i]\}$$

and

$$M_i(f) := \sup\{f(x) : x \in [x_{i-1}, x_i]\}$$

for $i = 1, \ldots, n$.

Lower sum and upper sum

Clearly, $m(f) \leq m_i(f) \leq M_i(f) \leq M(f)$ for $i = 1, \ldots, n$. Let

$$L(P, f) := \sum_{i=1}^{n} m_i(f)(x_i - x_{i-1}),$$

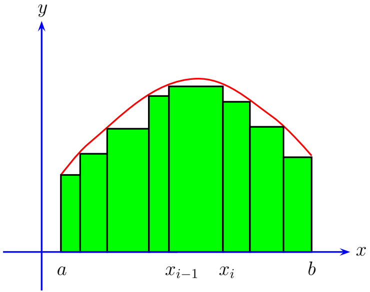
$$U(P, f) := \sum_{i=1}^{n} M_i(f)(x_i - x_{i-1}).$$

$L(P, f)$ is called the **lower sum** of f with respect to P and $U(P, f)$ is called the **upper sum** of f with respect to P.

Since $\sum_{i=1}^{n}(x_i - x_{i-1}) = b - a$, we obtain

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

Underlying assumption: The area of a rectangle $[x_{i-1}, x_i] \times [y_{i-1}, y_i]$ is equal to $(x_i - x_{i-1})(y_i - y_{i-1})$.



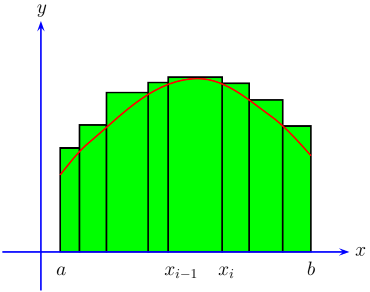


Figure: Approximating the 'area' under a curve by a lower sum which is sum of the areas of the inscribed rectangles, and by an upper sum which is the sum of the areas of the circumscribing rectangles.

Refinement of a partition

Facts:

(i) Let $f : [a, b] \to \mathbb{R}$ be a bounded function. If P is a partition of $[a, b]$ and P^* is a **refinement** of P, that is, if every point of P is also a point of P^*, then

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

Thus the lower sum increases and the upper sum decreases when a partition is refined. (See the figures on the next page. A formal proof is left as an exercise)

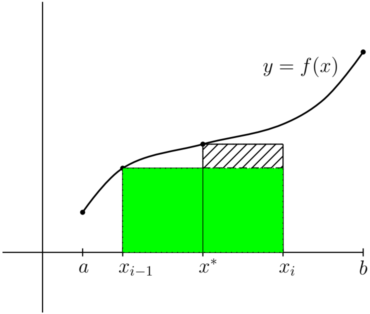
(ii) Let P_1, P_2 be any partitions of $[a, b]$, and let $P^* = P_1 \cup P_2$ be the **common refinement** of P_1 and P_2. Then by (i),

$$L(P_1, f) \leq L(P^*, f) \leq U(P^*, f) \leq U(P_2, f).$$

Thus any lower sum is less than or equal to any upper sum.

Effect of Refinement on Lower sum and Upper sum

To illustrate the effect of a refinement of the partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ on the corresponding lower sum and the upper sum of $f : [a, b] \to \mathbb{R}$, let us add an extra point x^* in the interval (x_{i-1}, x_i). We see that $L(P, f)$ can become bigger, while $U(P, f)$ may become smaller.



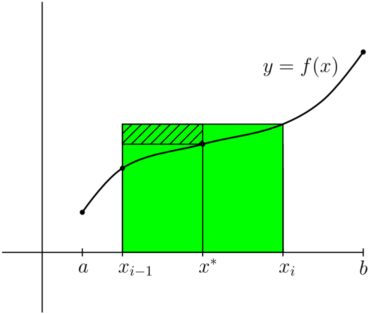


Figure: On the left, the shaded area is added to the lower sum. On the right, the shaded are is removed from the upper sum.

Riemann Integral

Let $a < b$, and $f : [a, b] \to \mathbb{R}$ be a bounded function. Define

$$L(f) := \sup\{L(P, f) : P \text{ is a partition of } [a, b]\},$$
$$U(f) := \inf\{U(P, f) : P \text{ is a partition of } [a, b]\}.$$

$L(f)$ is called the **lower Riemann integral** of f and $U(f)$ is called the **upper Riemann integral** of f. By (ii) above (and the definitions of sup and inf), we see that $L(f) \leq U(f)$.

Definition

A bounded function $f : [a, b] \to \mathbb{R}$ is said to be **(Riemann) integrable** on $[a, b]$ if $L(f) = U(f)$.

In this case, the **Riemann integral** of f on $[a, b]$ is the common value $U(f) = L(f)$, and it is denoted by

$$\int_a^b f \quad \text{or} \quad \int_a^b f(x)dx.$$

Example 1: Let $c \in \mathbb{R}$, and let $f(x) := c$ for all $x \in [a, b]$.

For a partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ and $i = 1, \ldots, n$,

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

and so

$$L(P, f) = U(P, f) = \sum_{i=1}^{n} c(x_i - x_{i-1}) = c(b - a).$$

Hence $L(f) = U(f) = c(b - a)$.

Thus f is integrable, and $\int_a^b f(x)\, dx = c(b - a)$.

Example 2: Define the **Dirichlet function** $f : [a, b] \to \mathbb{R}$ by

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

Then f is a bounded function on $[a, b]$.
Recall: Between any two real numbers, there is a rational number, and also an irrational number.

For a partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ and $i = 1, \ldots, n$,

$$m_i(f) := \inf_{x \in [x_{i-1}, x_i]} f(x) = 0 \;\text{ and }\; M_i(f) := \sup_{x \in [x_{i-1}, x_i]} f(x) = 1,$$

and so $L(P, f) = 0$ and $U(P, f) = b - a$. Hence $L(f) = 0$ and $U(f) = b - a$. Since $L(f) \neq U(f)$, f is not integrable.

Example 3: Let $f(x) := x$ for $x \in [a, b]$.

For a partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ and $i = 1, \ldots, n$,

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

and so

$$L(P, f) = \sum_{i=1}^{n} x_{i-1}(x_i - x_{i-1}) \text{ and } U(P, f) = \sum_{i=1}^{n} x_i(x_i - x_{i-1}).$$

Now $U(P, f) - L(P, f) = \sum_{i=1}^{n}(x_i - x_{i-1})^2$ and

$U(P, f) + L(P, f) = \sum_{i=1}^{n}(x_i^2 - x_{i-1}^2) = x_n^2 - x_0^2 = b^2 - a^2$.

Hence

$U(P, f) = (b^2 - a^2)/2 + \sum_{i=1}^{n}(x_i - x_{i-1})^2/2$, and

$L(P, f) = (b^2 - a^2)/2 - \sum_{i=1}^{n}(x_i - x_{i-1})^2/2$.

Let $n \in \mathbb{N}$, and let P_n denote the partition of $[a, b]$ in n equal parts. Then $x_i - x_{i-1} = (b-a)/n$ for $i = 1, \ldots, n$, and so

$$U(\mathrm{P}_n, f) = \frac{b^2 - a^2}{2} + \frac{1}{2}\sum_{i=1}^{n} \frac{(b-a)^2}{n^2} = \frac{b^2 - a^2}{2} + \frac{(b-a)^2}{2n},$$

$$L(\mathrm{P}_n, f) = \frac{b^2 - a^2}{2} - \frac{1}{2}\sum_{i=1}^{n} \frac{(b-a)^2}{n^2} = \frac{b^2 - a^2}{2} - \frac{(b-a)^2}{2n}.$$

Since $L(\mathrm{P}_n, f) \le L(f)$ and $U(f) \le U(\mathrm{P}_n, f)$, we obtain

$$\frac{b^2 - a^2}{2} - \frac{(b-a)^2}{2n} \le L(f) \le U(f) \le \frac{b^2 - a^2}{2} + \frac{(b-a)^2}{2n}.$$

Letting $n \to \infty$, we see that $L(f) = U(f) = (b^2 - a^2)/2$.

Thus f is integrable, and $\int_a^b f(x)\,dx = (b^2 - a^2)/2$.

A useful criterion for integrability

The above example illustrates the difficulty in proving the integrability of a bounded function f on $[a, b]$ by showing $U(f) = L(f)$. We now give a necessary and sufficient condition for the integrability of such a function, which is much easier to verify.

Theorem (Riemann condition)

Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then f is integrable on $[a, b]$ if and only if for every $\epsilon > 0$, there is a partition P_ϵ of $[a, b]$ such that $U(P_\epsilon, f) - L(P_\epsilon, f) < \epsilon$.

Proof: ($\Longrightarrow$) Let $\epsilon > 0$. There are partitions P_1, P_2 such that

$$U(P_1, f) < U(f) + \frac{\epsilon}{2} \quad \text{and} \quad L(f) - \frac{\epsilon}{2} < L(P_2, f).$$

Let P_ϵ denote the common refinement of P_1 and P_2. Then

$$U(P_\epsilon, f) - L(P_\epsilon, f) \leq U(P_1, f) - L(P_2, f) < U(f) - L(f) + \epsilon = \epsilon.$$

($\Longleftarrow$) Given any $\epsilon > 0$, by the Riemann condition, we obtain

$$0 \leq U(f) - L(f) \leq U(P_\epsilon, f) - L(P_\epsilon, f) < \epsilon.$$

Letting $\epsilon \to 0$, we see that $U(f) = L(f)$. □

Corollary

Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Then f is integrable if and only if there is a sequence (P_n) of partitions of $[a, b]$ such that $U(P_n, f) - L(P_n, f) \to 0$. In this case,

$$L(P_n, f) \to \int_a^b f(x)dx \quad \text{and} \quad U(P_n, f) \to \int_a^b f(x)dx.$$

Proof: ($\Longrightarrow$) For $n \in \mathbb{N}$, let $\epsilon = 1/n$, and use the Riemann condition to find a partition P_n of $[a, b]$ such that

$$U(P_n, f) - L(P_n, f) < 1/n.$$

($\Longleftarrow$) Given $\epsilon > 0$, find $n_0 \in \mathbb{N}$ such that

$$U(P_n, f) - L(P_n, f) < \epsilon \quad \text{for all } n \geq n_0.$$

Then the Riemann condition is satisfied with $P_\epsilon := P_{n_0}$.

Next, let (P_n) be a sequence of partitions such that $U(P_n, f) - L(P_n, f) \to 0$. Then for $n \in \mathbb{N}$,

$$0 \le L(f) - L(P_n, f) \le U(f) - L(P_n, f) \le U(P_n, f) - L(P_n, f),$$

and hence $L(P_n, f) \to L(f) = \int_a^b f(x)dx$. Similarly, for $n \in \mathbb{N}$,

$$0 \le U(P_n, f) - U(f) \le U(P_n, f) - L(f) \le U(P_n, f) - L(P_n, f),$$

and hence $U(P_n, f) \to U(f) = \int_a^b f(x)dx$. $\square$

Example: Fix $m \in \mathbb{N}$, and let $f(x) := x^m$ for $x \in [a, b]$. For $n \in \mathbb{N}$, let P_n be the partition of $[a, b]$ in n equal parts. Then $x_i - x_{i-1} = (b - a)/n$ for $i = 1, \ldots, n$. If m is odd, or if m is even and $a \ge 0$, then

$$U(\mathrm{P}_n, f) - L(\mathrm{P}_n, f) = \sum_{i=1}^{n} (x_i^m - x_{i-1}^m) \frac{(b-a)}{n} = \frac{(b-a)}{n}(b^m - a^m) \to 0,$$

and so f is integrable. (What if m is even and $a < 0$?)

In particular, if $m := 1$, then

$$\begin{aligned} U(\mathrm{P}_n, f) &= \frac{b-a}{n}\sum_{i=1}^{n}\left(a + i\,\frac{b-a}{n}\right) \\ &= a(b-a) + \frac{(b-a)^2}{n^2}\frac{n(n+1)}{2} \\ &\to a(b-a) + \frac{(b-a)^2}{2} = \frac{b^2-a^2}{2}, \end{aligned}$$

and so $\int_a^b x\,dx = (b^2-a^2)/2$, as we have already seen.

Similarly, if $m := 2$ and $a \geq 0$, then

$$\begin{aligned} U(\mathrm{P}_n, f) &= \frac{b-a}{n}\sum_{i=1}^{n}\left(a + i\,\frac{b-a}{n}\right)^2 \\ &\to a^2(b-a) + a(b-a)^2 + \frac{(b-a)^3}{3} = \frac{b^3-a^3}{3}. \end{aligned}$$

and so $\int_a^b x^2 dx = (b^3-a^3)/3$.

Results about Riemann integration

(Domain Additivity)

Let $f : [a, b] \to \mathbb{R}$ be a bounded function, and let $c \in (a, b)$. Then f is integrable on $[a, b]$ if and only if f is integrable on $[a, c]$ and $[c, b]$. In this case, $\int_a^b f = \int_a^c f + \int_c^b f$.

A proof can be given using the Riemann condition. We omit the details. See [GL-1] if interested.

In defining $\int_a^b f$, we have assumed $a < b$ so far.

Conventions: (i) Let $b = a$. Then every $f : [a, a] \to \mathbb{R}$ is integrable, and $\int_a^a f(x)\, dx := 0$.

(ii) Let $b < a$ and $f : [b, a] \to \mathbb{R}$ be integrable. Then $\int_a^b f(x)\, dx := -\int_b^a f(x)\, dx$.

It follows that $\int_c^d f = \int_a^d f - \int_a^c f$ for any $c, d \in [a, b]$.

Integrable functions

(i) If $f : [a, b] \to \mathbb{R}$ is monotonic, then f is integrable.

(ii) If $f : [a, b] \to \mathbb{R}$ is bounded and has at most a finite number of discontinuities in $[a, b]$, then f is integrable.

Proofs are a bit involved, especially for (ii), and thus skipped. See [GL-1] is interested.

Examples:

(i) Let $f(x) := [x]$ for $x \in [a, b]$. Since f is increasing on $[a, b]$, f is integrable on $[a, b]$.

(ii) Define $f : [0, 1] \to \mathbb{R}$ by $f(0) := 0$ and $f(x) := 1/n$ if $1/(n + 1) < x \le 1/n$ for $n \in \mathbb{N}$. Since f is increasing on $[0, 1]$, f is integrable on $[a, b]$.

(iii) A polynomial function is integrable on $[a, b]$.

(iv) Define $f : [0, 1] \to \mathbb{R}$ by $f(0) := 0$ and $f(x) := \sin(1/x)$ for $x \in (0, 1]$. Since f is bounded, and it is discontinuous only at 0, it follows that f is integrable on $[a, b]$.

Algebraic and Order Properties

Let $f, g : [a, b] \to \mathbb{R}$ be integrable functions. Then

(i) $f + g$ is integrable, and $\int_a^b (f + g) = \int_a^b f + \int_a^b g$.

(ii) $c\, f$ is integrable, and $\int_a^b c\, f = c \int_a^b f$ for $c \in \mathbb{R}$.

(iii) $f \cdot g$ is integrable.

(iv) If there is $\delta > 0$ such that $|f(x)| \geq \delta$ for all $x \in [a, b]$ (so that $1/f$ is bounded), then $1/f$ is integrable.

(v) If $f \leq g$, then $\int_a^b f \leq \int_a^b g$.

(vi) $|f|$ is integrable, and $|\int_a^b f| \leq \int_a^b |f|$.

Proving (ii) is easy. Proofs of (i), (iii) and (iv) are omitted. For (v), note that $U(P, f) \leq U(P, g)$ for any partition P of $[a, b]$, and so $\int_a^b f = U(f) \leq U(g) = \int_a^b g$. Finally, (vi) follows since $U(P, |f|) - L(P, |f|) \leq U(P, f) - L(P, f)$ for each P, and also since $-|f| \leq f \leq |f| \implies -\int_a^b |f| \leq \int_a^b f \leq \int_a^b |f|$.

Area of planar regions

Suppose $f : [a, b] \to \mathbb{R}$ is a bounded function. If $f \geq 0$, i.e., if $f(x) \geq 0$ for all $x \in [a, b]$, then we say that the planar region

$$R_f := \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\}.$$

has an area if f is integrable on $[a, b]$, and in this case, define

$$\mathrm{Area}(R_f) = \int_a^b f(x)dx.$$

Note that this is a nonnegative real number, thanks to (v). Note also that if $f, g : [a, b] \to \mathbb{R}$ are integrable on $[a, b]$, then so are $\max(f, g)$ and $\min(f, g)$. To see this, observe that

$$\max(f, g) = \frac{f + g + |f - g|}{2} \text{ and } \min(f, g) = \frac{f + g - |f - g|}{2}.$$

It follows that in general, if $f : [a, b] \to \mathbb{R}$ is integrable, then $\int_a^b f(x)dx = \mathrm{Area}(R_{f^+}) - \mathrm{Area}(R_{f^-})$, where $f^+ = \max\{f, 0\}$ and $f^- = -\min\{f, 0\}$. Thus $\int_a^b f(x)dx$ is like "**signed area**".

The Fundamental Theorem of Calculus (FTC)

(The Fundamental Theorem of Calculus: Part I)

Let f be integrable on $[a, b]$. For $x \in [a, b]$, define

$$F(x) := \int_a^x f(t)\, dt.$$

Then F is continuous on $[a, b]$. Moreover, if f is continuous at $c \in [a, b]$, then F is differentiable at c, and $F'(c) = f(c)$.

Proof: Let $\alpha \in \mathbb{R}$ satisfy $|f| \leq \alpha$ on $[a, b]$, and let $c \in [a, b]$. By the domain additivity of the Riemann integral,

$$|F(x) - F(c)| = \left| \int_c^x f(t)\, dt \right| \leq \alpha |x - c| \quad \text{for all } x \in [a, b].$$

If $x_n \to c$, then $F(x_n) \to F(c)$. Hence F is continuous at c.

MA 105 : Calculus
Division 1, Lecture 9

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Riemann integration
    - Partition of an interval, Subintervals of a partition
    - Upper sum and Lower sum of a function w.r.t a partition
    - Upper integral and Lower integral
    - Definition of integrable functions and of the Riemann integral
    - Example of integrable and non-integrable functions
- Results about Riemann integration
    - Criterion for integrability: Riemann condition. Corollary involving sequences of partitions
    - Domain additivity
    - (Classes of) integrable functions
    - Algebraic and Order Properties
    - Fundamental Theorem of Calculus: Statement of Part I

The Fundamental Theorem of Calculus (FTC)

(The Fundamental Theorem of Calculus: Part I)

Let f be integrable on $[a, b]$. For $x \in [a, b]$, define

$$F(x) := \int_a^x f(t)\, dt.$$

Then F is continuous on $[a, b]$. Moreover, if f is continuous at $c \in [a, b]$, then F is differentiable at c, and $F'(c) = f(c)$.

Proof: Let $\alpha \in \mathbb{R}$ satisfy $|f| \leq \alpha$ on $[a, b]$, and let $c \in [a, b]$. By the domain additivity of the Riemann integral,

$$|F(x) - F(c)| = \left| \int_c^x f(t)\, dt \right| \leq \alpha |x - c| \quad \text{for all } x \in [a, b].$$

If $x_n \to c$, then $F(x_n) \to F(c)$. Hence F is continuous at c.

Suppose f is continuous at $c \in [a, b]$. For $x \in [a, b]$, $x \neq c$, define $F_1(x) := \big(F(x) - F(c)\big)/(x - c)$, and $F_1(c) := f(c)$.

We show that the function F_1 is continuous at c. For $x \neq c$,

$$F_1(x) - f(c) = \left(\frac{1}{x-c}\int_c^x f(t)dt\right) - f(c) = \frac{1}{x-c}\int_c^x \big(f(t) - f(c)\big)dt$$

since $f(c) = \int_c^x f(c)dt/(x - c)$. Let $\epsilon > 0$. Since f is continuous at c, there exists $\delta > 0$ with

$$t \in [a, b] \quad \text{and} \quad |t - c| < \delta \implies |f(t) - f(c)| < \epsilon.$$

Hence if $x \in [a, b]$ and $0 < |x - c| < \delta$, then

$$|F_1(x) - f(c)| = \frac{\left|\int_c^x \big(f(t) - f(c)\big)dt\right|}{|x - c|} \leq \frac{\epsilon\,|x - c|}{|x - c|} = \epsilon.$$

By the C-lemma, F' exists at c and $F'(c) = F_1(c) = f(c)$. $\square$

Note: Suppose f is continuous on $[a, b]$. Then F is differentiable on $[a, b]$ and $F'(x) = f(x)$ for $x \in [a, b]$, that is,

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x) \quad \text{for all } x \in [a, b].$$

Thus F is in fact **continuously differentiable** on $[a, b]$.

Examples:

(i) Let $f(x) := [x]$ for $x \in [-1, 1]$. Define $F(x) := \int_{-1}^{x}[t]dt$ for $x \in [-1, 1]$. Check that $F(x) = -1 - x$ if $x \in [-1, 0]$, and $F(x) = F(0) = -1$ if $x \in (0, 1]$. Note: f is not continuous on $[-1, 1]$, but F is continuous on $[-1, 1]$.

(ii) Let $f(x) := |x|$ for $x \in [-1, 1]$. Define $F(x) := \int_{-1}^{x}|t|dt$ for $x \in [-1, 1]$. Check that $F(x) = \int_{-1}^{x}(-t)dt = (1 - x^2)/2$ if $x \in [-1, 0]$, and $F(x) = F(0) + \int_0^x t\,dt = (1 + x^2)/2$ if $x \in (0, 1]$. Note: f is not differentiable on $[-1, 1]$, but F is differentiable; in fact, $F'(x) = |x| = f(x)$ for $x \in [-1, 1]$.

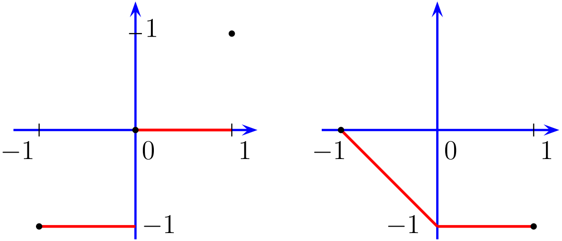


Figure: $y = [x]$ and $y = \int_{-1}^{x} [t]dt$

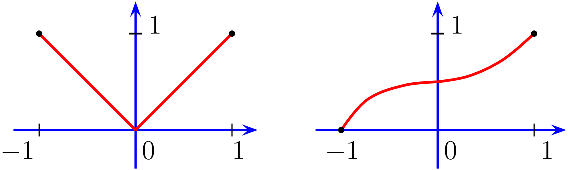


Figure: $y = |x|$ and $y = \int_{-1}^{x} |t|dt$

Antiderivative

Definition

Let I be an interval containing more than one point, and consider a function $f : I \to \mathbb{R}$. We say that f has an antiderivative on I if there is a differentiable function $F : I \to \mathbb{R}$ such that $F' = f$. Such a function F is called an antiderivative or a primitive of f, or an indefinite integral of f.

Lemma: If I is an interval and $f : I \to \mathbb{R}$ has an antiderivative F, then F is unique up to addition by a constant.

Proof: If $F' = f = G'$, then $(F - G)' = 0$ on the interval I, and hence $F - G$ is a constant function (by the MVT). $\square$

As a consequence of the FTC (Part I), we see that if $f : [a, b] \to \mathbb{R}$ is continuous, then f has an antiderivative given by $F(x) := \int_a^x f(t)dt$ for $x \in [a, b]$. (Note: $F(a) = 0$.)

(The Fundamental Theorem of Calculus: Part II)

Let $f : [a, b] \to \mathbb{R}$ be a differentiable function such that f' is integrable on $[a, b]$. Then

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

Proof: Let $P := \{x_0, x_1, \ldots, x_n\}$ be a partition of $[a, b]$. By the MVT, there is $c_i \in (x_{i-1}, x_i)$ for $i = 1, \ldots, n$ such that

$$f(b) - f(a) = \sum_{i=1}^{n} \big(f(x_i) - f(x_{i-1})\big) = \sum_{i=1}^{n} f'(c_i)(x_i - x_{i-1}).$$

Hence $L(P, f') \le f(b) - f(a) \le U(P, f')$ for every partition P of $[a, b]$, and so $\int_a^b f' = L(f') \le f(b) - f(a) \le U(f') = \int_a^b f'$. Thus $\int_a^b f' = f(b) - f(a)$. $\square$

Compare the conclusions of the MVT and the FTC, Part II:

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

What does the Fundamental Theorem of Calculus say?

(i) Part I: If f is continuous on $[a, b]$, then

$$\frac{d}{dx}\int_a^x f(t)\,dt = f(x) \quad \text{for all } x \in [a, b].$$

(ii) Part II: If f' exists and is integrable on $[a, b]$, then

$$\int_a^b \left(\frac{d}{dx}f(x)\right)dx = f(b) - f(a).$$

Thus the operations of differentiation and integration are 'inverse' to each other.

Side Remark: Some Pathologies

This optional slide (which was not shown during the lecture) elaborates on some remarks I made and also contains answers to some of the questions asked by the students during or after the class. The material here will not have bearing on any of the exams.

- An integrable function need not have an antiderivative. In fact, if $f : [a, b] \to \mathbb{R}$ has an antiderivative, then f has the IVP on $[a, b]$ (see, e.g, Theorem 4.16 in [GL-1]). In particular, $f : [-1, 1] \to \mathbb{R}$ given by $f(x) = [x]$ does not have an antiderivative, although f is clearly integrable.

- If $f : [a, b] \to \mathbb{R}$ has an antiderivative F, then f need not be integrable. In fact, f need not even be bounded. Example: $f = F'$ with $F : [-1, 1] \to \mathbb{R}$ given by $F(x) = x^2 \sin(1/x^2)$ for $x \neq 0$ and $F(0) = 0$ [See also Remark 6.25 (iii) in [GL-1].]

- Composites $g \circ f$ of integrable functions $f : [a, b] \to [c, d]$ and $g : [c, d] \to \mathbb{R}$ need not be integrable, unless g is continuous. See, e.g., Example 6.17 and Remark 6.18 in [GL-1].

Consequences of the FTC

(Integration by parts)

Let $f : [a, b] \to \mathbb{R}$ be a differentiable function such that f' is integrable. Suppose $g : [a, b] \to \mathbb{R}$ is integrable and has an antiderivative G on $[a, b]$. Then

$$\begin{aligned}\int_a^b f(x)g(x)dx &= f(b)G(b) - f(a)G(a) - \int_a^b f'(x)G(x)dx \\ &= f(x)G(x)\Big|_{x=a}^{x=b} - \int_a^b f'(x)G(x)\, dx.\end{aligned}$$

Proof: For $x \in [a, b]$, define $H(x) := f(x)G(x)$, so that $H'(x) = f(x)G'(x) + f'(x)G(x) = f(x)g(x) + f'(x)G(x)$. Then H' is integrable on $[a, b]$. Hence by the FTC (Part II),

$$\int_a^b H'(x)dx = H(b) - H(a) = f(b)G(b) - f(a)G(a). \quad \square$$

(Integration by substitution)

Let $f : [a, b] \to \mathbb{R}$ be continuous, and let $\phi : [\alpha, \beta] \to \mathbb{R}$ be a continuously differentiable function such that $\phi([\alpha, \beta]) = [a, b]$. Then $(f \circ \phi)\phi' : [\alpha, \beta] \to \mathbb{R}$ is integrable, and

$$\int_{\phi(\alpha)}^{\phi(\beta)} f(x)dx = \int_{\alpha}^{\beta} f\big(\phi(t)\big)\phi'(t)dt.$$

In particular, if $\phi'(t) \neq 0$ for all $t \in (\alpha, \beta)$, then

$$\int_{a}^{b} f(x)dx = \int_{\alpha}^{\beta} f(\phi(t))|\phi'(t)|dt.$$

Proof: Define $F : [a, b] \to \mathbb{R}$ by $F(x) := \int_a^x f(u)du$ for $x \in [a, b]$, and $H : [\alpha, \beta] \to \mathbb{R}$ by $H(t) := F(\phi(t))$ for $t \in [\alpha, \beta]$.

By the chain rule and by the FTC (Part I),

$$H'(t) = F'(\phi(t))\phi'(t) = f(\phi(t))\phi'(t) \quad \text{for } t \in [\alpha, \beta].$$

Then H' is integrable on $[\alpha, \beta]$. Hence by the FTC (Part II),

$$\int_{\alpha}^{\beta} H'(t)dt = H(\beta) - H(\alpha) = F(\phi(\beta)) - F(\phi(\alpha)) = \int_{\phi(\alpha)}^{\phi(\beta)} f(u)du,$$

that is,

$$\int_{\phi(\alpha)}^{\phi(\beta)} f(x)dx = \int_{\alpha}^{\beta} f\big(\phi(t)\big)\phi'(t)dt.$$

Now suppose $\phi'(t) \neq 0$ for all $t \in (\alpha, \beta)$. By the Intermediate Value Property of ϕ', either $\phi'(t) > 0$ for all $t \in (\alpha, \beta)$, or $\phi'(t) < 0$ for all $t \in (\alpha, \beta)$. In the former case, ϕ is strictly increasing on $[\alpha, \beta]$, $\phi(\alpha) = a$, $\phi(\beta) = b$ and $|\phi'| = \phi'$. In the latter case, ϕ is strictly decreasing on $[\alpha, \beta]$, $\phi(\alpha) = b$, $\phi(\beta) = a$ and $|\phi'| = -\phi'$. Hence the desired result follows. $\square$

Examples:

(i) To evaluate $\int_0^1 x\sqrt{1-x}\,dx$, let $f(x) := x$ and $g(x) := \sqrt{1-x}$ for $x \in [0,1]$. Then $f'(x) = 1$ for $x \in [0,1]$. Also, if $G(x) := -(2/3)(1-x)^{3/2}$, then $G'(x) = g(x)$ for $x \in [0,1]$, that is, $G' = g$. Integration by Parts yields

$$\int_0^1 x\sqrt{1-x}\,dx = 0-0-\int_0^1 \left(-\frac{2}{3}\right)(1-x)^{3/2}dx = \frac{2}{3}\int_0^1 (1-x)^{3/2}dx.$$

If we let $F(x) := -(2/5)(1-x)^{5/2}$ for $x \in [0,1]$, then $F'(x) = (1-x)^{3/2}$ for $x \in [0,1]$. By the FTC, Part II, the integral equals $(2/3)(F(1)-F(0)) = (2/3)(2/5) = 4/15$.

(ii) To evaluate $\int_0^1 t\sqrt{1-t^2}\,dt$, let $\phi(t) := 1-t^2$ for $t \in [0,1]$ and $f(x) := \sqrt{x}$ for $x \in [0,1]$. Then $\phi(0) = 1$, $\phi(1) = 0$, and $\phi'(t) = -2t$ for $t \in [0,1]$. Let $F(x) := (2/3)x^{3/2}, x \in [0,1]$. Then $F'(x) = \sqrt{x}$, $x \in [0,1]$. Integration by Substitution yields

$$\int_0^1 t\sqrt{1-t^2}\,dt = \frac{1}{2}\int_0^1 f(\phi(t))|\phi'(t)|dt = \frac{1}{2}\int_0^1 \sqrt{x}\,dx = \frac{1}{3}.$$

(Approximate) Riemann Sum

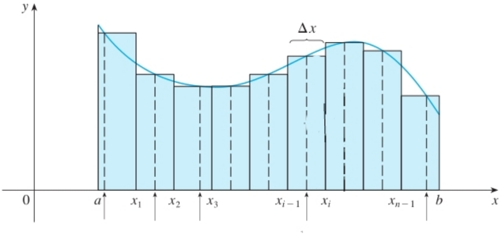
Given a bounded function $f : [a, b] \to \mathbb{R}$ and a partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$, the sum

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

is called a Riemann sum for f corresponding to P, where $t_i \in [x_{i-1}, x_i]$, $i = 1, \ldots, n$.

Since $m_i(f) \leq f(t_i) \leq M_i(f)$ for $i = 1, \ldots, n$, we note that

$$L(P, f) \leq S(P, f) \leq U(P, f) \quad \text{for any } t_i \in [x_{i-1}, x_i].$$



Approximation of a Riemann Integral

The **mesh** of a partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ is defined by $\mu(P) := \max\{x_i - x_{i-1} : i = 1, \ldots, n\}$.

Theorem: Let f be integrable on $[a, b]$, and let $\epsilon > 0$ be given. Then there is $\delta > 0$ such that $U(P, f) - L(P, f) < \epsilon$ for every partition P satisfying $\mu(P) < \delta$.

This improves upon the Riemann condition. Proof is omitted.

Corollary. Let f be integrable on $[a, b]$, and let (P_n) be a sequence of partitions of $[a, b]$ such that $\mu(P_n) \to 0$. Then $U(P_n, f) - L(P_n, f) \to 0$. Further, if $S(P_n, f)$ is a Riemann sum corresponding to P_n and f, then $S(P_n, f) \to \int_a^b f(x)dx$.

Proof: Let $\epsilon > 0$ be given, and let $\delta > 0$ be as in the above theorem. For this δ, we find $n_0 \in \mathbb{N}$ such that $\mu(P_n) < \delta$ for all $n \geq n_0$. Then $U(P_n, f) - L(P_n, f) < \epsilon$ for all $n \geq n_0$. Thus $U(P_n, f) - L(P_n, f) \to 0$.

Since

$$
\begin{aligned}
L(P_n, f) &\leq S(P_n, f) \leq U(P_n, f) \quad \text{and} \\
L(P_n, f) &\leq L(f) = \int_a^b f(x)dx = U(f) \leq U(P_n, f),
\end{aligned}
$$

we see that

$$
\left| S(P_n, f) - \int_a^b f(x)dx \right| \leq U(P_n, f) - L(P_n, f) \to 0. \quad \square
$$

The above corollary is useful in two ways.

(i) If $\int_a^b f(x)dx$ is known, then we can find $\lim_{n\to\infty} a_n$, if $a_n := S(P_n, f)$ for $n \in \mathbb{N}$ and $\mu(P_n) \to 0$.

(ii) If $\int_a^b f(x)dx$ is not known, then $S(P_n, f)$ gives us an approximation of $\int_a^b f(x)dx$ if $\mu(P_n) \to 0$.

Suppose $f : [a, b] \to \mathbb{R}$ is known to be an integrable function. This is the case, for example, if f is monotonic on $[a, b]$, or if f is discontinuous at only a finite number of points in $[a, b]$.

How can we find the Riemann integral $\int_a^b f(x)dx$?

Try to find an antiderivative of f on $[a, b]$, that is, find a function F on $[a, b]$ such that $F' = f$ on $[a, b]$. In that case, $\int_a^b f(x)dx = \int_a^b F'(x)dx = F(b) - F(a)$ by the FTC (Part II).

For instance, let $r \in \mathbb{Q}$, $r \geq 0$, and $f(x) := x^r$ for $x \in [a, b]$. Let $F(x) := x^{r+1}/(r+1)$ for $x \in [a, b]$. Then $F' = f$ on $[a, b]$. Hence $\int_a^b f(x)dx = F(b) - F(a) = (b^{r+1} - a^{r+1})/(r+1)$.

Next, suppose $0 \notin [a, b]$. Then the above calculation holds if $r \in \mathbb{Q}$ and $r < 0$, provided $r \neq -1$. If $r := -1$, we do not yet know an antiderivative of the function f. Hence we look for an approximation of $\int_a^b f(x)dx$.

Examples:

(i) For $n \in \mathbb{N}$, let $a_n := \dfrac{1^r + 2^r + \cdots + n^r}{n^{r+1}}$, where $r \in \mathbb{Q}$ and $r \geq 0$. Let us find $\lim_{n\to\infty} a_n$. Note that

$$a_n = \sum_{i=1}^{n} \frac{i^r}{n^{r+1}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i}{n}\right)^r = \sum_{i=1}^{n}\left(\frac{i}{n}\right)^r\left(\frac{i}{n} - \frac{i-1}{n}\right).$$

Define $f : [0,1] \to \mathbb{R}$ by $f(x) := x^r$.
Since f is increasing on $[0,1]$, it is integrable. For $n \in \mathbb{N}$, let $P_n := \{0, 1/n, \ldots, n/n\}$ and $t_i := i/n$ for $i = 1, 2, \ldots, n$.
Then $a_n = S(P_n, f)$. Since $\mu(P_n) = 1/n \to 0$, it follows that

$$S(P_n, f) \to \int_0^1 f(x)dx, \text{ that is },$$

$$a_n = \sum_{i=1}^{n} \frac{i^r}{n^{r+1}} \to \int_0^1 x^r\, dx = \frac{1}{r+1}.$$

(ii) Let $f(x) := 1/x$ for $x \in [1, 2]$. (This is the case $r := -1$ mentioned earlier.)

Since f is continuous on $[1, 2]$, it is integrable. For $n \in \mathbb{N}$, let

$$\mathrm{P}_n := \left\{1,\ 1 + \frac{1}{n},\ 1 + \frac{2}{n}, \ldots, 1 + \frac{n-1}{n},\ 2\right\}$$

be the partition of $[1, 2]$ in n equal parts.

Let $t_{n,i} = 1 + (i-1)/n$ for $i = 1, \ldots, n$, so that

$$S(\mathrm{P}_n, f) = \sum_{i=1}^{n} \frac{1}{1 + (i-1)/n} \left(\frac{i}{n} - \frac{i-1}{n}\right) = \sum_{i=1}^{n} \frac{1}{n+i-1}.$$

Since $\mu(\mathrm{P}_n) = 1/n \to 0$, we see that $S(\mathrm{P}_n, f) \to \int_a^b f(x)dx$, that is,

$$\sum_{i=1}^{n} \frac{1}{n+i-1} \to \int_1^2 \frac{1}{x}\, dx \text{ as } n \to \infty.$$

MA 105 : Calculus
Division 1, Lecture 10

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Fundamental Theorem of Calculus: Part I and its proof
- Examples
- Notion of an antiderivative
- Fundamental Theorem of Calculus: Part II and its proof
- Consequences of the FTC
  - Integration by parts
  - Integration by substitution
  - Examples
- (Approximate) Riemann sum
  - Mesh of a partition
  - Convergence of Riemann sums of an integrable function
  - Examples

Logarithmic and Exponential Functions

Using Riemann integration, we can define 'new' functions. The FTC (Part I) allows us to investigate their properties. If $x \geq 1$, then the function $t \longmapsto 1/t$ is continuous on $[1, x]$, and if $0 < x < 1$, it is continuous on $[x, 1]$.

Definition

The **(natural) logarithmic function** defined by

$$\ln x := \int_1^x \frac{1}{t} dt \quad \text{for } x \in (0, \infty).$$

Remark: We have seen that for any $n \in \mathbb{Z}$ with $n \neq -1$, the function $f(t) = t^n$ can easily be integrated on any interval not containing zero. With the introduction of ln as above, every rational function $p(t)/q(t)$ where the polynomial $q(t)$ factors as a product of linear polynomials, can be integrated using the method of partial fractions.

Properties of (natural) logarithmic function $\ln : (0, \infty) \to \mathbb{R}$

(i) $\ln 1 = 0$, $\ln x < 0$ if $x \in (0, 1)$, and $\ln x > 0$ if $x \in (1, \infty)$.
— Clear from the definition.

(ii) ln is differentiable and $(\ln)'(x) = 1/x$ for $x \in (0, \infty)$.
— Follows from the FTC

(iii) ln is strictly increasing and strictly concave on $(0, \infty)$.
— $(\ln)'(x) > 0$ and $(\ln)''(x) = -1/x^2 < 0 \ \forall\, x \in (0, \infty)$.

(iv) $\ln x_1 + \ln x_2 = \ln x_1 x_2$ for all $x_1, x_2 \in (0, \infty)$.
— Let $x_1, x_2 \in (0, \infty)$. By substituting $t = x_1 s$, we obtain

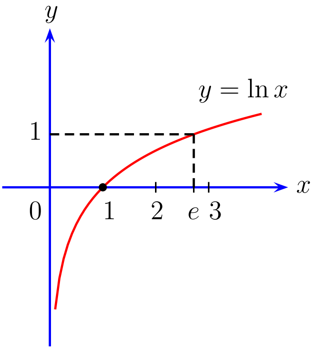
$$\ln x_1 x_2 = \int_1^{x_1 x_2} \frac{dt}{t} = \int_1^{x_1} \frac{dt}{t} + \int_{x_1}^{x_1 x_2} \frac{dt}{t} = \ln x_1 + \ln x_2.$$

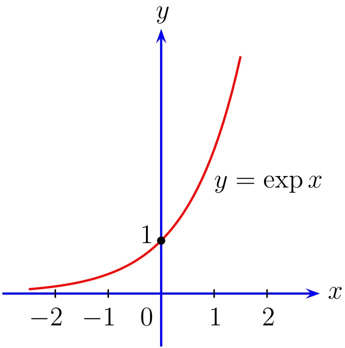
(v) $\ln x \to \infty$ as $x \to \infty$, and $\ln x \to -\infty$ as $x \to 0^+$.
— $\ln 2^n = n \ln 2$ for $n \in \mathbb{N}$ and $\ln(1/x) = -\ln x$.

(vi) $\ln : (0, \infty) \to \mathbb{R}$ is one-one and onto.
— Follows from (iii), (v) and the IVP of the function ln.

The **exponential function** $\exp : \mathbb{R} \to (0, \infty)$ is the inverse of the logarithmic function $\ln : (0, \infty) \to \mathbb{R}$. Thus

$$\exp x = y \iff \ln y = x \quad \text{for } x \in \mathbb{R}.$$





Properties of the exponential function $\exp : \mathbb{R} \to \mathbb{R}$:

(i) $\exp 0 = 1$ and $\exp x > 0$ for every $x \in \mathbb{R}$.

(ii) exp is differentiable and $(\exp)'(x) = \exp x$ for $x \in \mathbb{R}$:
If $x = \ln y$, then $(\exp)'(x) = 1/(\ln)'(y) = y = \exp x$.

(iii) exp is strictly increasing and strictly convex on $\mathbb{R}$.
— $\exp' x > 0$ and $\exp'' x > 0$ for $x \in \mathbb{R}$.

(iv) $\exp(x_1 + x_2) = (\exp x_1)(\exp x_2)$ for all $x_1, x_2 \in \mathbb{R}$.
— If $\exp x_1 = y_1$ and $\exp x_2 = y_2$, then
$\ln(y_1 y_2) = \ln y_1 + \ln y_2 = x_1 + x_2$.

(v) $\exp : \mathbb{R} \to (0, \infty)$ is one-one and onto.

(vi) $\exp x \to \infty$ as $x \to \infty$, and $\exp x \to 0$ as $x \to -\infty$:

Let e denote the unique number in $(0, \infty)$ such that $\ln e = 1$.
Then $2 < e < 4$: $\ln 2 = \int_1^2 (1/t)dt < \int_1^2 1\, dt = 1 = \ln e$, and
$\ln 4 = \int_1^4 (1/t)dt = \int_1^2 (1/t)dt + \int_2^4 (1/t)dt > \frac{1}{2} + \frac{2}{4} = 1 = \ln e$.
Now $\exp n = e^n > 2^n \to \infty$ and $0 < \exp(-n) < 2^{-n} \to 0$ as $n \to \infty$. This together with (iii) yields (vi).

Real powers of a positive real number

Let $a > 0$. Observe that for any $r \in \mathbb{Q}$,

$$\ln a^r = r \ln a \quad \text{and hence} \quad a^r = \exp(r \ln a).$$

To see this, consider $f(x) = \ln x^r - r \ln x$ for $x \in (0, \infty)$.
Then f is differentiable and

$$f'(x) = \frac{1}{x^r} r x^{r-1} - r\frac{1}{x} = 0 \quad \text{for all } x \in (0, \infty),$$

and so $f(x) = f(1) = \ln 1^r - r \ln 1 = 0 - 0 = 0 \; \forall \, x \in (0, \infty)$.

Definition

Let $a > 0$. For $x \in \mathbb{R}$, define

$$a^x := \exp(x \ln a)$$

In particular, $e^x = \exp(x \ln e) = \exp x$ for all $x \in \mathbb{R}$.

We can use the above definition of real powers together with the properties of exponential functions to derive "standard properties" of the power function with a fixed base $a > 0$, i.e., the function

$$f_a : \mathbb{R} \to (0, \infty) \quad \text{defined by} \quad f_a(x) := a^x$$

and also of the power function with a fixed exponent $b \in \mathbb{R}$, i.e., the function

$$g_b : (0, \infty) \to \mathbb{R} \quad \text{defined by} \quad g_b(x) := x^b.$$

In particular, both f_a and g_b are differential functions and

$$f_a'(x) = (\ln a)a^x \quad \text{and} \quad g_b'(x) = bx^{b-1}.$$

Further, f_a is bijective if $a \neq 1$, whereas g_b is bijective if $b \neq 0$. For more details, see Section 7.1 of [GL-1].

Inverse Trigonometric and Trigonometric Functions

In our development of calculus thus far, we have freely used the trigonometric functions such as $\sin, \cos, \tan$, and their basic properties without giving a precise definition of these real-valued functions of a real variable. A self-contained and rigorous definition can be given along the similar lines as that of the logarithmic and exponential functions. This is also partly motivated by algebraic considerations.

Noting that the function given by $t \longmapsto 1/(t^2+1) \in \mathbb{R}$ is continuous on $\mathbb{R}$, we make the following

Definition

The **arctangent function** $\arctan : \mathbb{R} \to \mathbb{R}$ is defined by

$$\arctan x := \int_0^x \frac{1}{1+t^2} dt \quad \text{for } x \in \mathbb{R}.$$

Properties of the Arctangent Function

From the definition of $\arctan : \mathbb{R} \to \mathbb{R}$, the following properties are easily derived.

- $\arctan 0 = 0$,
- $\arctan x \geq 0$ if $x > 0$, while $\arctan x \leq 0$ if $x < 0$.
- arctan is a differentiable function on $\mathbb{R}$ and

$$(\arctan)'x = \frac{1}{1+x^2} \quad \text{for every } x \in \mathbb{R}.$$

- arctan is strictly increasing on $\mathbb{R}$, strictly convex on $(-\infty, 0)$, strictly concave on $(0, \infty)$, and 0 is a point of inflection for arctan.
- arctan is an odd function.

Theorem

$\arctan : \mathbb{R} \to \mathbb{R}$ *is a bounded function. In fact,*

$$-2 < \arctan x < 2 \quad \textit{for all } x \in \mathbb{R}.$$

Boundedness of the Arctangent Function

For $x \in (1, \infty)$,

$$\arctan x = \int_0^1 \frac{1}{1+t^2}\,dt + \int_1^x \frac{1}{1+t^2}\,dt.$$

Since $1 \geq t^2$ for $t \in [0,1]$, while $t^2 \geq 1$ for $t \in [1, x]$,

$$\int_0^1 \frac{1}{1+1}\,dt + \int_1^x \frac{1}{t^2+t^2}\,dt \leq \arctan x \leq \int_0^1 \frac{1}{1}\,dt + \int_1^x \frac{1}{t^2}\,dt.$$

The definite integrals above are easy to evaluate, and thus

$$1 - \frac{1}{2x} \leq \arctan x \leq 2 - \frac{1}{x} \quad \text{for all } x \in (1, \infty).$$

Since the function arctan is strictly increasing and odd, it follows that $-2 < \arctan x < 2$ for all $x \in \mathbb{R}$. $\square$

Definition of π

We can now give a self-contained definition of π without getting into vicious circles! By the above results, there is a well-defined real number $\pi \leq 4$ such that

$$\pi = 2 \sup\{\arctan x : x \in (0, \infty)\}.$$

Since arctan is increasing and since $\arctan x \geq 1 - (1/2x)$ for all $x \in (1, \infty)$, we see that

$$\frac{\pi}{2} = \sup\{\arctan x : x \in (1, \infty)\} \geq \sup\left\{1 - \frac{1}{2x} : x \in (0, \infty)\right\} = 1.$$

Thus the real number π satisfies $2 \leq \pi \leq 4$.

With this definition, it can be seen that arctan maps $\mathbb{R}$ bijectively onto $(-\pi/2, \pi/2)$ and $\arctan 1 = \pi/4$. Thus, the inverse function $\tan : (-\pi/2, \pi/2) \to \mathbb{R}$ is well-defined and it can easily be shown that this has the "standard properties".

Sine and Cosine Functions

To begin with, we define the sine function and the cosine function on the interval $(-\pi/2, \pi/2)$ by

$$\sin x := \frac{\tan x}{\sqrt{1+\tan^2 x}} \quad \text{and} \quad \cos x := \frac{1}{\sqrt{1+\tan^2 x}}$$

It is clear from the definition that

$$\tan x = \frac{\sin x}{\cos x} \quad \text{for } x \in (-\pi/2, \pi/2).$$

Further, the "standard properties" of the tangent function imply the "standard properties" of the sine and cosine functions. Next, define

$$\sin(\pi/2) := 1 \quad \text{and} \quad \cos(\pi/2) := 0.$$

Finally, extend sin and cos to $\mathbb{R}$ by requiring

$$\sin(x+\pi) := -\sin x \quad \text{and} \quad \cos(x+\pi) := -\cos x \quad \text{for } x \in \mathbb{R}.$$

For more on this approach, see Section 7.2 of [GL-1].

Area under a Curve

Let $f : [a, b] \to \mathbb{R}$ be a bounded function. Suppose $f \geq 0$ on $[a, b]$, and let

$$R_f := \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b \text{ and } 0 \leq y \leq f(x)\}.$$

We say that R_f has an **area** if f is Riemann integrable, and then we define

$$\text{Area}\,(R_f) := \int_a^b f(x)dx.$$

If $f : [a, b] \to \mathbb{R}$ is any function, then $f = f^+ - f^-$, where

$$f^+ := \frac{|f| + f}{2} \quad \text{and} \quad f^- := \frac{|f| - f}{2}.$$

Note that $f^+ \geq 0$ and $f^- \geq 0$.

Positive and Negative Parts of a Function

In fact, for $x \in [a, b]$,

$$f^+(x) = \max\{f(x), 0\} \text{ and } f^-(x) = -\min\{f(x), 0\}.$$

The functions f^+ and f^- are known as the **positive part** of f and the **negative** part of f, respectively. Clearly, f is bounded if and only if f^+ and f^- are both bounded. Also, f is integrable if and only if f^+ and f^- are both integrable, and then

$$\begin{aligned} \int_a^b f(x)dx &= \int_a^b f^+(x)dx - \int_a^b f^-(x)dx \\ &= \text{Area}\,(R_{f^+}) - \text{Area}\,(R_{f^-}), \end{aligned}$$

which can be called the 'signed area' delineated by the curve $y = f(x)$, $x \in [a, b]$.

Area between Curves

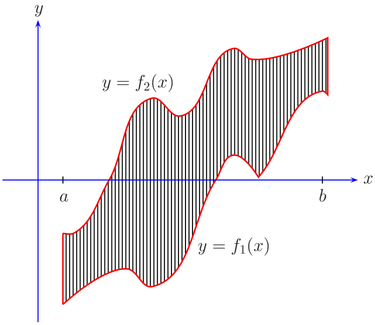
Let $f_1, f_2 : [a, b] \to \mathbb{R}$ be integrable functions such that $f_1 \le f_2$. Let $R := \{(x, y) \in \mathbb{R}^2 : a \le x \le b \text{ and } f_1(x) \le y \le f_2(x)\}$ be the **region between the curves** $y = f_1(x)$ and $y = f_2(x)$. Define

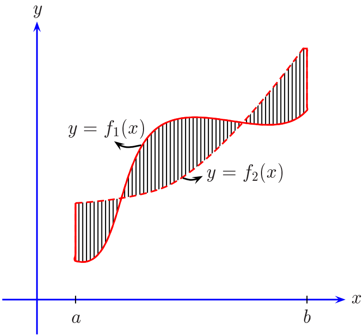
$$\text{Area}\,(R) := \text{Area}\,(R_{f_2 - f_1}) = \int_a^b \big(f_2(x) - f_1(x)\big)\,dx.$$

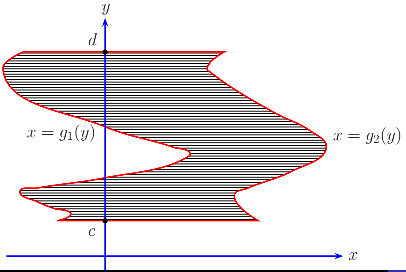
Let $g_1, g_2 : [c, d] \to \mathbb{R}$ be integrable functions such that $g_1 \le g_2$. Let $R := \{(x, y) \in \mathbb{R}^2 : c \le y \le d \text{ and } g_1(y) \le x \le g_2(y)\}$ be the **region between the curves** $x = g_1(y)$ and $x = g_2(y)$. Define

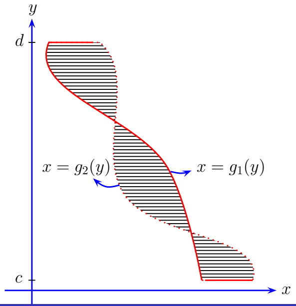
$$\text{Area}\,(R) := \int_c^d \big(g_2(y) - g_1(y)\big)\,dy.$$

If two curves cross each other a finite number of times, then we must find areas of several regions between them separately, and add them up.


y
x
a
b
y = f₂(x)
y = f₁(x)


y
x
a
b
y = f₁(x)
y = f₂(x)


y
x
d
c
x = g₁(y)
x = g₂(y)


y
x
d
c
x = g₂(y)
x = g₁(y)

Examples

(i) Let R denote the region enclosed by the loop of the curve $y^2 = x(1-x)^2$, that is, the region bounded by the curves $y = -\sqrt{x}(1-x)$ and $y = \sqrt{x}(1-x)$.
Now $\sqrt{x}(1-x) = -\sqrt{x}(1-x) \iff x = 0$ or 1, and $\sqrt{x}(1-x) \geq -\sqrt{x}(1-x)$ for $x \in [0,1]$. Hence

$$\text{Area}\,(R) = \int_0^1 \left(\sqrt{x}(1-x) - (-\sqrt{x}(1-x))\right) dx = \frac{8}{15}.$$

(ii) Let R denote the region bounded by the curves $x = -2y^2$ and $x = 1 - 3y^2$.
Now $-2y^2 = 1 - 3y^2 \iff y = \pm 1$, and $-2y^2 \leq 1 - 3y^2$ if $y \in [-1,1]$. Hence

$$\text{Area}\,(R) = \int_{-1}^1 \left(1 - 3y^2 - (-2y^2)\right) dy = \int_{-1}^1 (1 - y^2)\, dy = \frac{4}{3}.$$

Polar coordinates

Review:

The function $\cos^{-1} : [-1, 1] \to [0, \pi]$ is one-one and onto.

Let $P := (x, y) \neq (0, 0)$. There are unique $r,\ \theta \in \mathbb{R}$ such that

$$r > 0,\ \theta \in (-\pi, \pi],\ x = r\cos\theta \text{ and } y = r\sin\theta.$$

In fact, $r := \sqrt{x^2 + y^2}$ and

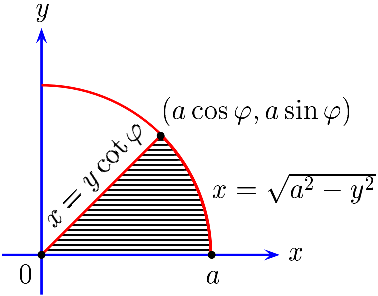
$$\theta := \begin{cases} \cos^{-1}(x/r) & \text{if } y \geq 0, \\ -\cos^{-1}(x/r) & \text{if } y < 0. \end{cases}$$

(If $y < 0$, then $|x/r| < 1$, and $-\cos^{-1}(x/r) \in (-\pi, 0)$.)

The pair (r, θ) is defined as the **polar coordinates** of P.

Area of a sector of a disk

Let $0 \le \varphi \le \pi/2$, and let R denote the sector of a disc of radius a, marked by the points $(0,0), (a,0)$ and $(a\cos\varphi, a\sin\varphi)$, that is, the region bounded by the curves $x = (\cot\varphi)y$ and $x = \sqrt{a^2 - y^2}$ for $y \in [0, a\sin\varphi]$, and by the x-axis.



$$\text{Then Area}\,(R) = \int_0^{a\sin\varphi} \left(\sqrt{a^2 - y^2} - (\cot\varphi)y\right) dy = \frac{a^2\varphi}{2}.$$

By symmetry, this result holds for $\varphi \in (\pi/2, \pi]$ as well.

MA 105 : Calculus
Division 1, Lecture 11

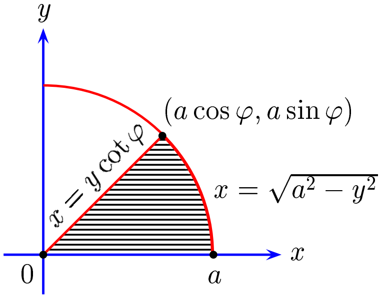
Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- The natural logarithm function and its properties
- The exponential function and its properties
- Real powers of a positive real number
- Definition of the arctangent function via integration, and its properties
- A self-contained definition of using the arctangent function
- An outline of the development of trigonometric functions beginning with arctan
- Applications of integration
    - Area of planar regions between two curves. Examples
    - Review of Polar coordinates
    - Area of the sector of a disk

Area of a sector of a disk

Let $0 \leq \varphi \leq \pi/2$, and let R denote the sector of a disc of radius a, marked by the points $(0,0), (a,0)$ and $(a\cos\varphi, a\sin\varphi)$, that is, the region bounded by the curves $x = (\cot\varphi)y$ and $x = \sqrt{a^2 - y^2}$ for $y \in [0, a\sin\varphi]$, and by the x-axis.



$$\text{Then Area}\,(R) \;=\; \int_0^{a\sin\varphi} \left(\sqrt{a^2 - y^2} - (\cot\varphi)y\right) dy = \frac{a^2\varphi}{2}.$$

By symmetry, this result holds for $\varphi \in (\pi/2, \pi]$ as well.

Curves given by Polar Equations

Let R denote the region bounded by the curve $r = p(\theta)$ and the rays $\theta = \alpha$, $\theta = \beta$, where $-\pi \le \alpha < \beta \le \pi$. Thus

$$R := \{(r\cos\theta, r\sin\theta) : \alpha \le \theta \le \beta \text{ and } 0 \le r \le p(\theta)\}.$$

Suppose $p : [\alpha, \beta] \to \mathbb{R}$ is integrable. (If $\alpha = -\pi$ and $\beta = \pi$, then we suppose $p(-\pi) = p(\pi)$.)

- Partition $[\alpha, \beta]$ into $\alpha = \theta_0 < \theta_1 < \cdots < \theta_n = \beta$.
- Pick sample points $\gamma_i \in [\theta_{i-1}, \theta_i]$ for $i = 1, \ldots, n$.
- Area between the rays $\theta = \theta_{i-1}$ and $\theta = \theta_i$ is approximated by the area of a sector of a disc of radius $r_i := p(\gamma_i)$, that is, by $\dfrac{p(\gamma_i)^2 (\theta_i - \theta_{i-1})}{2}$.
- The sum of areas of these sectors is a Riemann sum, namely $\displaystyle\sum_{i=1}^{n} \frac{p(\gamma_i)^2 (\theta_i - \theta_{i-1})}{2}$.

We define

$$\text{Area}\,(R) := \frac{1}{2}\int_{\alpha}^{\beta} p(\theta)^2 d\theta = \frac{1}{2}\int_{\alpha}^{\beta} r^2 d\theta.$$

Examples: (i) Let $a > 0$. Area of the disc enclosed by the circle $r = a$ is equal to $1/2 \int_{-\pi}^{\pi} a^2 d\theta = \pi a^2$.

(ii) Let $a > 0$, and let R denote the region enclosed by the cardioid $r = a(1 + \cos\theta)$. Then

$$\text{Area}\,(R) = \frac{1}{2}\int_{-\pi}^{\pi} a^2(1+\cos\theta)^2 d\theta = \frac{3a^2\pi}{2}.$$

(iii) Let R denote the region that lies inside the circle $r = 3\sin\theta$ and outside the cardioid $r = 1 + \sin\theta$, where $\theta \in [0, \pi]$. Now $3\sin\theta = 1 + \sin\theta \iff \theta \in \{\pi/6, 5\pi/6\}$, and $1 + \sin\theta \le 3\sin\theta$ if $\theta \in [\pi/6, 5\pi/6]$. Hence

$$\text{Area}\,(R) = \frac{1}{2}\int_{\pi/6}^{5\pi/6} \left((3\sin\theta)^2 - (1+\sin\theta)^2\right) d\theta = \pi.$$

Volume of a solid

Let D be a bounded subset of $\mathbb{R}^3$. A cross-section of D obtained by cutting D by a plane in $\mathbb{R}^3$ is called a **slice** of D.

Let $a < b$, and suppose D lies between the planes $x = a$ and $x = b$, which are perpendicular to the x-axis. For $s \in [a, b]$, consider the slice of D by the plane $x = s$, namely $\{(x, y, z) \in D : x = s\}$, and suppose it has an 'area' $A(s)$.

To find the volume of D, we proceed as follows.

- Partition $[a, b]$ into $a = x_0 < x_1 < \cdots < x_n = b$.
- Pick sample points $s_i \in [x_{i-1}, x_i]$ for $i = 1, \ldots, n$.
- Volume between the planes $x = x_{i-1}$ and $x = x_i$ is approximated by the volume of a rectangular slab of width $x_i - x_{i-1}$ and base area $A(s_i)$, that is, by $A(s_i)(x_i - x_{i-1})$.
- The sum of volumes of these slabs is $\displaystyle\sum_{i=1}^{n} A(s_i)(x_i - x_{i-1})$.

Slice Method: We define the **volume** of D by

$$\mathrm{Vol}\,(D) := \int_a^b A(x)\,dx,$$

provided the 'area function' $A : [a, b] \to \mathbb{R}$ is integrable.

Examples

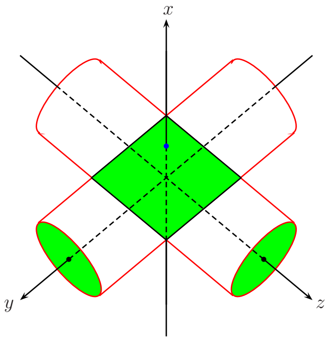
(i) If D is a cylinder with cross-sectional area A and height h, then $\mathrm{Vol}\,(D) = Ah$. (The 'area function' is the constant A.)

(ii) Let $a > 0$, and let D denote the solid enclosed by the cylinders $x^2 + y^2 = a^2$ and $x^2 + z^2 = a^2$. Then D lies between the planes $x = -a$ and $x = a$. For $s \in [-a, a]$, the slice $\{(x, y, z) \in D : x = s\}$ is the square

$$\left\{(s, y, z) \in \mathbb{R}^3 : |y| \le \sqrt{a^2 - s^2} \text{ and } |z| \le \sqrt{a^2 - s^2}\right\},$$

and so $A(s) = \left(2\sqrt{a^2 - s^2}\right)^2 = 4\,(a^2 - s^2)$. Hence

$$\mathrm{Vol}\,(D) = \int_{-a}^{a} A(s)ds = 4\int_{-a}^{a} \left(a^2 - s^2\right) ds = \frac{16a^3}{3}.$$



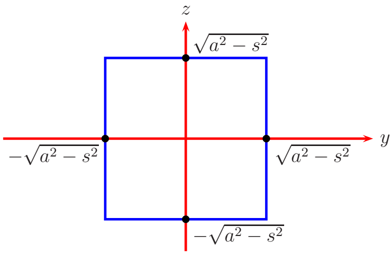


Figure: Solid enclosed by two cylinders and a slice resulting in a square region

Solids of Revolution

If a subset D of $\mathbb{R}^3$ is generated by revolving a planar region about an axis, then D is known as a **solid of revolution**.

Examples

- Let $a > 0$. The spherical ball

  $$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq a^2\}$$

  is obtained by revolving the semidisc

  $$\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq a^2 \text{ and } y \geq 0\}$$

  about the x-axis.
- Let $b > 0$. The cylindrical solid

  $$\{(x, y, z) \in \mathbb{R}^3 : x^2 + z^2 \leq b^2 \text{ and } 0 \leq y \leq h\}$$

  is obtained by revolving the rectangle $[0, b] \times [0, h]$ about the y-axis.

Volume of a Solid of Revolution: Washer Method

- Let D be the solid obtained by revolving the region between the curves $y = f_1(x)$ and $y = f_2(x)$, and the lines $x = a$ and $x = b$, about the x-axis, where $0 \le f_1 \le f_2$.
- The slice of D at $x \in [a, b]$ looks like a circular washer, that is, a disc of radius $f_2(x)$ from which a smaller disc of radius $f_1(x)$ has been removed, and so the area of the slice is $A(x) := \pi \left(f_2(x)^2 - f_1(x)^2\right)$.
- Suppose f_1 and f_2 are integrable on $[a, b]$. Then the area function A is integrable on $[a, b]$, and by the slice method,

$$\mathrm{Vol}\,(D) = \int_a^b A(x)dx = \pi \int_a^b \left(f_2(x)^2 - f_1(x)^2\right) dx.$$

This special case of the slice method is called the washer method.

If the inner radius of a washer is equal to 0, then the washer is in fact a disk. If this is the case for every $x \in [a, b]$, then the washer method is also known as the disk method.

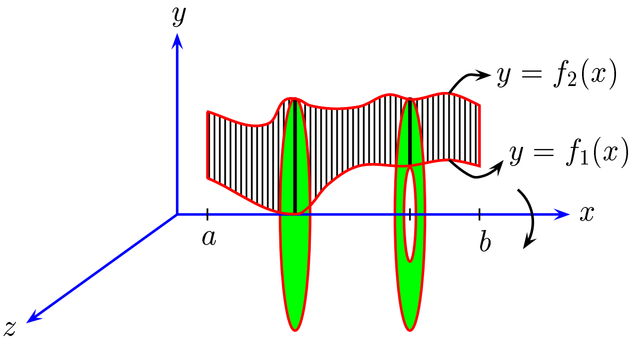


Figure: Illustrations of a disk and of a washer

Examples

(i) Let D denote the solid obtained by rotating the region between the curves $y = x$ and $y = x^2$ about the x-axis.

- Let $f_1(x) := x^2$ and $f_2(x) := x$ for $x \in [0, 1]$. The curves $y = f_1(x)$ and $y = f_2(x)$ intersect at $x = 0$ and $x = 1$, and $0 \le f_1 \le f_2$ on $[0, 1]$.
- By the washer method,

$$\mathrm{Vol}\,(D) = \pi \int_0^1 \left((x)^2 - (x^2)^2\right) dx = \pi \int_0^1 (x^2 - x^4) dx = \frac{2\pi}{15}.$$

(ii) Let D denote the spherical ball with centre at $(0,0,0)$, and radius $a > 0$. Then D is obtained by revolving the semidisc $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 \le a^2 \text{ and } y \ge 0\}$ about the x-axis.

- Let $f_1(x) := 0$ and $f_2(x) := \sqrt{a^2 - x^2}$ for $x \in [-a, a]$. The curves $y = f_1(x)$ and $y = f_2(x)$ intersect at $x = -a$ and $x = a$, and $0 \le f_1 \le f_2$ on $[-a, a]$.
- By the disc method,

$$\mathrm{Vol}\,(D) = \pi \int_{-a}^{a} \left(\left(\sqrt{a^2 - x^2} \right)^2 - 0^2 \right) dx = \pi \int_{-a}^{a} (a^2 - x^2)\, dx = \frac{4\pi a^3}{3}.$$

Let D be the solid obtained by revolving the region between the curves $x = g_1(y)$ and $x = g_2(y)$, $c \le y \le d$, about the y-axis, where $0 \le g_1 \le g_2$. Then, as before,

$$\mathrm{Vol}\,(D) = \pi \int_c^d \left(g_2(y)^2 - g_1(y)^2\right) dy.$$

Example

Let D denote the solid obtained by revolving the region in the first quadrant between the parabolas $y = x^2$ and $y = 2 - x^2$ about the y-axis. Now $\sqrt{y} = \sqrt{2-y} \iff y = 1$. By the disk method,

$$\mathrm{Vol}\,(D) = \pi \int_0^1 (\sqrt{y})^2 dy + \pi \int_1^2 \left(\sqrt{2-y}\right)^2 dy = \pi\left(\frac{1}{2} + \frac{1}{2}\right) = \pi.$$

Volume of a Solid of Revolution: Shell Method

Let D be a bounded subset of $\mathbb{R}^3$. A cross-section of D obtained by piercing a cylinder through D is called a **sliver** of D.

- Let $f_1, f_2 : [a, b] \to \mathbb{R}$ be such that $f_1 \le f_2$, and suppose $0 \le a < b$. Let D denote the solid generated by revolving the region between the curves $y = f_1(x)$ and $y = f_2(x)$, and the lines $x = a$ and $x = b$ about the y-axis.
- For $s \in [a, b]$, the sliver $\{(x, y, z) \in D : x^2 + z^2 = s^2\}$ of D by the cylinder $x^2 + z^2 = s^2$ is a right circular cylinder having height $f_2(s) - f_1(s)$ and radius s, and so its surface area is $2\pi s\big(f_2(s) - f_1(s)\big)$. (To be justified later)
- Suppose f_1 and f_2 are integrable on $[a, b]$. Since D is 'made up' of these cylindrical slivers (or shells), we define

$$\mathrm{Vol}\,(D) := 2\pi \int_a^b x\big(f_2(x) - f_1(x)\big)\,dx,$$

given by the shell method.

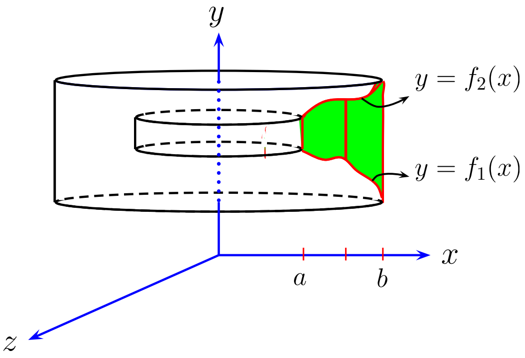


Figure: Illustration of the Shell Method

$$\text{Vol}\,(D) := 2\pi \int_a^b x\big(f_2(x) - f_1(x)\big)\,dx.$$

Shell Method (continued)

Let $g_1, g_2 : [c, d] \to \mathbb{R}$ be integrable functions such that $g_1 \le g_2$, and suppose $0 \le c < d$. Let D denote the solid obtained by revolving the region between the curves $x = g_1(y)$ and $x = g_2(y)$ and between the lines $y = c$ and $y = d$ about the x-axis.

For $t \in [c, d]$, the sliver $\{(x, y, z) \in D : y^2 + z^2 = t^2\}$ of D by the cylinder $y^2 + z^2 = t^2$ is a right circular cylinder having height $g_2(t) - g_1(t)$ and radius t, and so its surface area is $2\pi t\big(g_2(t) - g_1(t)\big)$.

Since D is 'made up' of these cylindrical slivers (or shells), we define

$$\mathrm{Vol}\,(D) := 2\pi \int_c^d y\big(g_2(y) - g_1(y)\big)\,dy.$$

Examples

(i) Let D denote the solid obtained by revolving the region bounded by the curves $y = 2x^2 - x^3$ and $y = 0$ about the y-axis. Now $2x^2 - x^3 = 0 \iff x = 0$ or $x = 2$. By the shell method,

$$\text{Vol}\,(D) = 2\pi \int_0^2 x(2x^2 - x^3 - 0)dx = \frac{56\pi}{5}.$$

(ii) Let D denote the solid obtained by revolving the rectangle $[0, b] \times [0, h]$ about the y-axis. By the shell method,

$$\text{Vol}\,(D) = 2\pi \int_0^b x\,(h - 0)\,dx = \pi hb^2.$$

(iii) Let E denote the solid obtained by revolving the rectangle $[0, b] \times [0, h]$ about the x-axis. By the shell method,

$$\text{Vol}\,(E) = 2\pi \int_0^h y\,(b - 0)\,dy = \pi bh^2.$$

(iv) Let D denote the solid in $\mathbb{R}^3$ obtained by revolving the region in the first quadrant bounded by the curve $y = x^3$ and the line $y = 4x$ about the line $x = -1$.

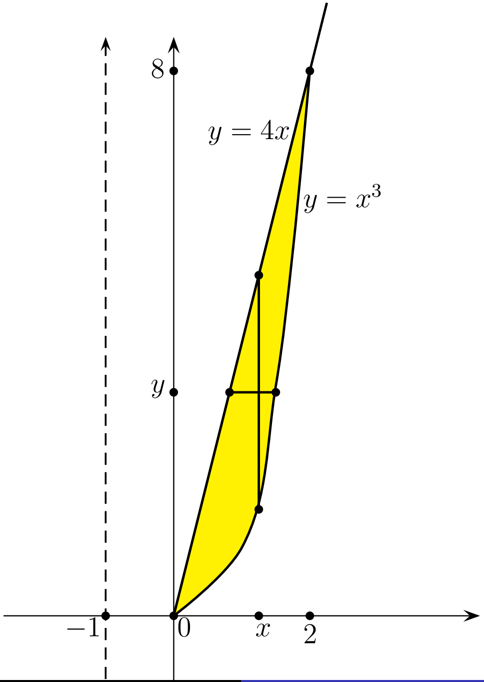
Now in the first quadrant, $x^3 = 4x \iff x = 0$ or $x = 2$.

- By the washer method,

$$\begin{aligned}\mathrm{Vol}\,(D) &= \pi \int_0^8 \left(\left(y^{1/3} + 1\right)^2 - \left((y/4) + 1\right)^2 \right) dy \\ &= \pi \int_0^8 \left(y^{2/3} + 2y^{1/3} - (y^2/16) - (y/2)\right) dy = \frac{248\pi}{15}.\end{aligned}$$

- By the shell method,

$$\begin{aligned}\mathrm{Vol}\,(D) &= 2\pi \int_0^2 (x+1)\left(4x - x^3\right) dx \\ &= 2\pi \int_0^2 (4x + 4x^2 - x^3 - x^4)\, dx = \frac{248\pi}{15}.\end{aligned}$$



8
$y = 4x$
$y = x^3$
y
-1
0
x
2

Remarks on Volume of a Solid of Revolution

Let D denote a bounded solid of revolution.

Remark 1. In the washer method, the slices (which look like washers) are taken perpendicular to the axis of revolution. On the other hand, in the shell method, the slivers (which look like cylindrical shells) are taken parallel to the axis of revolution.

Remark 2. The basic expression in the washer method is $\pi(r_2^2 - r_1^2)$, where r_2 and r_1 are outer and inner radii of the washer. The basic expression in the shell method is $2\pi r(h_2 - h_1)$, where r is the radius of the sliver, while $h_2 - h_1$ is the height of the sliver.

Remark 3. The volume of D found by the washer method and by the shell method must be the same! This result would follow from a general definition of the volume of a solid in $\mathbb{R}^3$. This can serve as a check on your calculations.

Parametrized Curve

A **parametrized curve** or a **path** C in $\mathbb{R}^2$ is given by $(x(t), y(t))$, where $x, y : [\alpha, \beta] \to \mathbb{R}$ are continuous functions.

Here $[\alpha, \beta]$ is called the **parameter interval**.

We wish to define the 'length' of C.

Basic assumption: The (Euclidean) length of a line segment joining points (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$ is equal to

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

We shall assume that C is **smooth**, that is, the functions x, y are **continuously differentiable** on $[\alpha, \beta]$. This means that x, y are differentiable on $[\alpha, \beta]$, and their derivatives x', y' are continuous on $[\alpha, \beta]$.

MA 105 : Calculus
Division 1, Lecture 12

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Curves given by polar equations. Area of related planar regions
- Volume of a solid. Slice Method, Examples
- Solids of Revolution
    - Washer Method
    - Shell Method
    - Examples
- Parametrized curves

Parametrized Curve

A **parametrized curve** or a **path** C in $\mathbb{R}^2$ is given by $(x(t), y(t))$, where $x, y : [\alpha, \beta] \to \mathbb{R}$ are continuous functions.

Here $[\alpha, \beta]$ is called the **parameter interval**.

We wish to define the 'length' of C.

Basic assumption: The (Euclidean) length of a line segment joining points (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$ is equal to

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

We shall assume that C is **smooth**, that is, the functions x, y are **continuously differentiable** on $[\alpha, \beta]$. This means that x, y are differentiable on $[\alpha, \beta]$, and their derivatives x', y' are continuous on $[\alpha, \beta]$.

Arc Length of a Smooth Curve

- Partition $[\alpha, \beta]$ into $\alpha = t_0 < t_1 < \cdots < t_n = \beta$.
- Let $P_i := \big(x(t_i), y(t_i)\big)$ for $i = 1, \ldots, n$, and draw the line segments joining P_0 to P_1, P_1 to P_2, ..., P_{n-1} to P_n.
- The sum of the lengths of these line segments is

$$\sum_{i=1}^{n} \sqrt{\left(x(t_i) - x(t_{i-1})\right)^2 + \left(y(t_i) - y(t_{i-1})\right)^2}$$

$$= \sum_{i=1}^{n} \sqrt{\left(x'(s_i)\right)^2 + \left(y'(u_i)\right)^2}\,(t_i - t_{i-1}),$$

for some s_i, $u_i \in (t_{i-1}, t_i)$ for $i = 1, \ldots, n$ by the MVT.

- We define the **arc length** of C by

$$\ell(C) := \int_{\alpha}^{\beta} \sqrt{x'(t)^2 + y'(t)^2}\, dt.$$

Special Cases

Special cases:

(i) Let a curve C be given by $y = f(x)$, $x \in [a, b]$.
Here $\alpha := a$, $\beta := b$, $x(t) := t$ and $y(t) := f(t)$ for $t \in [a, b]$.
Suppose f is continuously differentiable on $[a, b]$. Then

$$\ell(C) := \int_a^b \sqrt{1 + f'(x)^2}\, dx.$$

(ii) Let a curve C be given by $x = g(y)$, $y \in [c, d]$.
Here $\alpha := c$, $\beta := d$, $x(t) := g(t)$ and $y(t) := t$ for $t \in [c, d]$.
Suppose g is continuously differentiable on $[c, d]$. Then

$$\ell(C) := \int_c^d \sqrt{g'(y)^2 + 1}\, dy.$$

Arc Length in Polar coordinates

Let C be given by a polar equation $r = p(\theta)$, $\theta \in [\alpha, \beta]$. As a parametrized curve, C is given by $(x(\theta), y(\theta))$, where

$$x(\theta) := p(\theta)\cos\theta \quad \text{and} \quad y(\theta) := p(\theta)\sin\theta, \quad \theta \in [\alpha, \beta].$$

Suppose the function p is continuously differentiable on $[\alpha, \beta]$. For $\theta \in [\alpha, \beta]$, we note that $\sqrt{x'(\theta)^2 + y'(\theta)^2}$ is equal to

$$\begin{aligned} & \sqrt{\left(p'(\theta)\cos\theta - p(\theta)\sin\theta\right)^2 + \left(p'(\theta)\sin\theta + p(\theta)\cos\theta\right)^2} \\ = \ & \sqrt{p(\theta)^2 + p'(\theta)^2}. \end{aligned}$$

Hence

$$\ell(C) := \int_\alpha^\beta \sqrt{p(\theta)^2 + p'(\theta)^2}\, d\theta.$$

Examples

(i) Let C be given by $y = x^2$, $x \in [0, 1]$. Then

$$\begin{aligned} \ell(C) &= \int_0^1 \sqrt{1 + (2x)^2}\, dx = \frac{1}{2} \int_0^2 \sqrt{1 + u^2}\, du \\ &= \frac{1}{2}\sqrt{5} + \frac{1}{4} \ln\left(2 + \sqrt{5}\right). \end{aligned}$$

(Use Integration by Parts. Also, if $f(u) := \ln(u + \sqrt{1 + u^2})$ for $u \in \mathbb{R}$, then note that $f'(u) = 1/\sqrt{1 + u^2}$ for $u \in \mathbb{R}$, and so $\int_0^x \sqrt{1 + u^2}\, du = \frac{1}{2}\left(x\sqrt{1 + x^2} + \ln\left(x + \sqrt{1 + x^2}\right)\right)$ for $x \in \mathbb{R}$.)

(ii) Let C be given by $x = (2y^6 + 1)/8y^2$, $y \in [1, 2]$. Then

$$\int_1^2 \left(1 + \left(y^3 - \frac{1}{4y^3}\right)^2\right)^{1/2} dy = \int_1^2 \left(y^3 + \frac{1}{4y^3}\right) dy = \frac{123}{32}.$$

(iii) Let $a > 0$ and $\varphi \in [0, \pi]$. Let C denote the arc of a circle of radius a given by $x(\theta) := a \cos\theta$, $y(\theta) := a \sin\theta$ for $\theta \in [0, \varphi]$. Then C is given by the polar equation $r = p(\theta)$, where $p(\theta) = a$ for $\theta \in [0, \phi]$, and so

$$\ell(C) = \int_0^{\varphi} \sqrt{a^2 + 0^2}\, d\theta = a\,\varphi.$$

Hence the length of a circle of radius a is $\int_{-\pi}^{\pi} a\, d\theta = 2\pi a$.

(iv) Let C be given by $r = 1 + \cos\theta$ for $\theta \in [0, \pi]$. Then

$$\begin{aligned}\ell(C) &= \int_0^{\pi} \sqrt{(1+\cos\theta)^2 + (-\sin\theta)^2}\, d\theta \\ &= \int_0^{\pi} \sqrt{2(1+\cos\theta)}\, d\theta = 2\int_0^{\pi} \cos\frac{\theta}{2}\, d\theta = 4.\end{aligned}$$

(Note: $\cos(\theta/2) \geq 0$ for $\theta \in [0, \pi]$.)

Curves in $\mathbb{R}^3$

Suppose C is a smooth parametrized curve in $\mathbb{R}^3$ given by $(x(t), y(t), z(t))$, where x, y, $z : [\alpha, \beta] \to \mathbb{R}$ are continuously differentiable functions on $[\alpha, \beta]$.

In analogy with the definition of the arc length of a curve in $\mathbb{R}^2$, we define the **arc length** of C by

$$\ell\,(C) := \int_\alpha^\beta \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2}\,dt.$$

Example

Let C denote a **helix** in $\mathbb{R}^3$ given by $x(t) := a\cos t,\ \ y(t) := a\sin t,\ \ z(t) := b\,t,\ \ t \in [\alpha, \beta]$, where $a,\ b, \in \mathbb{R}$, $a > 0$ and $b \neq 0$. Then

$$\ell\,(C) = \int_\alpha^\beta \sqrt{(-a\sin t)^2 + (a\cos t)^2 + b^2}\,dt = (\beta - \alpha)\sqrt{a^2 + b^2}.$$

Surface of Revolution

A **surface of revolution** is generated when a curve C in $\mathbb{R}^2$ is revolved about a line L in $\mathbb{R}^2$.

First suppose the curve C is a slanted line segment P_1P_2 of length λ_2, and C does not cross L. Let d_1 and d_2 denote the distances of P_1 and P_2 from L with $d_1 \leq d_2$. Then the surface of revolution is a frustum F of a right circular cone with base radii d_1 and d_2, and slant height λ_2. We find its surface area.

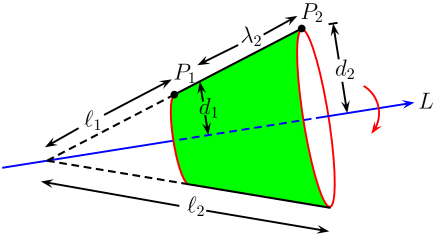


Figure: Frustum of a right circular cone

Consider a cone with base radius d and slant height ℓ. If we slit open this cone, we obtain a sector of a disk of radius ℓ.

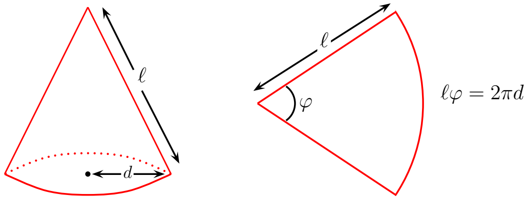


Figure: Right circular cone and sector of a disk

Since $\ell\varphi = 2\pi d$, the surface area of the cone is equal to

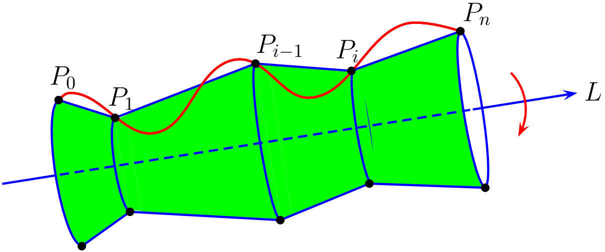
$$\frac{1}{2}\ell^2\varphi = \frac{1}{2}\ell^2\frac{2\pi d}{\ell} = \pi\ell d.$$

Hence the surface area of the frustrum F of the cone is $\pi\ell_2 d_2 - \pi\ell_1 d_1 = \pi(d_1 + d_2)(\ell_2 - \ell_1) = \pi(d_1 + d_2)\lambda_2$.

Now suppose C is parametrized by $(x(t), y(t))$, $t \in [\alpha, \beta]$.

- Partition $[\alpha, \beta]$ into $\alpha = t_0 < t_1 < \cdots < t_n = \beta$.
- Let $P_i := \big(x(t_i), y(t_i)\big)$ for $i = 0, 1, \ldots, n$, and draw the line segments P_0P_1, P_1P_2, …, $P_{n-1}P_n$.

Let $d_0, d_1, d_2, \ldots, d_n$ be the distances of $P_0, P_1, P_2, \ldots, P_n$ from the line L. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the lengths of the line segments $P_0P_1, P_1P_2, \ldots, P_{n-1}P_n$. Suppose they don't cross L.



Fix $i \in \{1, \ldots, n\}$. When the line segment $P_{i-1}P_i$ is revolved about the line L, it generates a frustum F_i (of a right circular cone) whose surface area is $\pi(d_{i-1} + d_i)\lambda_i$.

Let $\rho(t)$ denote the distance of the point $(x(t), y(t))$ on the curve C from the line L. Then $d_i = \rho(t_i)$ for $i = 0, 1, \ldots, n$.

Thus the sum of the surface areas of the frustrums $F_1, \ldots, F_n$ is

$$\pi \sum_{i=1}^{n} \big(\rho(t_{i-1}) + \rho(t_i)\big)\lambda_i,$$

If the functions x' and y' are continuously differentiable on $[\alpha, \beta]$, then the length λ_i of the line segment $P_{i-1}P_i$ is given by

$$\begin{aligned}\lambda_i &= \sqrt{\big(x(t_i) - x(t_{i-1})\big)^2 + \big(y(t_i) - y(t_{i-1})\big)^2} \\ &= \sqrt{x'(s_i)^2 + y'(u_i)^2}\,(t_i - t_{i-1})\end{aligned}$$

for some $s_i,\ u_i \in (t_{i-1}, t_i)$ for $i = 1, \ldots, n$ (by the MVT).

Area of Surface of Revolution

Let C be a smooth curve parametrized by $(x(t), y(t))$, $t \in [\alpha, \beta]$. Suppose the curve C does not cross the line L given by $ax + by + c = 0$. We define the **area of the surface** S generated by revolving C about the line L by

$$\text{Area}\,(S) := 2\pi \int_{\alpha}^{\beta} \rho(t)\sqrt{x'(t)^2 + y'(t)^2}\, dt,$$

where $\rho(t)$ is the distance of $(x(t), y(t))$ from the line L,

that is, $\rho(t) := |ax(t) + by(t) + c|/\sqrt{a^2 + b^2}$ for $t \in [a, b]$.

Note: Since the curve C does not cross the line L, the curve C lies entirely on one of the sides of the line L, that is, either $ax(t) + by(t) + c \geq 0$ for all $t \in [\alpha, \beta]$, or $ax(t) + by(t) + c \leq 0$ for all $t \in [\alpha, \beta]$.

Special Cases:

(i) Let the line L be the x-axis, and let the curve C be given by $y = f(x)$ for $x \in [a, b]$, where f is continuously differentiable. If $f \geq 0$ on $[a, b]$ or $f \leq 0$ on $[a, b]$, then

$$\text{Area}(S) = 2\pi \int_a^b |f(x)|\sqrt{1 + f'(x)^2}dx.$$

(ii) Let the line L be the y-axis, and let the curve C be given by $x = g(y)$ for $y \in [c, d]$, where g is continuously differentiable. If $g \geq 0$ on $[c, d]$ or $g \leq 0$ on $[c, d]$, then

$$\text{Area}(S) = 2\pi \int_c^d |g(y)|\sqrt{1 + g'(y)^2}dy.$$

(iii) Let the line L be given by $\theta = \gamma$, where $\gamma \in (-\pi, \pi]$, and let the curve C be given by $r = p(\theta)$ for $\theta \in [\alpha, \beta]$, where p is continuously differentiable on $[\alpha, \beta]$. Suppose C does not cross L. Now the curve C is also given by $(p(\theta)\cos\theta, p(\theta)\sin\theta)$ for $\theta \in [\alpha, \beta]$.

Also, $\rho(\theta) = p(\theta)|\sin(\theta - \gamma)|$ for $\theta \in [\alpha, \beta]$.

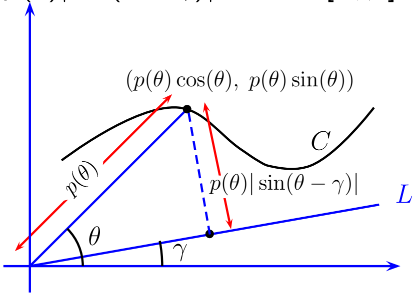


Figure: Distance of a point on a polar curve from a ray.

Thus $\text{Area}(S) = 2\pi \displaystyle\int_\alpha^\beta p(\theta)|\sin(\theta-\gamma)|\sqrt{p(\theta)^2 + p'(\theta)^2}\,d\theta.$

Examples:

(i) Let S denote the surface generated by revolving the curve $y = (x^3/3) + (1/4x)$, $x \in [1, 3]$, about the line $y = -1$. Then

$$\begin{aligned} \text{Area}(S) &= 2\pi \int_1^3 (y+1)\sqrt{1+(y')^2}\, dx \\ &= 2\pi \int_1^3 \left(\frac{x^3}{3} + \frac{1}{4x} + 1\right) \sqrt{1 + \left(x^2 - \frac{1}{4x^2}\right)^2}\, dx \\ &= 2\pi \int_1^3 \left(\frac{x^3}{3} + \frac{1}{4x} + 1\right) \left(x^2 + \frac{1}{4x^2}\right) dx \\ &= 1823\pi/18. \end{aligned}$$

(iii) Let $0 < b < a$ and let C denote the circle given by $(a + b\cos t, b\sin t)$, $t \in [-\pi, \pi]$. Let S denote the surface generated by revolving the curve C about the y-axis. Then $a + b\cos t > 0$ for all $t \in [-\pi, \pi]$, and so

$$\begin{aligned}\text{Area}\,(S) &= 2\pi \int_{-\pi}^{\pi} (a + b\cos t)\sqrt{(-b\sin t)^2 + (b\cos t)^2}\, dt \\ &= 2\pi b \int_{-\pi}^{\pi} (a + b\cos t) dt \\ &= 4\pi^2 ab.\end{aligned}$$

Note: S is in fact the surface of a torus in $\mathbb{R}^3$.

(iii) Let $a > 0$, and S denote the surface generated by revolving the semicircle $p(\theta) = a$, $\theta \in [0, \pi]$, about the x-axis. Then

$$\text{Area}(S) = 2\pi \int_0^{\pi} a\sin\theta \sqrt{a^2 + 0^2}\, d\theta = 4\pi a^2.$$

Note: S is in fact the sphere of radius a in $\mathbb{R}^3$.

Euclidean Spaces

Let $m \in \mathbb{N}$, and consider the m dimensional **Euclidean space**

$$\mathbb{R}^m := \{(x_1, \ldots, x_m) : x_j \in \mathbb{R} \text{ for } j = 1, \ldots, m\}.$$

If $m = 1$, then an element of $\mathbb{R}^1 := \mathbb{R}$ is called a **scalar** and if $m > 1$, then an element of $\mathbb{R}^m$ is called a **vector**.
We shall consider sequences in $\mathbb{R}^m$, and functions from subsets of $\mathbb{R}^m$ to $\mathbb{R}$, and we shall study their properties.
We shall set up the basic structure in $\mathbb{R}^m$, and then restrict ourselves to $m := 2, 3$.

For $\boldsymbol{x} := (x_1, \ldots, x_m)$, $\boldsymbol{y} := (y_1, \ldots, y_m) \in \mathbb{R}^m$, and $a \in \mathbb{R}$, let

$$\begin{aligned} \boldsymbol{x} + \boldsymbol{y} &:= (x_1 + y_1, \ldots, x_m + y_m) \in \mathbb{R}^m, \\ a\,\boldsymbol{x} &:= (a\,x_1, \ldots, a\,x_m) \in \mathbb{R}^m, \end{aligned}$$

called the **sum** of $\boldsymbol{x}$ and $\boldsymbol{y}$, and the **scalar multiple** of a and $\boldsymbol{x}$.

Norm of an element of $\mathbb{R}^m$

Let $\boldsymbol{x} := (x_1, \ldots, x_m) \in \mathbb{R}^m$. We define the **norm** of $\boldsymbol{x}$ by

$$\|\boldsymbol{x}\| := (x_1^2 + \cdots + x_m^2)^{1/2}.$$

For $m = 1$, the norm of $x \in \mathbb{R}$ is its absolute value $|x|$. Note:

$$\max\{|x_1|, \ldots, |x_m|\} \leq \|\boldsymbol{x}\| \leq |x_1| + \cdots + |x_m|.$$

Let $\mathbf{0} := (0, \ldots, 0)$. Then $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \mathbf{0}$.

For $\boldsymbol{x} := (x_1, \ldots, x_m) \in \mathbb{R}^m$ and $\boldsymbol{y} := (y_1, \ldots, y_m) \in \mathbb{R}^m$, let

$$\boldsymbol{x} \cdot \boldsymbol{y} \quad := \quad x_1 y_1 + \cdots + x_m y_m \in \mathbb{R},$$

called the **dot product** or the **scalar product** of $\boldsymbol{x}$ and $\boldsymbol{y}$.

Theorem (Cauchy-Schwarz Inequality and Triangle Inequality)

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, $|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$ and $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

Proof: If $\|\boldsymbol{x}\| = 0$ or $\|\boldsymbol{y}\| = 0$, then $\boldsymbol{x} = \mathbf{0}$ or $\boldsymbol{y} = \mathbf{0}$, so done.

Suppose $\|\boldsymbol{x}\| \neq 0$ and $\|\boldsymbol{y}\| \neq 0$. Then

$$\frac{|x_j|}{\|\boldsymbol{x}\|}\frac{|y_j|}{\|\boldsymbol{y}\|} \leq \frac{1}{2}\left(\frac{|x_j|^2}{\|\boldsymbol{x}\|^2} + \frac{|y_j|^2}{\|\boldsymbol{y}\|^2}\right) \quad \text{for } j = 1, \ldots, m,$$

since $ab \leq (a^2 + b^2)/2$ for all $a, b \in \mathbb{R}$. Hence

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \sum_{j=1}^{m} |x_j|\,|y_j| \leq \frac{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}{2}(1+1) = \|\boldsymbol{x}\|\|\boldsymbol{y}\|.$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$. As a consequence,

$$\begin{aligned} \|\boldsymbol{x} + \boldsymbol{y}\|^2 &= \sum_{j=1}^{m}(x_j + y_j)^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\,\boldsymbol{x} \cdot \boldsymbol{y} \\ &\leq \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\,\|\boldsymbol{x}\|\|\boldsymbol{y}\| = (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2. \end{aligned}$$

Thus $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$. $\square$

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$. We say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal** or **perpendicular** (to each other) if

$$\boldsymbol{x} \cdot \boldsymbol{y} = 0.$$

If $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{y} \neq \mathbf{0}$, then the **angle** between them is the unique $\theta \in [0, \pi]$ such that $\cos\theta = \boldsymbol{x} \cdot \boldsymbol{y} / \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|$. (Cosine Rule)

Let $m \in \mathbb{N}$ and $\boldsymbol{x}_0 \in \mathbb{R}^m$. For $\boldsymbol{x} \in \mathbb{R}^m$, $\|\boldsymbol{x} - \boldsymbol{x}_0\|$ is called the **distance** of $\boldsymbol{x}$ from $\boldsymbol{x}_0$. For $r > 0$, the subset

$$B(\boldsymbol{x}_0, r) := \{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x} - \boldsymbol{x}_0\| < r\}$$

of $\mathbb{R}^m$ is called the **neighbourhood** of $\boldsymbol{x}_0$ of radius r.

If $m := 1$, then $B(\boldsymbol{x}_0, r)$ is the interval $(x_0 - r, x_0 + r)$ in $\mathbb{R}$.

If $m := 2$ and $\boldsymbol{x}_0 := (x_0, y_0)$, then $B(\boldsymbol{x}_0, r)$ is the disk $\{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < r^2\}$ in $\mathbb{R}^2$.

If $m := 3$ and $\boldsymbol{x}_0 := (x_0, y_0, z_0)$, then $B(\boldsymbol{x}_0, r)$ is the ball $\{(x, y, z) \in \mathbb{R}^3 : (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 < r^2\}$ in $\mathbb{R}^3$.

MA 105 : Calculus
Division 1, Lecture 13

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Arc length of a smooth curve in the plane
- Special cases, including polar curves. Examples
- Curves in 3-space and their arc-length. Helix as an example
- Surfaces of revolution
- Surface area of a cone and of a frustum of a cone
- Area of surface of revolution obtained by revolving a smooth parametric curve about a line
- Special cases and examples.

- Euclidean spaces
- Notion of addition and scalar multiplication of vectors
- Dot product. (Cauchy-) Schwarz inequality and Triangle inequality
- Notions of orthogonal vectors, angle between nonzero vectors, neighbourhood of a vector

Remaining Tut Problem

In the last tutorial, all problems of Tut Sheet 5, except # 4, 5, 7, were covered. Of those that remain, (7) was worked out in Lecture 11 and (5) in the last lecture. We will now work out:

(4) Find the arc length of the each of the curves described below.

(i) the cycloid $x = t - \sin t,\ y = 1 - \cos t,\ 0 \le t \le 2\pi$

(ii) $y = \int_0^x \sqrt{\cos 2t}\, dt,\ 0 \le x \le \pi/4.$

Solution: The arc length in (i) is given by

$$\int_0^{2\pi} \sqrt{(1-\cos(t))^2 + \sin^2(t)}dt = \int_0^{2\pi} 2|\sin(t/2)|dt = 8.$$

and using the FTC, the arc length in (ii) is given by

$$\int_0^{\pi/4} \sqrt{1+y'^2}dx = \int_0^{\pi/4} \sqrt{1+\cos(2x)}dx = \sqrt{2}\int_0^{\pi/4} |\cos(x)|dx = 1.$$

Thus Tutorial Sheet 5 is now completely done!

Euclidean Spaces

Let $m \in \mathbb{N}$, and consider the m dimensional **Euclidean space**

$$\mathbb{R}^m := \{(x_1, \ldots, x_m) : x_j \in \mathbb{R} \text{ for } j = 1, \ldots, m\}.$$

If $m = 1$, then an element of $\mathbb{R}^1 := \mathbb{R}$ is called a **scalar** and if $m > 1$, then an element of $\mathbb{R}^m$ is called a **vector**.
We shall consider sequences in $\mathbb{R}^m$, and functions from subsets of $\mathbb{R}^m$ to $\mathbb{R}$, and we shall study their properties.
We shall set up the basic structure in $\mathbb{R}^m$, and then restrict ourselves to $m := 2, 3$.

For $\boldsymbol{x} := (x_1, \ldots, x_m)$, $\boldsymbol{y} := (y_1, \ldots, y_m) \in \mathbb{R}^m$, and $a \in \mathbb{R}$, let

$$\begin{aligned} \boldsymbol{x} + \boldsymbol{y} &:= (x_1 + y_1, \ldots, x_m + y_m) \in \mathbb{R}^m, \\ a\,\boldsymbol{x} &:= (a\,x_1, \ldots, a\,x_m) \in \mathbb{R}^m, \end{aligned}$$

called the **sum** of $\boldsymbol{x}$ and $\boldsymbol{y}$, and the **scalar multiple** of a and $\boldsymbol{x}$.

Norm of an element of $\mathbb{R}^m$

Let $\boldsymbol{x} := (x_1, \ldots, x_m) \in \mathbb{R}^m$. We define the **norm** of $\boldsymbol{x}$ by

$$\|\boldsymbol{x}\| := (x_1^2 + \cdots + x_m^2)^{1/2}.$$

For $m = 1$, the norm of $x \in \mathbb{R}$ is its absolute value $|x|$. Note:

$$\max\{|x_1|, \ldots, |x_m|\} \le \|\boldsymbol{x}\| \le |x_1| + \cdots + |x_m|.$$

Let $\boldsymbol{0} := (0, \ldots, 0)$. Then $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \boldsymbol{0}$.

For $\boldsymbol{x} := (x_1, \ldots, x_m) \in \mathbb{R}^m$ and $\boldsymbol{y} := (y_1, \ldots, y_m) \in \mathbb{R}^m$, let

$$\boldsymbol{x} \cdot \boldsymbol{y} \;:=\; x_1 y_1 + \cdots + x_m y_m \in \mathbb{R},$$

called the **dot product** or the **scalar product** of $\boldsymbol{x}$ and $\boldsymbol{y}$.

Theorem (Cauchy-Schwarz Inequality and Triangle Inequality)

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, $|\boldsymbol{x} \cdot \boldsymbol{y}| \le \|\boldsymbol{x}\| \|\boldsymbol{y}\|$ and $\|\boldsymbol{x} + \boldsymbol{y}\| \le \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$. We say that $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal** or **perpendicular** (to each other) if

$$\boldsymbol{x} \cdot \boldsymbol{y} = 0.$$

If $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{y} \neq \mathbf{0}$, then the **angle** between them is the unique $\theta \in [0, \pi]$ such that $\cos\theta = \boldsymbol{x} \cdot \boldsymbol{y} / \|\boldsymbol{x}\| \, \|\boldsymbol{y}\|$. (Cosine Rule)

Let $m \in \mathbb{N}$ and $\boldsymbol{x}_0 \in \mathbb{R}^m$. For $\boldsymbol{x} \in \mathbb{R}^m$, $\|\boldsymbol{x} - \boldsymbol{x}_0\|$ is called the **distance** of $\boldsymbol{x}$ from $\boldsymbol{x}_0$. For $r > 0$, the subset

$$B(\boldsymbol{x}_0, r) := \{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x} - \boldsymbol{x}_0\| < r\}$$

of $\mathbb{R}^m$ is called the **neighbourhood** of $\boldsymbol{x}_0$ of radius r.

If $m := 1$, then $B(\boldsymbol{x}_0, r)$ is the interval $(x_0 - r, x_0 + r)$ in $\mathbb{R}$.

If $m := 2$ and $\boldsymbol{x}_0 := (x_0, y_0)$, then $B(\boldsymbol{x}_0, r)$ is the disk $\{(x, y) \in \mathbb{R}^2 : (x - x_0)^2 + (y - y_0)^2 < r^2\}$ in $\mathbb{R}^2$.

If $m := 3$ and $\boldsymbol{x}_0 := (x_0, y_0, z_0)$, then $B(\boldsymbol{x}_0, r)$ is the ball $\{(x, y, z) \in \mathbb{R}^3 : (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 < r^2\}$ in $\mathbb{R}^3$.

Sequence in $\mathbb{R}^2$ and its Limit

A **sequence** in $\mathbb{R}^2$ is a function from $\mathbb{N}$ to $\mathbb{R}^2$. We denote the nth term of such a sequence by (x_n, y_n), and the sequence itself by $\big((x_n, y_n)\big)$.

Definition

A sequence $\big((x_n, y_n)\big)$ in $\mathbb{R}^2$ is **convergent** in $\mathbb{R}^2$ if both (x_n) and (y_n) are convergent sequences in $\mathbb{R}$. In this case, we write

$$(x_n, y_n) \to (x_0, y_0), \text{ where } x_n \to x_0 \text{ and } y_n \to y_0.$$

For $n \in \mathbb{N}$,

$$\max\{|x_n - x_0|, |y_n - y_0|\} \leq \|(x_n, y_n) - (x_0, y_0)\| \leq |x_n - x_0| + |y_n - y_0|.$$

Hence $(x_n, y_n) \to (x_0, y_0) \iff \|(x_n, y_n) - (x_0, y_0)\| \to 0.$

Functions on subsets of $\mathbb{R}^2$

Let D be a subset of $\mathbb{R}^2$, and let f be a real-valued function defined on D. The subset

$$\{(x, y, f(x, y)) \in \mathbb{R}^3 : (x, y) \in D\}$$

is called the **graph** of f. It is the surface $z = f(x, y)$ in $\mathbb{R}^3$.

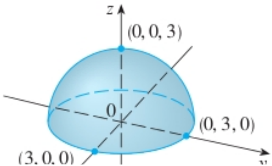
Let $c \in \mathbb{R}$. Then the set
$\{\big(x, y, c\big) \in \mathbb{R}^3 : (x, y) \in D$ and $f(x, y) = c\}$ is called a **contour line** of f. It is the intersection of the graph of f by the horizontal plane $z = c$ in $\mathbb{R}^3$.

Also, $\{(x, y) \in D : f(x, y) = c\} \subset \mathbb{R}^2$ is called a **level curve** of f. It is the projection of the corresponding contour line on the plane $z = 0$ in $\mathbb{R}^3$.

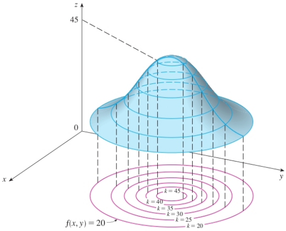
Contour lines and level curves help us visualize the behaviour of a function f on its domain in $\mathbb{R}^2$. Such a visualization is not possible for functions defined on subsets of $\mathbb{R}^3$.

Examples

(i) Let $D := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 9\}$, and let $f(x, y) := \sqrt{9 - x^2 - y^2}$ for $(x, y) \in D$. The graph of f is the upper hemisphere of radius 3 centered at the origin.

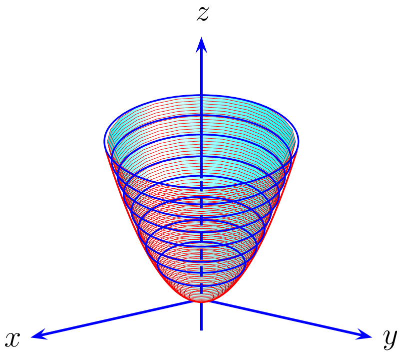


(ii) Contour lines



and level curves.

(iii) Let $D := \mathbb{R}^2$, and let $f(x, y) := x^2 + y^2$ for $(x, y) \in D$. The graph of f is a paraboloid.



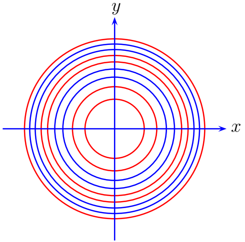


Figure: Contour lines and the corresponging level curves

(iv) Let $D := \{(x, y) \in \mathbb{R}^2 : x \neq y\}$, and let $f(x, y) := (x + y)/(x - y)$ for $(x, y) \in D$. Then the contour line corresponding to $c = 2$ is $\{(x, x/3, 2) : x \in \mathbb{R} \setminus \{0\}\}$.

Continuity of a function defined on a subset of $\mathbb{R}^2$

Definition

Let $D \subset \mathbb{R}^2$, $f : D \to \mathbb{R}$ and $(x_0, y_0) \in D$. We say that f is **continuous** at (x_0, y_0) if

$$(x_n, y_n) \to (x_0, y_0) \implies f(x_n, y_n) \to f(x_0, y_0)$$

for every sequence $\big((x_n, y_n)\big)$ in D.

We say that f is **continuous on** D if f is continuous at each point of D.

We say that f is **discontinuous** on D if it is not continuous at some point of D.

Basic properties of continuous functions

- Let $D \subset \mathbb{R}^2$, and $f, g : D \to \mathbb{R}$. Let $(x_0, y_0) \in D$, and let f and g be continuous at (x_0, y_0). Then $f + g$ and $f \cdot g$ are continuous at (x_0, y_0). Further, f/g is continuous at (x_0, y_0) if $g((x_0, y_0)) \neq 0$. (Limit theorems for sequences.)
- Let $D \subset \mathbb{R}^2, E \subset \mathbb{R}$. Suppose $f : D \to \mathbb{R}$ and $g : E \to \mathbb{R}$ are functions such that $f(D) \subset E$, f is continuous at $(x_0, y_0) \in D$, and g is continuous at $f(x_0, y_0) \in E$. Then the **composite function** $g \circ f : D \to \mathbb{R}$ is continuous at (x_0, y_0). (The proof is easy.)

Examples

(i) Let p be a **polynomial function in two variables**, that is, let $k = 0$ or $k \in \mathbb{N}$ and for $(x, y) \in \mathbb{R}^2$, let

$$p(x, y) := \sum_{i \geq 0,\, j \geq 0,\, i+j \leq k} a_{i,j} x^i y^j, \quad \text{where } a_{i,j} \in \mathbb{R}.$$

Then p is a continuous function on $\mathbb{R}^2$.

(ii) A **rational function in two variables**, that is, a function given by $r(x, y) := p(x, y)/q(x, y)$, where p and q are polynomial functions in two variables, is continuous at each $(x_0, y_0) \in \mathbb{R}^2$ at which $q((x_0, y_0)) \neq 0$.

(iii) Let $p(x, y)$ be a polynomial function in two variables, and define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) := e^{p(x,y)}$ for $(x, y) \in \mathbb{R}^2$. Then f is continuous on $\mathbb{R}^2$ since the function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(z) := e^z$ for $z \in \mathbb{R}$, is continuous on $\mathbb{R}$.

Equivalent condition for continuity of a function

Let $D \subset \mathbb{R}^2$, $f : D \to \mathbb{R}$ and $(x_0, y_0) \in D$. Then f is continuous at $(x_0, y_0) \iff$ the following ϵ-δ condition holds: For every $\epsilon > 0$, there exists $\delta > 0$ such that $(x, y) \in D$ and

$$\|(x, y) - (x_0, y_0)\| < \delta \implies |f(x, y) - f(x_0, y_0)| < \epsilon.$$

Examples:

Let $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose $f(0, 0) := 0$.

(i) Suppose $f(x, y) := xy/(x^2 + y^2)$ for $(x, y) \neq (0, 0)$.
Let $(x_n, y_n) := (1/n, 1/n)$, $n \in \mathbb{N}$. Then $(x_n, y_n) \to (0, 0)$, but $f(x_n, y_n) \to 1/2 \neq f(0, 0)$. Hence f is not continuous at $(0, 0)$.

(ii) Suppose $f(x, y) := x^2y/(x^2 + y^2)$ for $(x, y) \neq (0, 0)$. Now

$$|f(x, y) - f(0, 0)| = |x|\frac{|xy|}{x^2 + y^2} \leq \frac{|x|}{2} \quad \text{for } (x, y) \in \mathbb{R}^2.$$

If $(x_n, y_n) \to (0, 0)$, then $f(x_n, y_n) \to 0 = f(0, 0)$. Hence f is continuous at 0.

(iii) Suppose $f(x, y) := x^2y/(x^4 + y^2)$ for $(x, y) \neq (0, 0)$.
If $(x_n, y_n) := (1/n, 1/n)$, then $f(x_n, y_n) = n/(1 + n^2) \to 0$.
Can we say f is continuous at $(0, 0)$?
Let $(x_n, y_n) := (1/n, 1/n^2)$, then $(x_n, y_n) \to (0, 0)$, but $f(x_n, y_n) \to 1/2 \neq f(0, 0)$. Hence f is not continuous at $(0, 0)$.

(iv) Suppose $f(x, y) := x^3y/(x^4 + y^2)$ for $(x, y) \neq (0, 0)$. Now

$$|f(x, y) - f(0, 0)| = |x|\frac{|x^2y|}{x^4 + y^2} \leq \frac{|x|}{2} \quad \text{for } (x, y) \in \mathbb{R}^2.$$

As in (ii) above, f is continuous at $(0, 0)$.

Limits of Functions of Two Variables

Definition

Let $D \subset \mathbb{R}^2$ and $f : D \to \mathbb{R}$. Suppose $(x_0, y_0) \in \mathbb{R}^2$ is such that $B((x_0, y_0), r) \setminus \{(x_0, y_0)\} \subset D$ for some $r > 0$ [or suppose, more generally, that (x_0, y_0) is a limit point of D, which means that every neighborhood of (x_0, y_0) contains a point of D other than (x_0, y_0)]. We say that $\lim_{(x,y)\to(x_0,y_0)} f(x, y)$ exists if there is $\ell \in \mathbb{R}$ such that $f(x_n, y_n) \to \ell$ for every sequence $((x_n, y_n))$ in D with $(x_n, y_n) \neq (x_0, y_0)$ for all $n \in \mathbb{N}$ and $(x_n, y_n) \to (x_0, y_0)$.

Examples: (i) Let

$$f(x,y) := \begin{cases} x + y & \text{if } (x,y) \in \mathbb{R}^2,\ x \neq y, \\ 1 & \text{if } (x,y) \in \mathbb{R}^2,\ x = y. \end{cases}$$

If $(x_n, y_n) := (1/n, 1/n)$, then $f(x_n, y_n) = 1$, and if $(x_n, y_n) := (1/n, -1/n)$, then $f(x_n, y_n) = 0$ for all $n \in \mathbb{N}$. Hence $\lim_{(x,y)\to(0,0)} f(x, y)$ does not exist.

(ii) Suppose $f(x,y) := xy/(x^2+y^2)$ for $(x,y) \neq (0,0)$.
If $(x_n, y_n) := (1/n, 1/n)$, then $f(x_n, y_n) = 1/2$, and
if $(x_n, y_n) := (1/n, -1/n)$, then $f(x_n, y_n) = -1/2$ for $n \in \mathbb{N}$.
Hence $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.

Aliter: Let $(x_n, y_n) := (1/n, (-1)^n/n)$, then for $n \in \mathbb{N}$,
$f(x_n, y_n) = (-1)^n/2$, and so the sequence $(f(x_n, y_n))$ is divergent. Hence $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.

(iii) Suppose $f(x,y) := x^2y/(x^4+y^2)$ for $(x,y) \neq (0,0)$.
Suppose $(x_n, y_n) \to (0,0)$ and $(x_n, y_n) \neq (0,0)$. If $x_n := 0$ or $y_n := 0$, then $f(x_n, y_n) = 0$. Also, if $\alpha \in \mathbb{R}$, $\alpha \neq 0$ and $y_n := \alpha x_n$ for $n \in \mathbb{N}$, then $f(x_n, y_n) = \alpha x_n/(x_n^2 + \alpha^2) \to 0$.

But if $(x_n, y_n) := (1/n, 1/n^2)$ for $n \in \mathbb{N}$, then $f(x_n, y_n) = 1/2$.
Hence $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.

Continuity and Limit

Let us relate the concepts of continuity and limit for functions of two variables.

Let $D \subset \mathbb{R}^2$. We say that $(x_0, y_0) \in \mathbb{R}^2$ is an **interior point** of D if there is $r > 0$ such that $B((x_0, y_0), r) \subset D$.

Suppose (x_0, y_0) be an interior point of D [or more generally, suppose $(x_0, y_0) \in D$ and (x_0, y_0) is a limit point of D]. Then $f : D \to \mathbb{R}$ is continuous at $(x_0, y_0) \iff \lim_{(x,y)\to(x_0,y_0)} f(x, y)$ exists and it is equal to $f(x_0, y_0)$.

Thus for f to be continuous at an interior point (x_0, y_0) of D,

(i) $f(x_0, y_0)$ has to be defined,

(ii) $\lim_{(x,y)\to(x_0,y_0)} f(x, y)$ must exist, and

(iii) $\lim_{(x,y)\to(x_0,y_0)} f(x, y) = f(x_0, y_0)$.

Limit theorems for functions of two real variables

Suppose $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \ell_f$ and $\lim_{(x,y)\to(x_0,y_0)} g(x,y) = \ell_g$.
The limit theorems for sequences in $\mathbb{R}$ show the following:

(i) $\lim_{(x,y)\to(x_0,y_0)} (f \pm g)(x,y) = \ell_f \pm \ell_g$ resp.

(ii) $\lim_{(x,y)\to(x_0,y_0)} (f \cdot g)(x,y) = \ell_f \cdot \ell_g$.

(iii) If $\ell_g \neq 0$, then $\lim_{(x,y)\to(x_0,y_0)} \left(\frac{f}{g}\right)(x,y) = \frac{\ell_f}{\ell_g}$.

(iv) If $f(x,y) \leq g(x,y)$ for all $(x,y) \neq (x_0,y_0)$ near (x_0,y_0), then $\ell_f \leq \ell_g$.

(v) (Sandwich theorem) If $f(x,y) \leq h(x,y) \leq g(x,y)$ for all (x,y) near (x_0,y_0) and $(x,y) \neq (x_0,y_0)$, and if $\ell_f = \ell = \ell_g$, then $\lim_{(x,y)\to(x_0,y_0)} h(x,y) = \ell$.

Equivalent condition for a limit to exist

(Recall the ϵ-δ condition for Continuity)

Let $D \subset \mathbb{R}^2$, $f : D \to \mathbb{R}$ and $(x_0, y_0) \in D$. Then f is continuous at (x_0, y_0) $\iff$ the following ϵ-δ condition holds: For every $\epsilon > 0$, there exists $\delta > 0$ such that

$$(x, y) \in D, \ \|(x, y) - (x_0, y_0)\| < \delta \implies |f(x, y) - f(x_0, y_0)| < \epsilon.$$

Similarly,

(ϵ-δ condition for Limit)

Let $D \subset \mathbb{R}^2$ and $f : D \to \mathbb{R}$ and $(x_0, y_0) \in \mathbb{R}^2$. Suppose $B((x_0, y_0), r) \setminus \{(x_0, y_0)\} \subset D$ for some $r > 0$ [or more generally, suppose, (x_0, y_0) is a limit point of D]. Let $\ell \in \mathbb{R}$. Then $\displaystyle\lim_{(x,y)\to(x_0,y_0)} f(x, y) = \ell \iff$ the following ϵ-δ condition holds: For every $\epsilon > 0$, there is $\delta > 0$ such that $(x, y) \in D$, $0 < \|(x, y) - (x_0, y_0)\| < \delta \implies |f(x, y) - \ell| < \epsilon$.

Examples:

(i) Let $f(x,y) := x^2y/(x^2+y^2)$ for $(x,y) \neq (0,0)$. Now

$$|f(x,y)| = |x|\frac{|xy|}{x^2+y^2} \leq \frac{|x|}{2} \quad \text{for } (x,y) \in \mathbb{R}^2.$$

Let $\epsilon > 0$. If $\delta := 2\epsilon$, then

$$0 < \|(x,y)\| < \delta \implies |f(x,y) - 0| \leq \frac{|x|}{2} \leq \frac{\|(x,y)\|}{2} < \frac{\delta}{2} = \epsilon.$$

Hence $\lim\limits_{(x,y)\to(0,0)} f(x,y) = 0$.

Aliter: If we define $f(0,0) := 0$, then f is continuous at $(0,0)$ as we have seen before, and so $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ exists and is equal to 0.

(ii) Let $f(x,y) := \dfrac{x^2+y^2}{1-\sqrt{x^2+y^2+1}}$ for $(x,y) \neq (0,0)$.

'Rationalizing' the fraction, we obtain

$f(x,y) = -1-\sqrt{x^2+y^2+1}$ for $(x,y) \neq (0,0)$, and so $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y) = -1-\sqrt{1} = -2$. (Check: $\delta := \epsilon$ works.)

(iii) Let $f(x,y) := y \sin \dfrac{1}{x^2+y^2}$ for $(x,y) \neq (0,0)$.

Since $|f(x,y)-0| \leq |y| \leq \|(x,y)\|$ for $(x,y) \neq (0,0)$, $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y) = 0$. (Clearly, $\delta := \epsilon$ works.)

Thumb rule: Suppose $f(x,y) = p(x,y)/q(x,y)$ for $(x,y) \neq (0,0)$, where $p(x,y)$ and $q(x,y)$ are homogeneous polynomials of degrees m and n respectively, and $p(0,0) = 0 = q(0,0)$. If $m > n$, then $f(x,y) \to 0$ as $(x,y) \to (0,0)$, and if $m \leq n$, then this limit does not exist.

MA 105 : Calculus
Division 1, Lecture 14

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Completion of Tut Sheet 5
- Sequence in $\mathbb{R}^2$ and its limit
- Functions on subsets of $\mathbb{R}^2$. Examples
- Continuity of real-valued functions defined on subsets of $\mathbb{R}^2$
- Basic properties of continuous functions
- Limits of functions of two variables. Examples
- Continuity and Limit
- Limit theorems for functions of two variables

Continuity and Limit

Let us relate the concepts of continuity and limit for functions of two variables.

Let $D \subset \mathbb{R}^2$. We say that $(x_0, y_0) \in \mathbb{R}^2$ is an **interior point** of D if there is $r > 0$ such that $B((x_0, y_0), r) \subset D$.

Suppose (x_0, y_0) be an interior point of D [or more generally, suppose $(x_0, y_0) \in D$ and (x_0, y_0) is a limit point of D]. Then $f : D \to \mathbb{R}$ is continuous at $(x_0, y_0) \iff \lim\limits_{(x,y)\to(x_0,y_0)} f(x, y)$ exists and it is equal to $f(x_0, y_0)$.

Thus for f to be continuous at an interior point (x_0, y_0) of D,

(i) $f(x_0, y_0)$ has to be defined,

(ii) $\lim\limits_{(x,y)\to(x_0,y_0)} f(x, y)$ must exist, and

(iii) $\lim\limits_{(x,y)\to(x_0,y_0)} f(x, y) = f(x_0, y_0)$.

Limit theorems for functions of two real variables

Suppose $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = \ell_f$ and $\lim_{(x,y)\to(x_0,y_0)} g(x,y) = \ell_g$.
The limit theorems for sequences in $\mathbb{R}$ show the following:

(i) $\lim_{(x,y)\to(x_0,y_0)} (f \pm g)(x,y) = \ell_f \pm \ell_g$ resp.

(ii) $\lim_{(x,y)\to(x_0,y_0)} (f \cdot g)(x,y) = \ell_f \cdot \ell_g$.

(iii) If $\ell_g \neq 0$, then $\lim_{(x,y)\to(x_0,y_0)} \left(\dfrac{f}{g}\right)(x,y) = \dfrac{\ell_f}{\ell_g}$.

(iv) If $f(x,y) \le g(x,y)$ for all $(x,y) \neq (x_0,y_0)$ near (x_0,y_0), then $\ell_f \le \ell_g$.

(v) (Sandwich theorem) If $f(x,y) \le h(x,y) \le g(x,y)$ for all (x,y) near (x_0,y_0) and $(x,y) \neq (x_0,y_0)$, and if $\ell_f = \ell = \ell_g$, then $\lim_{(x,y)\to(x_0,y_0)} h(x,y) = \ell$.

Equivalent condition for a limit to exist

(Recall the ϵ-δ condition for Continuity)

Let $D \subset \mathbb{R}^2$, $f : D \to \mathbb{R}$ and $(x_0, y_0) \in D$. Then f is continuous at $(x_0, y_0) \iff$ the following ϵ-δ condition holds: For every $\epsilon > 0$, there exists $\delta > 0$ such that

$$(x, y) \in D,\ \|(x, y) - (x_0, y_0)\| < \delta \implies |f(x, y) - f(x_0, y_0)| < \epsilon.$$

Similarly,

(ϵ-δ condition for Limit)

Let $D \subset \mathbb{R}^2$ and $f : D \to \mathbb{R}$ and $(x_0, y_0) \in \mathbb{R}^2$. Suppose $B((x_0, y_0), r) \setminus \{(x_0, y_0)\} \subset D$ for some $r > 0$ [or more generally, suppose, (x_0, y_0) is a limit point of D]. Let $\ell \in \mathbb{R}$. Then $\displaystyle\lim_{(x,y)\to(x_0,y_0)} f(x, y) = \ell \iff$ the following ϵ-δ condition holds: For every $\epsilon > 0$, there is $\delta > 0$ such that $(x, y) \in D$, $0 < \|(x, y) - (x_0, y_0)\| < \delta \implies |f(x, y) - \ell| < \epsilon$.

Examples:

(i) Let $f(x,y) := x^2y/(x^2+y^2)$ for $(x,y) \neq (0,0)$. Now

$$|f(x,y)| = |x|\frac{|xy|}{x^2+y^2} \leq \frac{|x|}{2} \quad \text{for } (x,y) \in \mathbb{R}^2.$$

Let $\epsilon > 0$. If $\delta := 2\epsilon$, then

$$0 < \|(x,y)\| < \delta \implies |f(x,y) - 0| \leq \frac{|x|}{2} \leq \frac{\|(x,y)\|}{2} < \frac{\delta}{2} = \epsilon.$$

Hence $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y) = 0$.

Aliter: If we define $f(0,0) := 0$, then f is continuous at $(0,0)$ as we have seen before, and so $\displaystyle\lim_{(x,y)\to(0,0)} f(x,y)$ exists and is equal to 0.

(ii) Let $f(x,y) := \dfrac{x^2+y^2}{1-\sqrt{x^2+y^2+1}}$ for $(x,y) \neq (0,0)$.

'Rationalizing' the fraction, we obtain

$f(x,y) = -1-\sqrt{x^2+y^2+1}$ for $(x,y) \neq (0,0)$, and so $\lim\limits_{(x,y)\to(0,0)} f(x,y) = -1-\sqrt{1} = -2$. (Check: $\delta := \epsilon$ works.)

(iii) Let $f(x,y) := y \sin \dfrac{1}{x^2+y^2}$ for $(x,y) \neq (0,0)$.

Since $|f(x,y)-0| \leq |y| \leq \|(x,y)\|$ for $(x,y) \neq (0,0)$, $\lim\limits_{(x,y)\to(0,0)} f(x,y) = 0$. (Clearly, $\delta := \epsilon$ works.)

Thumb rule: Suppose $f(x,y) = p(x,y)/q(x,y)$ for $(x,y) \neq (0,0)$, where $p(x,y)$ and $q(x,y)$ are homogeneous polynomials of degrees m and n respectively, and $p(0,0) = 0 = q(0,0)$. If $m > n$, then $f(x,y) \to 0$ as $(x,y) \to (0,0)$, and if $m \leq n$, then this limit does not exist.

Partial Differentiation

Recall: Let $D \subset \mathbb{R}$, and let c be an interior point of D. A function $f : D \to \mathbb{R}$ is said to have a **derivative** at c if $\lim\limits_{h \to 0} \dfrac{f(c+h) - f(c)}{h}$ exists, and then it is denoted by $f'(c)$.

Definition

Let $D \subset \mathbb{R}^2$, and let (x_0, y_0) be an interior point of D. A function $f : D \to \mathbb{R}$ is said to have a **partial derivative with respect to** x at (x_0, y_0) if

$$\lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h}$$

exists, and then it is denoted by $f_x(x_0, y_0)$ or by $\dfrac{\partial f}{\partial x}(x_0, y_0)$.

This definition also makes sense if there is $r > 0$ such that the horizontal line segment $\{(x, y_0) : x \in (x_0 - r, x_0 + r)\} \subset D$.

Geometric interpretation

Let C denote the curve obtained by intersecting the graph of f by the horizontal plane $y = y_0$. The partial derivative of f with respect to x at (x_0, y_0) equals the slope of the tangent to the curve C at (x_0, y_0). It can be interpreted as the rate of change in f along the x-axis at (x_0, y_0).

Computationally, $f_x(x_0, y_0)$ is obtained by differentiating f with respect to x at x_0, treating y as the constant y_0.

Similarly, a function $f : D \to \mathbb{R}$ is said to have a **partial derivative with respect to** y at (x_0, y_0) if

$$\lim_{k \to 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}$$

exists, and then it is denoted by $f_y(x_0, y_0)$ or by $\dfrac{\partial f}{\partial y}(x_0, y_0)$.

This definition also makes sense if there is $r > 0$ such that the vertical line segment $\{(x_0, y) : y \in (y_0 - r, y_0 + r)\} \subset D$.

Let $a < b$ and $c < d$. Consider

$$D := \{(x, c) \in \mathbb{R}^2 : a \le x \le b\} \cup \{(a, y) \in \mathbb{R}^2 : c \le y \le d\},$$

and let $f : D \to \mathbb{R}$. Although there is no interior point of D, we can still define $f_x^+(a, c)$, $f_x^-(b, c)$, $f_y^+(a, c)$ and $f_y^-(a, d)$ as the appropriate left hand and the right hand partial derivatives.

Partial derivatives of sums, products, quotients and compositions of functions of two variables can be found in exactly the same manner as the derivatives of a function of one variable.

Definition

If the partial derivatives of $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ of f exist at (x_0, y_0), then

$$(\nabla f)(x_0, y_0) := (f_x(x_0, y_0), f_y(x_0, y_0)) \in \mathbb{R}^2$$

is called the **gradient** of f at (x_0, y_0).

Partial Differentiation: Examples

(i) Let $f(x,y) := x^2 + y^2$ for $(x,y) \in \mathbb{R}^2$. Partial derivatives of f exist at all points of $\mathbb{R}^2$. In fact, for $(x_0, y_0) \in \mathbb{R}^2$,

$$f_x(x_0, y_0) = 2x_0 \quad \text{and} \quad f_y(x_0, y_0) = 2y_0.$$

(ii) Let $f(x,y) := \sqrt{x^2 + y^2}$ for $(x,y) \in \mathbb{R}^2$.

Let $(x_0, y_0) \neq (0,0)$. Then

$$f_x(x_0, y_0) = \frac{x_0}{\sqrt{x_0^2 + y_0^2}} \quad \text{and} \quad f_y(x_0, y_0) = \frac{y_0}{\sqrt{x_0^2 + y_0^2}}.$$

But f does not have either partial derivative at $(0,0)$ since $\lim\limits_{h \to 0} |h|/h$ does not exist. Note: f is continuous at $(0,0)$.

(iii) Let $f(x,y) := xy/(x^2 + y^2)$ if $(x,y) \neq (0,0)$, and $f(0,0) = 0$. It is easy to see that $f_x(0,0) = 0 = f_y(0,0)$. We have already seen that f is not continuous at $(0,0)$.

Higher Order Partial Derivatives

Let $D \subset \mathbb{R}^2$, and $f : D \to \mathbb{R}$. Suppose $f_x(x_0, y_0)$ exists for every $(x_0, y_0) \in D$. If the function $f_x : D \to \mathbb{R}$ has a partial derivative with respect to x at (x_0, y_0), then it is denoted by $f_{xx}(x_0, y_0)$ or by $\dfrac{\partial^2 f}{\partial x^2}(x_0, y_0)$.

Also, if $f_x : D \to \mathbb{R}$ has a partial derivative with respect to y at (x_0, y_0), then it is denoted by $f_{xy}(x_0, y_0)$ or by $\dfrac{\partial^2 f}{\partial y \partial x}(x_0, y_0)$.

Similarly, we define $f_{yy}(x_0, y_0)$ and $f_{yx}(x_0, y_0)$, or $\dfrac{\partial^2 f}{\partial y^2}(x_0, y_0)$ and $\dfrac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$.

In general, the **mixed partial derivatives** $f_{xy}(x_0, y_0)$ and $f_{yx}(x_0, y_0)$ may not be equal.

Mixed Partials

Example: Let $f(x,y) := x\,y\dfrac{x^2-y^2}{x^2+y^2}$ if $(x,y) \neq (0,0)$, and $f(0,0) := 0$. Then $f_x(0,y_0) = -y_0$ for $y_0 \in \mathbb{R}$, and $f_y(x_0,0) = x_0$ for $x_0 \in \mathbb{R}$. Hence $f_{xy}(0,0) = -1$ and $f_{yx}(0,0) = 1$.

Theorem (Mixed Partials Theorem)

Let $D \subset \mathbb{R}^2$, and let (x_0,y_0) be an interior point of D. Then there is $r > 0$ such that
$S := \{(x,y) \in \mathbb{R}^2 : |x - x_0| < r \text{ and } |y - y_0| < r\} \subset D$.
Consider $f : S \to \mathbb{R}$, and suppose f_x and f_y exist on S. If one of the mixed partials f_{xy} or f_{yx} exists on S, and it is continuous at (x_0,y_0), then the other mixed partial exists at (x_0,y_0), and $f_{xy}(x_0,y_0) = f_{yx}(x_0,y_0)$.

We omit a proof of this result since it is rather involved.

Directional Derivatives

Let $D \subset \mathbb{R}^2$, and let (x_0, y_0) be an interior point of D.
Let $\boldsymbol{u} := (u_1, u_2) \in \mathbb{R}^2$ be a **unit vector**, that is, $\|\boldsymbol{u}\| = 1$.

Definition

A function $f : D \to \mathbb{R}$ is said to have a **directional derivative along** $\boldsymbol{u}$ at (x_0, y_0) if

$$\lim_{t \to 0} \frac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0)}{t}$$

exists, and then it is denoted by $(\boldsymbol{D_u} f)(x_0, y_0)$.

This definition also makes sense if there is $r > 0$ such that the (slanted) line segment $\{(x_0 + tu_1, y_0 + tu_2) : t \in (-r, r)\} \subset D$.

$(\boldsymbol{D_u} f)(x_0, y_0)$ is the rate of change in f along $\boldsymbol{u}$ at (x_0, y_0).

Clearly, if $\boldsymbol{u} := (1,0)$, then $\boldsymbol{D_u} f = f_x$, and if $\boldsymbol{u} := (0,1)$, then $\boldsymbol{D_u} f = f_y$. Also, $\boldsymbol{D_{-u}} f = -\boldsymbol{D_u} f$ for every unit vector $\boldsymbol{u}$.

(Carathéodory's lemma for a directional derivative)

Let (x_0, y_0) be an interior point of $D \subset \mathbb{R}^2$. Let E denote the subset $\{(x_0 + tu_1, y_0 + tu_2) \in D : t \in \mathbb{R}\}$ of D. Then $(\boldsymbol{D_u} f)(x_0, y_0)$ exists $\iff$ there is a function $f_1 : E \to \mathbb{R}$ which is continuous at (x_0, y_0) and satisfies $f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0) = t\, f_1(x_0 + tu_1, y_0 + tu_2)$ for all $(x_0 + tu_1, y_0 + tu_2) \in E$.

In this case, $(\boldsymbol{D_u} f)(x_0, y_0) = f_1(x_0, y_0)$.

Proof. Suppose $(\boldsymbol{D_u} f)(x_0, y_0)$ exists. Define $f_1 : E \to \mathbb{R}$ by

$$f_1(x_0+tu_1, y_0+tu_2) := \begin{cases} \dfrac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0)}{t} & \text{if } t \neq 0, \\ (\boldsymbol{D_u} f)(x_0, y_0) & \text{if } t = 0. \end{cases}$$

Then f_1 is continuous at (x_0, y_0) since
$f_1(x_0 + tu_1, y_0 + tu_2) \to (\boldsymbol{D_u} f)(x_0, y_0) = f_1(x_0, y_0)$ as $t \to 0$.

Conversely, suppose there is $f_1 : E \to \mathbb{R}$ as stated. Then

$$\begin{aligned}\lim_{t\to 0} \frac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0)}{t} &= \lim_{t\to 0} f_1(x_0 + tu_1, y_0 + tu_2) \\ &= f_1(x_0, y_0),\end{aligned}$$

since f_1 is continuous at (x_0, y_0). Hence $(\boldsymbol{D_u} f)(x_0, y_0)$ exists.

In this case, it follows that $(\boldsymbol{D_u} f)(x_0, y_0) = f_1(x_0, y_0)$. $\square$

Examples of directional derivatives

Examples:

(i) Let $f(x,y) := x^2 + y^2$ for $(x,y) \in \mathbb{R}^2$.

Let $(x_0, y_0) \in \mathbb{R}^2$ and $\boldsymbol{u} := (u_1, u_2)$ be a unit vector. For $t \neq 0$,

$$\begin{aligned} & \frac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0)}{t} \\ = & \frac{(x_0 + tu_1)^2 + (y_0 + tu_2)^2 - x_0^2 - y_0^2}{t} \\ = & \ 2x_0u_1 + 2y_0u_2 + t. \end{aligned}$$

Letting $t \to 0$, we obtain

$(\boldsymbol{D_u} f)(x_0, y_0) = 2x_0u_1 + 2y_0u_2$.

Note: $(\boldsymbol{D_u} f)(x_0, y_0) = (\nabla f)(x_0, y_0) \cdot \boldsymbol{u}$ for every unit vector $\boldsymbol{u}$.

(ii) Let $f(x,y) := \sqrt{x^2+y^2}$ for $(x,y) \in \mathbb{R}^2$. Let $(x_0,y_0) \in \mathbb{R}^2$ and $\boldsymbol{u} := (u_1,u_2)$ be a unit vector. For $t \neq 0$,

$$\begin{aligned} & \frac{f(x_0+tu_1, y_0+tu_2) - f(x_0,y_0)}{t} \\ = \quad & \frac{\sqrt{(x_0+tu_1)^2+(y_0+tu_2)^2} - \sqrt{x_0^2+y_0^2}}{t} \\ = \quad & \frac{2x_0u_1+2y_0u_2+t}{\sqrt{(x_0+tu_1)^2+(y_0+tu_2)^2} + \sqrt{x_0^2+y_0^2}}. \end{aligned}$$

Hence if $(x_0,y_0) \neq (0,0)$, then

$$(\boldsymbol{D}_{\boldsymbol{u}}f)(x_0,y_0) = \frac{x_0u_1+y_0u_2}{\sqrt{x_0^2+y_0^2}} = (\nabla f)(x_0,y_0)\cdot \boldsymbol{u}.$$

But $(\boldsymbol{D}_{\boldsymbol{u}}f)(0,0)$ does not exist since $\lim_{t\to 0} t/|t|$ does not exist.

(iii) Let $f(x,y) := xy/(x^2+y^2)$ if $(x,y) \neq (0,0)$, and $f(0,0) := 0$. For a unit vector $\boldsymbol{u} := (u_1, u_2)$ and $t \neq 0$,

$$\frac{f(0+tu_1, 0+tu_2) - f(0,0)}{t} = \frac{u_1 u_2}{t}.$$

Hence $(D_{\boldsymbol{u}} f)(0,0)$ exists $\iff u_1 = 0$ or $u_2 = 0$.

(iv) Let $f(x,y) := x^2y/(x^4+y^2)$ if $(x,y) \neq (0,0)$, and $f(0,0) := 0$. For a unit vector $\boldsymbol{u} := (u_1, u_2)$ and $t \neq 0$,

$$\frac{f(0+tu_1, 0+tu_2) - f(0,0)}{t} = \frac{u_1^2 u_2}{t^2 u_1^4 + u_2^2}.$$

Hence $(D_{\boldsymbol{u}} f)(0,0) = 0$ if $u_2 = 0$, and $(D_{\boldsymbol{u}} f)(0,0) = u_1^2/u_2$ if $u_2 \neq 0$. In particular, $(\nabla f)(0,0) = (f_x(0,0), f_y(0,0)) = (0,0)$.

Thus $(D_{\boldsymbol{u}} f)(0,0) \neq (\nabla f)(0,0) \cdot \boldsymbol{u}$ unless $u_1 = 0$ or $u_2 = 0$.

(v) Let $f(x, y) := x^3y/(x^4 + y^2)$ if $(x, y) \neq (0, 0)$, and $f(0, 0) := 0$. For a unit vector $\boldsymbol{u} := (u_1, u_2)$ and $t \neq 0$,

$$\frac{f(0 + tu_1, 0 + tu_2) - f(0, 0)}{t} = \frac{tu_1^3 u_2}{t^2 u_1^4 + u_2^2}.$$

Hence $(\boldsymbol{D_u} f)(0, 0) = 0$ if $u_2 = 0$ and also if $u_2 \neq 0$. In particular, $(\nabla f)(0, 0) = (f_x(0, 0), f_y(0, 0)) = (0, 0)$

Thus $(\boldsymbol{D_u} f)(0, 0) = (\nabla f)(0, 0) \cdot \boldsymbol{u}$ for every unit vector $\boldsymbol{u}$.

Higher order directional derivatives can also be considered. But we shall not do so here.

Differentiation of a function of two real variables

Let $D \subset \mathbb{R}^2$, and $f : D \to \mathbb{R}$. Suppose (x_0, y_0) is an interior point of D. We have considered partial derivatives, and more generally, directional derivatives of f at (x_0, y_0). Each of them gives us an information about the rate of change in f at (x_0, y_0) in a given direction. Let us see how we can obtain complete information about the rate of change in f at (x_0, y_0).

Let us recall the one variable situation. If $D \subset \mathbb{R}$, $f : D \to \mathbb{R}$, and c is an interior point of D, then the derivative of f at c is

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

Analogously, we would have liked to define the 'derivative' of f at (x_0, y_0) by

$$\lim_{(h,k)\to(0,0)} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0)}{(h, k)}.$$

But this does not make sense since division by a vector $(h, k) \in \mathbb{R}^2$ is not defined even when $(h, k) \neq (0, 0)$.

We note that if $f : D \to \mathbb{R}$, c is an interior point of D and $\alpha \in \mathbb{R}$, then $f'(c) = \alpha \in \mathbb{R}$ if and only if

$$\lim_{h \to 0} \frac{f(c+h) - f(c) - \alpha h}{h} = 0,$$

that is,

$$\lim_{h \to 0} \frac{f(c+h) - f(c) - \alpha h}{|h|} = 0.$$

Hence if $D \subset \mathbb{R}^2$, (x_0, y_0) is an interior point of D and $f : D \to \mathbb{R}$, then it is natural to see if there is $(\alpha, \beta) \in \mathbb{R}^2$ such that

$$\lim_{(h,k) \to (0,0)} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0) - \alpha h - \beta k}{\|(h, k)\|} = 0.$$

Differentiability and Total Derivative

Suppose $D \subset \mathbb{R}^2$ and (x_0, y_0) is an interior point of D. Consider a funtion $f : D \to \mathbb{R}$.

Definition

We say that f is differentiable at (x_0, y_0) if there is $(\alpha, \beta) \in \mathbb{R}^2$ such that

$$\lim_{(h,k)\to(0,0)} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0) - \alpha h - \beta k}{\sqrt{h^2 + k^2}} = 0.$$

In this case, the pair $(\alpha, \beta) \in \mathbb{R}^2$ is called the **total derivative** of f at (x_0, y_0). Does it give us the full picture?

First, letting $(h, k) \to (0, 0)$ along the x-axis, we obtain

$$\lim_{h\to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0) - \alpha h}{|h|} = 0,$$

that is, $f_x(x_0, y_0) = \alpha$.

Next, letting $(h, k) \to (0, 0)$ along the y-axis, we obtain

$$\lim_{k \to 0} \frac{f(x_0, y_0 + k) - f(x_0, y_0) - \beta k}{|k|} = 0,$$

that is, $f_y(x_0, y_0) = \beta$.

Thus if f is differentiable at (x_0, y_0), and (α, β) is its total derivative, then both partials of f exist at (x_0, y_0), and $f_x(x_0, y_0) = \alpha$, $f_y(x_0, y_0) = \beta$, that is, $(\alpha, \beta) = (\nabla f)(x_0, y_0)$, the gradient of f at (x_0, y_0).

More generally, let $\boldsymbol{u} := (u_1, u_2)$ be a unit vector. Letting $(h, k) \to (0, 0)$ along the parametrized line $x = tu_1$, $y = tu_2$,

$$\lim_{t \to 0} \frac{f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0) - \alpha t u_1 - \beta t u_2}{|t|} = 0,$$

that is, $(\boldsymbol{D}_{\boldsymbol{u}} f)(x_0, y_0) = \alpha u_1 + \beta u_2 = (\alpha, \beta) \cdot \boldsymbol{u} = (\nabla f)(x_0, y_0) \cdot \boldsymbol{u}$.

Linear Maps and Derivatives

Let $m \in \mathbb{N}$. A function $L : \mathbb{R}^m \to \mathbb{R}$ is called a linear map if $L(r\boldsymbol{x} + s\boldsymbol{y}) = r\,L(\boldsymbol{x}) + s\,L(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ and $r, s \in \mathbb{R}$.

(i) A function $L : \mathbb{R} \to \mathbb{R}$ is a linear map if and only if there is $\alpha \in \mathbb{R}$ such that $L(x) = \alpha x$ for all $x \in \mathbb{R}$.

Let $D \subset \mathbb{R}$, c an interior point of D, and $f : D \to \mathbb{R}$. Then f is differentiable at c if and only if there is a linear map $L : \mathbb{R} \to \mathbb{R}$ such that $\lim\limits_{h \to 0} \dfrac{f(c+h) - f(c) - L(h)}{h} = 0$, that is,

$$f(c+h) = f(c) + L(h) + R(h), \text{ where } \frac{R(h)}{h} \to 0 \text{ as } h \to 0.$$

Thus the 'remainder term' $R(h)$ is even smaller than h.

The linear map L is known as the **derivative** of f at c.

(ii) A function $L : \mathbb{R}^2 \to \mathbb{R}$ is a linear map if and only if there are $\alpha, \beta \in \mathbb{R}$ such that $L(x, y) = \alpha x + \beta y$ for all $(x, y) \in \mathbb{R}^2$.

Let $D \subset \mathbb{R}^2$, (x_0, y_0) an interior point of D, and $f : D \to \mathbb{R}$. Then f is differentiable at (x_0, y_0) if and only if there is a linear map $L : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\lim_{(h,k)\to(0,0)} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0) - L(h, k)}{\|(h, k)\|} = 0, \text{ that is,}$$

$$f(x_0+h, y_0+k) = f(x_0, y_0)+L(h, k)+R(h, k), \text{ where } \frac{R(h, k)}{\|(h, k)\|} \to 0$$

as $\|(h, k)\| \to 0$.
Thus the 'remainder term' $R(h, k)$ is even smaller than $\|(h, k)\|$.

The linear map L is known as the **total derivative** of f at (x_0, y_0).

We now consider a two-dimensional analogue of the Carathéodory Lemma. As before, it relates the concept of differentiability to the concept of continuity.

Proposition (Increment Lemma)

Let $D \subset \mathbb{R}^2$, and let (x_0, y_0) be an interior point of D. Then $f : D \to \mathbb{R}$ is differentiable at (x_0, y_0) if and only if there exist functions $f_1, f_2 : D \to \mathbb{R}$ such that f_1 and f_2 are continuous at (x_0, y_0), and for all $(x, y) \in D$,

$$f(x, y) - f(x_0, y_0) = (x - x_0)f_1(x, y) + (y - y_0)f_2(x, y).$$

In this case, $\nabla f(x_0, y_0) = \big(f_1(x_0, y_0), f_2(x_0, y_0)\big)$.

Proof.

Let α, $\beta \in \mathbb{R}$. For $(x, y) \in D$, consider the 'remainder function'

$$F(x, y) := f(x, y) - f(x_0, y_0) - \alpha(x - x_0) - \beta(y - y_0).$$

Suppose $f : D \to \mathbb{R}$ is differentiable at (x_0, y_0). Then $\alpha := f_x(x_0, y_0)$ and $\beta := f_y(x_0, y_0)$ exist.
Define $f_1, f_2 : D \to \mathbb{R}$ as follows. Let $f_1(x_0, y_0) := \alpha$, $f_2(x_0, y_0) := \beta$, and for $(x, y) \neq (x_0, y_0)$, let

$$\begin{aligned} f_1(x,y) &:= \alpha + \frac{(x-x_0)F(x,y)}{(x-x_0)^2+(y-y_0)^2}, \\ f_2(x,y) &:= \beta + \frac{(y-y_0)F(x,y)}{(x-x_0)^2+(y-y_0)^2}. \end{aligned}$$

Let $(x, y) \in D$ with $(x, y) \neq (x_0, y_0)$. Then

$$\begin{aligned} &(x-x_0)f_1(x,y) + (y-y_0)f_2(x,y) \\ &= \alpha(x-x_0) + \frac{(x-x_0)^2F(x,y)}{(x-x_0)^2+(y-y_0)^2} \\ &\quad +\beta(y-y_0) + \frac{(y-y_0)^2F(x,y)}{(x-x_0)^2+(y-y_0)^2} \\ &= \alpha(x-x_0) + \beta(y-y_0) + F(x,y) = f(x,y) - f(x_0,y_0), \end{aligned}$$

and this equation also holds for $(x, y) := (x_0, y_0)$, as desired.

Also, since

$$\frac{|x - x_0|}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} \le 1 \quad \text{and} \quad \frac{|y - y_0|}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} \le 1,$$

and since f is differentiable at (x_0, y_0), we obtain

$$|f_1(x,y) - \alpha|, \ |f_2(x,y) - \beta| \le \left| \frac{F(x,y)}{\sqrt{(x-x_0)^2 + (y-y_0)^2}} \right| \to 0$$

as $(x, y) \to (x_0, y_0)$. Hence f_1 and f_2 are continuous at (x_0, y_0).

Conversely, suppose there are $f_1, f_2 : D \to \mathbb{R}$ as stated. Let $\alpha := f_1(x_0, y_0)$ and $\beta := f_2(x_0, y_0)$. Then for $(x, y) \in D$,

$$F(x,y) = (x - x_0)(f_1(x,y) - \alpha) + (y - y_0)(f_2(x,y) - \beta),$$

and so $|F(x,y)|/\|(x - x_0, y - y_0)\| \le |f_1(x,y) - \alpha| + |f_2(x,y) - \beta|$.

Now $|f_1(x,y)-\alpha|+|f_2(x,y)-\beta| \to 0$, since f_1 and f_2 are continuous at (x_0, y_0). Hence

$$\left|\frac{F(x,y)}{\sqrt{(x-x_0)^2+(y-y_0)^2}}\right| \to 0 \quad \text{as } (x,y) \to (x_0,y_0).$$

Thus f is differentiable at (x_0, y_0), and
$\nabla f(x_0,y_0) = (\alpha, \beta) = (f_1(x_0,y_0), f_2(x_0,y_0))$. □

A pair (f_1, f_2) of functions stated in the Increment Lemma is called a **pair of increment functions** associated with the function f and the point (x_0, y_0). Unlike the increment function f_1 stated in the Carathéodory Lemma, such a pair (f_1, f_2) of functions is not unique. Indeed, if $g : D \to \mathbb{R}$ is any function that is continuous at (x_0, y_0), and for $(x, y) \in D$, we let

$$\begin{aligned} g_1(x,y) &:= f_1(x,y) + (y-y_0)g(x,y), \\ g_2(x,y) &:= f_2(x,y) - (x-x_0)g(x,y), \end{aligned}$$

then (g_1, g_2) is also a corresponding pair of increment functions.

We now consider an important consequences of the Increment Lemma.

Proposition (Differentiability $\Longrightarrow$ Continuity)

Let $D \subset \mathbb{R}^2$, and let (x_0, y_0) be an interior point of D. Suppose $f : D \to \mathbb{R}$ is differentiable at (x_0, y_0). Then f is continuous at (x_0, y_0).

Proof:
If (f_1, f_2) is a pair of increment functions associated with f and (x_0, y_0), then

$$f(x, y) = f(x_0, y_0) + (x - x_0)f_1(x, y) + (y - y_0)f_2(x, y)$$

for all $(x, y) \in D$. Consequently, the continuity of f at (x_0, y_0) follows from the continuity of f_1 and f_2 at (x_0, y_0). $\square$

MA 105 : Calculus
Division 1, Lecture 15

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Continuity and Limit
- Limit theorems for functions of two variables
- Equivalent conditions for continuity and limits. Examples
- Partial differentiation
- Examples
- Higher order partial derivatives. Mixed Partials Theorem
- Directional derivatives. Examples
- Differentiation of a function of two real variables
- Differentiability and Total Derivative
- Linear Maps and Derivatives
- Increment Lemma. Application: Differentiability implies Continuity

Differentiability and Total Derivative

Suppose $D \subset \mathbb{R}^2$ and (x_0, y_0) is an interior point of D. Consider a funtion $f : D \to \mathbb{R}$.

Definition

We say that f is differentiable at (x_0, y_0) if there is $(\alpha, \beta) \in \mathbb{R}^2$ such that

$$\lim_{(h,k)\to(0,0)} \frac{f(x_0 + h, y_0 + k) - f(x_0, y_0) - \alpha h - \beta k}{\sqrt{h^2 + k^2}} = 0.$$

In this case, the pair $(\alpha, \beta) \in \mathbb{R}^2$ is called the **total derivative** of f at (x_0, y_0). We have seen that if f is differentiable at (x_0, y_0), then both the partial derivatives at (x_0, y_0) exist and in fact, $(\alpha, \beta) = (\nabla f)(x_0, y_0)$. Moreover, directional derivatives at (x_0, y_0) exists along any unit vector $\boldsymbol{u}$ exist, and

$$(\boldsymbol{D_u} f)(x_0, y_0) = (\nabla f)(x_0, y_0) \cdot \boldsymbol{u}.$$

Increment Lemma

We have the following two-dimensional analogue of the Carathéodory Lemma.

Proposition (Increment Lemma)

Let $D \subset \mathbb{R}^2$, and let (x_0, y_0) be an interior point of D. Then $f : D \to \mathbb{R}$ is differentiable at (x_0, y_0) if and only if there exist functions $f_1, f_2 : D \to \mathbb{R}$ such that f_1 and f_2 are continuous at (x_0, y_0), and for all $(x, y) \in D$,

$$f(x, y) - f(x_0, y_0) = (x - x_0)f_1(x, y) + (y - y_0)f_2(x, y).$$

In this case, $\nabla f(x_0, y_0) = \big(f_1(x_0, y_0), f_2(x_0, y_0)\big)$.

A pair (f_1, f_2) of functions stated in the Increment Lemma is called a **pair of increment functions** associated with the function f and the point (x_0, y_0).

We now consider an important consequences of the Increment Lemma.

Proposition (Differentiability $\Longrightarrow$ Continuity)

Let $D \subset \mathbb{R}^2$, and let (x_0, y_0) be an interior point of D. Suppose $f : D \to \mathbb{R}$ is differentiable at (x_0, y_0). Then f is continuous at (x_0, y_0).

Proof:
If (f_1, f_2) is a pair of increment functions associated with f and (x_0, y_0), then

$$f(x, y) = f(x_0, y_0) + (x - x_0)f_1(x, y) + (y - y_0)f_2(x, y)$$

for all $(x, y) \in D$. Consequently, the continuity of f at (x_0, y_0) follows from the continuity of f_1 and f_2 at (x_0, y_0). $\square$

Differentiability: Necessary Conditions

Let (x_0, y_0) be an interior point of a subset D of $\mathbb{R}^2$, and let $f : D \to \mathbb{R}$. Suppose f is differentiable at (x_0, y_0). Then the following conditions necessarily hold, that is, they are necessary conditions.

- Both partial derivatives $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ exist.
- The directional derivative $(\boldsymbol{D_u} f)(x_0, y_0)$ exists for every unit vector $\boldsymbol{u} \in \mathbb{R}^2$.
- $(\boldsymbol{D_u} f)(x_0, y_0) = \big(f_x(x_0, y_0), f_y(x_0, y_0)\big) \cdot \boldsymbol{u} = (\nabla f)(x_0, y_0) \cdot \boldsymbol{u}$ for every unit vector $\boldsymbol{u} \in \mathbb{R}^2$.
- f is continuous at (x_0, y_0).

None of the above conditions are sufficient to conclude that a function is differentiable. In fact, we shall give an example of a function for which all the above conditions hold at (x_0, y_0), but the function is not differentiable at (x_0, y_0) (Example (vii)).

Proposition (Sufficient condition for differentiability)

Let $(x_0, y_0) \in \mathbb{R}^2$, $r > 0$, and let $f : S \to \mathbb{R}$, where $S := \{(x, y) \in \mathbb{R}^2 : |x - x_0| < r \text{ and } |y - y_0| < r\}$. Suppose one of the partial derivatives of f exists on S and is continuous at (x_0, y_0), while the other exists at (x_0, y_0). Then f is differentiable at (x_0, y_0).

Proof. Without loss of generality, suppose f_x exists on S and is continuous at (x_0, y_0), while f_y exists at (x_0, y_0).
For $(x, y) \in S$, $f(x, y) - f(x_0, y_0) = A(x, y) + B(y)$, where

$$A(x, y) := f(x, y) - f(x_0, y) \quad \text{and} \quad B(y) := f(x_0, y) - f(x_0, y_0).$$

Define $f_1, f_2 : S \to \mathbb{R}$ as follows.

$$f_1(x, y) := \begin{cases} \dfrac{A(x, y)}{x - x_0} & \text{if } x \neq x_0, \\ f_x(x_0, y) & \text{if } x = x_0, \end{cases} \qquad f_2(x, y) := \begin{cases} \dfrac{B(y)}{y - y_0} & \text{if } y \neq y_0, \\ f_y(x_0, y_0) & \text{if } y = y_0. \end{cases}$$

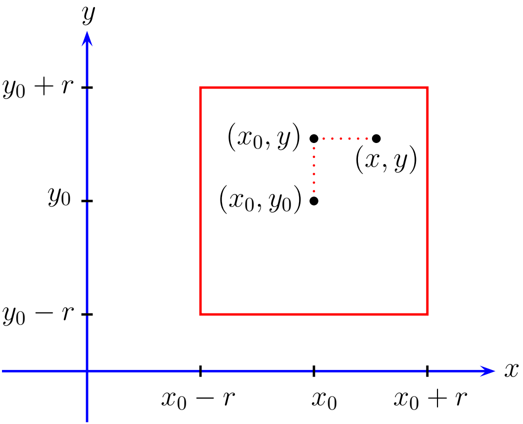


Figure: The dotted lines linking (x, y) to (x_0, y), and (x_0, y) to (x_0, y_0).

By considering the cases $x \neq x_0, \; x = x_0, \; y \neq y_0$ and $y = y_0$, it follows that for all $(x, y) \in S$,

$$f(x, y) - f(x_0, y_0) = (x - x_0) f_1(x, y) + (y - y_0) f_2(x, y).$$

To show f_1 is continuous at (x_0, y_0), consider $x \in (x_0 - r, x_0) \cup (x_0, x_0 + r)$ and $y \in (y_0 - r, y_0 + r)$. Since f_x exists on S, there is $c_y \in \mathbb{R}$ between x_0 and x such that

$$A(x, y) = f(x, y) - f(x_0, y) = (x - x_0) f_x(c_y, y).$$

by the MVT. Hence $f_1(x, y) = f_x(c_y, y)$. Also, by definition, $f_1(x_0, y) = f_x(x_0, y)$ for all $y \in (y_0 - r, y_0 + r)$. Since f_x is continuous at (x_0, y_0), we see that f_1 is continuous at (x_0, y_0).

Also, f_2 is continuous at (x_0, y_0) since $f_y(x_0, y_0)$ exists.

By the Increment Lemma, f is differentiable at (x_0, y_0). $\square$

How to check differentiability?

To check the differentiability of a function of two variables, we may see whether f violates one of the necessary conditions for differentiability. Also, we may see whether f satisfies a sufficient condition for differentiability. Alternatively, we may turn to the definition of differentiability itself.

Examples:

(i) Let $f(x, y) := \sqrt{x^2 + y^2}$ for $(x, y) \in \mathbb{R}^2$.
Then f is continuous at $(0, 0)$. But $(\boldsymbol{D_u} f)(0, 0)$ does not exist for any unit vector $\boldsymbol{u}$. Hence f is not differentiable at $(0, 0)$.

(ii) Let $f(x, y) := 1$ if $0 < y < x^2$ and $f(x, y) := 0$ otherwise.
If $\boldsymbol{u} := (u_1, u_2)$, then $f(tu_1, tu_2) = 0$ for all t in an interval about 0, and so $(\boldsymbol{D_u} f)(0, 0) = 0 = (\nabla f)(0, 0) \cdot \boldsymbol{u}$. But f is not continuous at $(0, 0)$ since $f(0, 0) = 0$ and $f(1/n, 1/2n^2) = 1$ for all $n \in \mathbb{N}$. Hence f is not differentiable at $(0, 0)$.

(iii) Let $f(x,y) := x^2y/(x^2+y^2)$ for $(x,y) \neq (0,0)$, and $f(0,0) := 0$. Then f is continuous at $(0,0)$, and $(\boldsymbol{D_u} f)(0,0) = u_1^2 u_2$ for every unit vector $\boldsymbol{u} := (u_1, u_2)$. But $(\boldsymbol{D_u} f)(0,0) \neq (\nabla f)(0,0) \cdot \boldsymbol{u}$ unless $u_1 = 0$ or $u_2 = 0$. Hence f is not differentiable at $(0,0)$.

(iv) Let $f(x,y) := x^2 + y^2$ for $(x,y) \in \mathbb{R}^2$. Let $(x_0, y_0) \in \mathbb{R}^2$. Then f is continuous at (x_0, y_0) and for every unit vector $\boldsymbol{u}$, $(\boldsymbol{D_u} f)(x_0, y_0) = (\nabla f)(x_0, y_0) \cdot \boldsymbol{u}$. Since f_x and f_y exist on $\mathbb{R}^2$ and f_x is continuous at (x_0, y_0), f is differentiable at (x_0, y_0).

(v) Let $f(x,y) := x^2y^2/(x^4+y^2)$ for $(x,y) \neq (0,0)$, and $f(0,0) := 0$. We check that f is continuous at $(0,0)$, and $(\boldsymbol{D_u} f)(0,0) = 0 = (\nabla f)(0,0) \cdot \boldsymbol{u}$ for every unit vector $\boldsymbol{u}$.

Let us find $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ for $(x_0, y_0) \neq (0,0)$.

For $(x_0, y_0) \neq (0,0)$, it can be checked that

$$f_x(x_0, y_0) = \frac{2x_0 y_0^2 (y_0^2 - x_0^4)}{(x_0^4 + y_0^2)^2} \quad \text{and} \quad f_y(x_0, y_0) = \frac{2x_0^6 y_0}{(x_0^4 + y_0^2)^2}.$$

Now f_x is continuous at $(0,0)$ as $y_0^4, 2x_0^4 y_0^2 \le (x_0^4 + y_0^2)^2$. Hence f is differentiable at $(0,0)$. (We note that f_y is not continuous at $(0,0)$ since $f_y(1/n, 1/n^2) = 1/2$ for $n \in \mathbb{N}$.) (Aliter: Here is a direct proof. For $(h,k) \neq (0,0)$,

$$0 \le \frac{h^2k^2 - 0 - 0\cdot h - 0\cdot k}{(h^4 + k^2)\sqrt{h^2 + k^2}} = h\,\frac{k^2}{(h^4 + k^2)}\frac{h}{\sqrt{h^2 + k^2}} \le |h| \to 0$$

as $(h,k) \to (0,0)$, and so f is differentiable at $(0,0)$.)

(vi) Let $f(x, y) := |xy|$ for $(x, y) \in \mathbb{R}^2$. Then f is continuous at $(0, 0)$, and $(\boldsymbol{D_u} f)(0, 0) = 0 = (\nabla f)(0, 0) \cdot \boldsymbol{u}$ for every unit vector $\boldsymbol{u}$. On the other hand, $f_x(0, y_0)$ does not exist if $y_0 \neq 0$ and $f_y(x_0, 0)$ does not exist if $x_0 \neq 0$. For $(h, k) \neq (0, 0)$,

$$Q(h, k) := \frac{f(h, k)}{\sqrt{h^2 + k^2}} = \frac{|hk|}{\sqrt{h^2 + k^2}} \leq |h| \to 0$$

as $(h, k) \to (0, 0)$. Hence f is differentiable at $(0, 0)$.

(vii) Let $f(x, y) := x^3 y/(x^4 + y^2)$ for $(x, y) \neq (0, 0)$, and $f(0, 0) := 0$. Then $(\boldsymbol{D_u} f)(0, 0) = 0 = (\nabla f)(0, 0) \cdot \boldsymbol{u}$ for every unit vector $\boldsymbol{u}$. Also, f is continuous at $(0, 0)$. Further, f_x and f_y are not continuous at $(0, 0)$. (Check!) For $(h, k) \neq (0, 0)$,

$$Q(h, k) := \frac{f(h, k)}{\sqrt{h^2 + k^2}} = \frac{h^3 k}{(h^4 + k^2)\sqrt{h^2 + k^2}} \not\to 0$$

as $(h, k) \to (0, 0)$ since $Q(1/n, 1/n^2) \to 1/2$. Hence f is not differentiable at $(0, 0)$.

We obtain basic properties of differentiable functions, and also the gradients of a sum, of a difference, of a product and of a quotient of functions from the Increment Lemma as follows.

Proposition

Let $D \subset \mathbb{R}^2$ and let (x_0, y_0) be an interior point of D. Suppose $f, g : D \to \mathbb{R}$ are functions that are differentiable at (x_0, y_0). Then $f \pm g$ and fg are differentiable at (x_0, y_0), and moreover,

$$\nabla(f \pm g)(x_0, y_0) = (\nabla f)(x_0, y_0) \pm (\nabla g)(x_0, y_0) \text{ resp.,}$$

$$\nabla(fg)(x_0, y_0) = g(x_0, y_0)(\nabla f)(x_0, y_0) + f(x_0, y_0)(\nabla g)(x_0, y_0).$$

If $g(x_0, y_0) \neq 0$, then f/g is differentiable at (x_0, y_0) and

$$\nabla\left(\frac{f}{g}\right)(x_0, y_0) = \frac{g(x_0, y_0)(\nabla f)(x_0, y_0) - f(x_0, y_0)(\nabla g)(x_0, y_0)}{g(x_0, y_0)^2}.$$

Proposition (Chain Rules)

Let $D \subset \mathbb{R}^2$ and let (x_0, y_0) be an interior point of D, and let $f : D \to \mathbb{R}$ be differentiable at (x_0, y_0).

(i) Let $E \subset \mathbb{R}$ be such that $f(D) \subset E$ and $z_0 := f(x_0, y_0)$ is an interior point of E. If $g : E \to \mathbb{R}$ is differentiable at $f(x_0, y_0)$, then the function $h := g \circ f : D \to \mathbb{R}$ is differentiable at (x_0, y_0),

$$h_x(x_0, y_0) = g'(z_0)\, f_x(x_0, y_0) \quad \text{and} \quad h_y(x_0, y_0) = g'(z_0)\, f_y(x_0, y_0).$$

(ii) Let $E \subset \mathbb{R}$, and let t_0 be an interior point of E. If $x, y : E \to \mathbb{R}$ are differentiable at t_0, and if $(x(t), y(t)) \in D$ for all $t \in E$ and $(x(t_0), y(t_0)) := (x_0, y_0)$, then the function $\phi : E \to \mathbb{R}$ defined by $\phi(t) := f(x(t), y(t))$ for $t \in E$ is differentiable at t_0, and

$$\phi'(t_0) = f_x(x_0, y_0)\, x'(t_0) + f_y(x_0, y_0)\, y'(t_0).$$

Proposition (Chain Rules: continued)

(iii) Let $E \subset \mathbb{R}^2$, and let (u_0, v_0) be an interior point of E. If $x, y : E \to \mathbb{R}$ are differentiable at (u_0, v_0), and if $(x(u, v), y(u, v)) \in D$ for all $(u, v) \in E$ and $(x(u_0, v_0), y(u_0, v_0)) := (x_0, y_0)$, then the function $F : E \to \mathbb{R}$ defined by $F(u, v) := f(x(u, v), y(u, v))$ for $(u, v) \in E$ is differentiable at (u_0, v_0),

$$F_u(u_0, v_0) = f_x(x_0, y_0)\, x_u(u_0, v_0) + f_y(x_0, y_0)\, y_u(u_0, v_0),$$

and

$$F_v(u_0, v_0) = f_x(x_0, y_0)\, x_v(u_0, v_0) + f_y(x_0, y_0)\, y_v(u_0, v_0).$$

It is often helpful to write the identities given by the Chain Rules in an informal but suggestive notation as follows.

(i) If $z = f(x, y)$ and $w = g(z)$, then w is a function of (x, y), and $\dfrac{\partial w}{\partial x} = \dfrac{dw}{dz}\dfrac{\partial z}{\partial x}, \quad \dfrac{\partial w}{\partial y} = \dfrac{dw}{dz}\dfrac{\partial z}{\partial y}.$

(ii) If $z = f(x, y)$ and if $x = x(t)$, $y = y(t)$, then z is a function of t, and

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

(iii) If $z = f(x, y)$ and if $x = x(u, v)$, $y = y(u, v)$, then z is a function of u and v, and

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u} \quad \text{and} \quad \frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v}.$$

It should be noted that the identities in (i), (ii), and (iii) above are valid when the concerned (partial) derivatives are evaluated at appropriate points and when, for instance, each of the (partial) derivatives exists in a neighbourhood of the relevant point and is continuous at that point.

The conclusions of Chain Rules (i), (ii) and (iii) can be written in terms of the gradient $(\nabla f)(x_0, y_0)$ of f at (x_0, y_0) as follows:

$$
\begin{aligned}
\text{(i)} \quad (\nabla h)(x_0, y_0) &= g'(z_0)(\nabla f)(x_0, y_0) \\
\text{(ii)} \quad \phi'(t_0) &= (\nabla f)(x_0, y_0) \cdot \big(x'(t_0), y'(t_0)\big) \\
\text{(iii)} \quad (\nabla F)(u_0, v_0) &= \Big((\nabla f)(x_0, y_0) \cdot \big(x_u(u_0, v_0), y_u(u_0, v_0)\big), \\
&\qquad (\nabla f)(x_0, y_0) \cdot \big(x_v(u_0, v_0), y_v(u_0, v_0)\big)\Big)
\end{aligned}
$$

The Chain Rules given above can be proved by using the Increment Lemma.

Examples illustrating the Chain Rules:

(i) Define $f : \mathbb{R}^2 \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ by $f(x, y) := xy$ for $(x, y) \in \mathbb{R}^2$, and $g(z) := \sin z$ for $z \in \mathbb{R}$. By the Chain Rule, the composite function $(g \circ f)(x, y) := \sin(xy)$ is differentiable at every point of $\mathbb{R}^2$, and

$$\frac{\partial (g \circ f)}{\partial x} = \frac{dg}{dz}\frac{\partial z}{\partial x} = (\cos xy)y \;\text{ and }\; \frac{\partial (g \circ f)}{\partial y} = \frac{dg}{dz}\frac{\partial z}{\partial y} = (\cos xy)x.$$

(ii) Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) := x^2 + y^2$ for $(x, y) \in \mathbb{R}^2$, and $x, y : \mathbb{R} \to \mathbb{R}$ by $x(t) := e^t$ and $y(t) := \sin t$ for $t \in \mathbb{R}$. By the Chain Rule, the composite function $\phi : \mathbb{R} \to \mathbb{R}$ given by $\phi(t) := f(x(t), y(t)) = e^{2t} + \sin^2 t$ is differentiable at every point of $\mathbb{R}$, and

$$\begin{aligned}\frac{d\phi}{dt} &= \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = (2x(t))e^t + (2y(t))\cos t \\ &= 2e^{2t} + 2\sin t \cos t.\end{aligned}$$

(iii) As in (ii) above, define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x, y) := x^2 + y^2$ for $(x, y) \in \mathbb{R}^2$, and $x, y : \mathbb{R}^2 \to \mathbb{R}$ by $x(u, v) := u^2 - v^2$ and $y(u, v) := 2uv$ for $(u, v) \in \mathbb{R}^2$. By the Chain Rule, the composite function $F : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\begin{aligned} F(u, v) &:= f(x(u, v), y(u, v)) = (u^2 - v^2)^2 + (2uv)^2 \\ &= u^4 + 2u^2v^2 + v^4 \end{aligned}$$

is differentiable at every point of $\mathbb{R}^2$, and

$$\begin{aligned} \frac{\partial F}{\partial u} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u} = 2x(2u) + 2y(2v) \\ &= 2(u^2 - v^2)(2u) + 2(2uv)(2v) = 4u(u^2 + v^2), \\ \frac{\partial F}{\partial v} &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v} = 2x(-2v) + 2y(2u) \\ &= 2(u^2 - v^2)(-2v) + 2(2uv)(2u) = 4v(u^2 + v^2). \end{aligned}$$

Geometric Interpretation of the Gradient

Suppose $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ is differentiable at an interior point (x_0, y_0) of D, and suppose $(\nabla f)(x_0, y_0) \neq (0, 0)$. Then for every unit vector $\boldsymbol{u}$,

$$(\boldsymbol{D_u} f)(x_0, y_0) = (\nabla f)(x_0, y_0) \cdot \boldsymbol{u} = \|(\nabla f)(x_0, y_0)\| \cos \theta,$$

where $\theta \in [0, \pi]$ is the angle between $(\nabla f)(x_0, y_0)$ and $\boldsymbol{u}$.
Hence

- $(\boldsymbol{D_u} f)(x_0, y_0)$ is maximum when $\cos \theta = 1$, that is, $\theta = 0$, so that $\boldsymbol{u} = \dfrac{(\nabla f)(x_0, y_0)}{\|(\nabla f)(x_0, y_0)\|}$.
- $(\boldsymbol{D_u} f)(x_0, y_0)$ is minimum when $\cos \theta = -1$, that is, $\theta = \pi$, so that $\boldsymbol{u} = -\dfrac{(\nabla f)(x_0, y_0)}{\|(\nabla f)(x_0, y_0)\|}$.
- $(\boldsymbol{D_u} f)(x_0, y_0) = 0$ when $\cos \theta = 0$, that is, $\theta = \pi/2$, so that $\boldsymbol{u}$ is perpendicular to $(\nabla f)(x_0, y_0)$.

Example:

Let $f(x,y) := 4 - x^2 - y^2$ for $(x,y) \in \mathbb{R}^2$. Then f is differentiable on $\mathbb{R}^2$, and $(\nabla f)(x_0,y_0) = (-2x_0, -2y_0)$ for all $(x_0,y_0) \in \mathbb{R}^2$.

Let $(x_0,y_0) := (1,1)$. Then $(\nabla f)(1,1) = (-2,-2)$.

Consider the surface $z = f(x,y)$. At $(1,1)$,

- the direction of steepest ascent is $(-2,-2)/\|(-2,-2)\| = (-1/\sqrt{2}, -1/\sqrt{2})$,
- the direction of steepest descent is $-(-2,-2)/\|(-2,-2)\| = (1/\sqrt{2}, 1/\sqrt{2})$,
- the directions $(u_, u_2)$ of no change (in height) satisfy $(-2,-2)\cdot(u_1,u_2) = 0$, and so, they are $(1/\sqrt{2}, -1/\sqrt{2})$ and $(-1/\sqrt{2}, 1/\sqrt{2})$.

See the picture on the next slide.

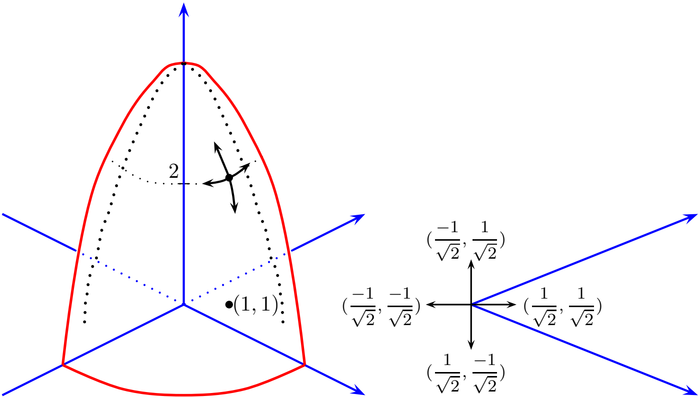


Figure: Directions of steepest ascent, of steepest decent and of no change in height

Tangent Line

Recall the one variable situation:
Let $a < b$ and $x_0 \in (a, b)$. Let a function $f : (a, b) \to \mathbb{R}$ be differentiable at x_0. Then the equation of the **tangent line** to the curve $y = f(x)$ in $\mathbb{R}^2$ at $(x_0, f(x_0))$ is given by

$$y - f(x_0) = f'(x_0)(x - x_0).$$

More generally, suppose $D \subset \mathbb{R}^2$ and $P_0 := (x_0, y_0)$ is an interior point of D. Let a function $F : D \to \mathbb{R}$ have partial derivatives at P_0, and let $(\nabla F)(P_0) \neq (0, 0)$. Suppose F defines a curve C in $\mathbb{R}^2$ (implicitly) by the equation $F(x, y) = 0$ for $(x, y) \in D$, and P_0 lies on C. Then the equation of the **tangent line** to C at P_0 is given by

$$F_x(P_0)(x - x_0) + F_y(P_0)(y - y_0) = 0.$$

Note: If $D := (a, b) \times \mathbb{R}$, $f : (a, b) \to \mathbb{R}$, and for $(x, y) \in D$, $F(x, y) := y - f(x)$, then $y_0 = f(x_0)$, $F_x(P_0) = -f'(x_0)$ and $F_y(P_0) = 1$. We recover the earlier equation of the tangent line.

Tangent Plane

Let us turn to the two variable situation:
Let $D \subset \mathbb{R}^2$, (x_0, y_0) be an interior point of D, and let a function $f : D \to \mathbb{R}$ be differentiable at (x_0, y_0). Then the equation of the **tangent plane** to the surface $z = f(x, y)$ in $\mathbb{R}^3$ at $(x_0, y_0, f(x_0, y_0))$ is given by

$$z - f(x_0, y_0) = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

More generally, suppose $E \subset \mathbb{R}^3$ and $P_0 := (x_0, y_0, z_0)$ is an interior point of E. Let a function $F : E \to \mathbb{R}$ have partial derivatives at P_0, and let $(\nabla F)(P_0) \neq (0, 0, 0)$. Suppose F (implicitly) defines a surface S in $\mathbb{R}^3$ by the equation $F(x, y, z) = 0$ for $(x, y, z) \in E$, and P_0 lies on S. Then the equation of the **tangent plane** to S at P_0 is given by

$$F_x(P_0)(x - x_0) + F_y(P_0)(y - y_0) + F_z(P_0)(z - z_0) = 0.$$

Note: As before, if we let $F(x, y, z) := z - f(x, y)$, then we recover the earlier equation of the tangent plane.

Normal Line

Now $(\nabla F)(P_0) \cdot (x - x_0, y - y_0, z - z_0) = 0$ for all (x, y, z) on the tangent plane to the surface S, that is, $(\nabla F)(P_0)$ is perpendicular to the tangent plane. The line passing through $P_0 = (x_0, y_0, z_0)$ and parallel to the nonzero vector $(\nabla F)(P_0) = (F_x(P_0), F_y(P_0), F_z(P_0))$ is called the **normal line** to the surface defined by $F(x, y, z) = 0$ at P_0. The parametric equations of this normal line are

$$x = x_0 + F_x(P_0)t, \quad y = y_0 + F_y(P_0)t, \quad z = z_0 + F_z(P_0)t, \quad t \in \mathbb{R}.$$

If all $F_x(P_0), F_y(P_0), F_z(P_0)$ are nonzero, then the equations are

$$\frac{x - x_0}{F_x(P_0)} = \frac{y - y_0}{F_y(P_0)} = \frac{z - z_0}{F_z(P_0)}.$$

Also, the parametric equations of the normal line to the surface defined by $z - f(x, y) = 0$ at (x_0, y_0, z_0) are

$$x = x_0 - f_x(x_0, y_0)t, \; y = y_0 - f_y(x_0, y_0)t, \; z = f(x_0, y_0) + t, \; t \in \mathbb{R}.$$

Tangent Vector to a Curve

We now describe a crucial property of the normal line to a surface. Let C be a parametrized smooth curve in $\mathbb{R}^3$ given by

$$t \longmapsto (x(t), y(t), z(t))\,, \quad t \in [\alpha, \beta].$$

Let $t_0 \in (\alpha, \beta)$ be such that $(x'(t_0), y'(t_0), z'(t_0)) \neq (0, 0, 0)$. Then the vector $\big(x'(t_0), y'(t_0), z'(t_0)\big)$ is called the **tangent vector** to the curve C at the point $\big(x(t_0), y(t_0), z(t_0)\big)$ on it.

Now suppose that the curve C lies on a surface S defined by $F(x, y, z) = 0$, that is, $\phi(t) := F\,(x(t), y(t), z(t)) = 0$ for all $t \in [\alpha, \beta]$. By the Chain Rule,

$$0 = \phi'(t_0) = F_x(P_0)x'(t_0) + F_y(P_0)y'(t_0) + F_z(P_0)z'(t_0).$$

Thus the normal line to the surface S at P_0 is perpendicular to tangent vectors to all curves on S that pass through P_0.

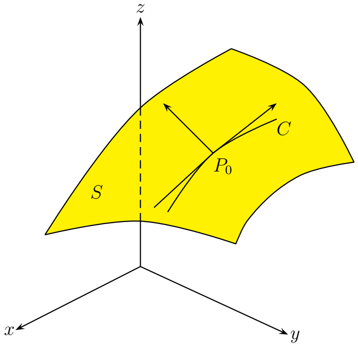


Figure: Normal line to the surface S at P_0 and a curve C on S

Special case: Suppose a surface S is given by $z = f(x, y)$ for $(x, y) \in D$. Let $Q_0 := (x_0, y_0) \in D$, and let a parametrized curve given by $(x(t), y(t))$, $t \in [\alpha, \beta]$, pass through Q_0. Let $z(t) := f(x(t), y(t))$, $t \in [\alpha, \beta]$. Then the curve C given by $(x(t), y(t), z(t))$, $t \in [\alpha, \beta]$, lies on S, and passes through $P_0 := f(Q_0)$.

Hence the normal line

$$x = x_0 - f_x(Q_0)t,\ y = y_0 - f_y(Q_0)t,\ z = f(Q_0) + t,\ t \in \mathbb{R},$$

to the surface S at P_0 is perpendicular to the tangent vector $(x'(t_0), y'(t_0), z'(t_0))$ to the curve C at P_0, that is, $(-f_x(Q_0), -f_y(Q_0), 1) \cdot (x'(t_0), y'(t_0), z'(t_0)) = 0$.
(This can also be independently verified as $z'(t_0) = f_x(Q_0)x'(t_0) + f_y(Q_0)y'(t_0)$ by the chain rule.)

In particular, let $c_0 := f(x_0, y_0)$, and suppose the level curve $\{(x, y) \in D : f(x, y) = c_0\}$ is parametrized by $(x(t), y(t))$ with $t \in [\alpha, \beta]$. Then the contour line C is parametrized by $(x(t), y(t), c_0)$ with $t \in [\alpha, \beta]$, and it lies on the surface S. Hence the normal line to S at $P_0 := (x_0, y_0, c_0)$ is perpendicular to the corresponding contour line C.

Examples:

(i) Let $F(x,y,z) := x^2+y^2+z^2-1$ for $(x,y,z) \in \mathbb{R}^3$. Then $(\nabla F)(x,y,z) = (2x,2y,2z)$ and $(\nabla F)(0,0,1) = (0,0,2)$. The tangent plane to the surface defined by $F(x,y,z)=0$ at the point $(0,0,1)$ is $0(x-0)+0(y-0)+2(z-1)=0$, that is, $z=1$. The normal line at $(0,0,1)$ is $x=0, y=0, z=1+2t$, where $t \in \mathbb{R}$, that is, the z-axis. Note that the surface is the unit sphere in $\mathbb{R}^3$, and the point $(0,0,1)$ is its north pole.

(ii) Let $F(x,y,z) := e^x + \sin y - \cos z$ for $(x,y,z) \in \mathbb{R}^3$. Then $(\nabla F)(x,y,z) = (e^x, \cos y, \sin z)$ for $(x,y,z) \in \mathbb{R}^3$ and $(\nabla F)(0,0,0) = (1,1,0)$. The tangent plane to the surface defined by $F(x,y,z)=0$ at $(0,0,0)$ is $x+y+0z=0$, that is, $x+y=0$. The normal line at $(0,0,0)$ is $x=t, y=t, z=0$, where $t \in \mathbb{R}$, that is, the intersection of the planes given by $x=y$ and $z=0$.

(iii) Let $F(x, y, z) := x^2 + y^2 - 1$ for $(x, y, z) \in \mathbb{R}^3$. Then $(\nabla F)(x, y, z) = (2x, 2y, 0)$ for $(x, y, z) \in \mathbb{R}^3$. The tangent plane to the surface defined by $F(x, y, z) = 0$ at the point $P_0 := (\cos\theta_0, \sin\theta_0, z_0)$ on the surface is $2\cos\theta_0(x - \cos\theta_0) + 2\sin\theta_0(y - \sin\theta_0) + 0(z - z_0) = 0$, that is, $\cos\theta_0\ x + \sin\theta_0\ y = 1$. The normal line at P_0 is $x = \cos\theta_0(1 + 2t)$, $y = \sin\theta_0(1 + 2t)$, $z = z_0$, where $t \in \mathbb{R}$. Note that the surface is a cylinder in $\mathbb{R}^3$.

A surface in $\mathbb{R}^3$ can be parametrically given by $\big(x(u, v), y(u, v), z(u, v)\big)$ for $(u, v) \in E$, where $E \subset \mathbb{R}^2$, and $x,\ y,\ z : E \to \mathbb{R}$ are continuous functions. A point $P_0 := (x_0, y_0, z_0)$ on this surface is given by $\big(x(u_0, v_0), y(u_0, v_0), z(u_0, v_0)\big)$ for some $(u_0, v_0) \in E$. In a subsequent lecture, we shall define a normal vector to this parametrically given surface at P_0, and also obtain the equation of the tangent plane to the surface at P_0.

MA 105 : Calculus
Division 1, Lecture 16

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Necessary condition for differentiability
- Sufficient condition for differentiability
- Examples
- Basic properties of differentiable functions
- Chain Rules (in three avatars). Examples
- Geometric Interpretation of the Gradient
- Tangent plane and normal line
- Tangent vector to a parametrized curve

Tangent Plane

Let $D \subset \mathbb{R}^2$, (x_0, y_0) be an interior point of D, and let a function $f : D \to \mathbb{R}$ be differentiable at (x_0, y_0). Then the equation of the **tangent plane** to the surface $z = f(x, y)$ in $\mathbb{R}^3$ at $(x_0, y_0, f(x_0, y_0))$ is given by

$$z - f(x_0, y_0) = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

More generally, suppose $E \subset \mathbb{R}^3$ and $P_0 := (x_0, y_0, z_0)$ is an interior point of E. Let a function $F : E \to \mathbb{R}$ have partial derivatives at P_0, and let $(\nabla F)(P_0) \neq (0, 0, 0)$. Suppose F (implicitly) defines a surface S in $\mathbb{R}^3$ by the equation $F(x, y, z) = 0$ for $(x, y, z) \in E$, and P_0 lies on S. Then the equation of the **tangent plane** to S at P_0 is given by

$$F_x(P_0)(x - x_0) + F_y(P_0)(y - y_0) + F_z(P_0)(z - z_0) = 0.$$

Note: As before, if we let $F(x, y, z) := z - f(x, y)$, then we recover the earlier equation of the tangent plane.

Normal Line

Now $(\nabla F)(P_0) \cdot (x - x_0, y - y_0, z - z_0) = 0$ for all (x, y, z) on the tangent plane to the surface S, that is, $(\nabla F)(P_0)$ is perpendicular to the tangent plane. The line passing through $P_0 = (x_0, y_0, z_0)$ and parallel to the nonzero vector $(\nabla F)(P_0) = (F_x(P_0), F_y(P_0), F_z(P_0))$ is called the **normal line** to the surface defined by $F(x, y, z) = 0$ at P_0. The parametric equations of this normal line are

$$x = x_0 + F_x(P_0)t, \quad y = y_0 + F_y(P_0)t, \quad z = z_0 + F_z(P_0)t, \quad t \in \mathbb{R}.$$

If all $F_x(P_0), F_y(P_0), F_z(P_0)$ are nonzero, then the equations are

$$\frac{x - x_0}{F_x(P_0)} = \frac{y - y_0}{F_y(P_0)} = \frac{z - z_0}{F_z(P_0)}.$$

Also, the parametric equations of the normal line to the surface defined by $z - f(x, y) = 0$ at (x_0, y_0, z_0) are

$$x = x_0 - f_x(x_0, y_0)t, \; y = y_0 - f_y(x_0, y_0)t, \; z = f(x_0, y_0) + t, \; t \in \mathbb{R}.$$

Tangent Vector to a Curve

We now describe a crucial property of the normal line to a surface. Let C be a parametrized smooth curve in $\mathbb{R}^3$ given by

$$t \longmapsto (x(t), y(t), z(t)), \quad t \in [\alpha, \beta].$$

Let $t_0 \in (\alpha, \beta)$ be such that $(x'(t_0), y'(t_0), z'(t_0)) \neq (0, 0, 0)$. Then the vector $\big(x'(t_0), y'(t_0), z'(t_0)\big)$ is called the **tangent vector** to the curve C at the point $\big(x(t_0), y(t_0), z(t_0)\big)$ on it.

Now suppose that the curve C lies on a surface S defined by $F(x, y, z) = 0$, that is, $\phi(t) := F\left(x(t), y(t), z(t)\right) = 0$ for all $t \in [\alpha, \beta]$. By the Chain Rule,

$$0 = \phi'(t_0) = F_x(P_0)x'(t_0) + F_y(P_0)y'(t_0) + F_z(P_0)z'(t_0).$$

Thus the normal line to the surface S at P_0 is perpendicular to tangent vectors to all curves on S that pass through P_0.

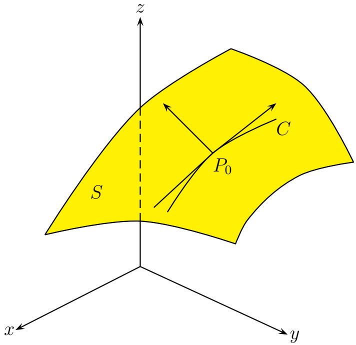


Figure: Normal line to the surface S at P_0 and a curve C on S

Special case: Suppose a surface S is given by $z = f(x, y)$ for $(x, y) \in D$. Let $Q_0 := (x_0, y_0) \in D$, and let a parametrized curve given by $(x(t), y(t))$, $t \in [\alpha, \beta]$, pass through Q_0. Let $z(t) := f(x(t), y(t))$, $t \in [\alpha, \beta]$. Then the curve C given by $(x(t), y(t), z(t))$, $t \in [\alpha, \beta]$, lies on S, and passes through $P_0 := f(Q_0)$.

Hence the normal line

$$x = x_0 - f_x(Q_0)t,\ y = y_0 - f_y(Q_0)t,\ z = f(Q_0) + t,\ t \in \mathbb{R},$$

to the surface S at P_0 is perpendicular to the tangent vector $(x'(t_0), y'(t_0), z'(t_0))$ to the curve C at P_0, that is, $(-f_x(Q_0), -f_y(Q_0), 1) \cdot (x'(t_0), y'(t_0), z'(t_0)) = 0$.
(This can also be independently verified as $z'(t_0) = f_x(Q_0)x'(t_0) + f_y(Q_0)y'(t_0)$ by the chain rule.)

In particular, let $c_0 := f(x_0, y_0)$, and suppose the level curve $\{(x, y) \in D : f(x, y) = c_0\}$ is parametrized by $(x(t), y(t))$ with $t \in [\alpha, \beta]$. Then the contour line C is parametrized by $(x(t), y(t), c_0)$ with $t \in [\alpha, \beta]$, and it lies on the surface S. Hence the normal line to S at $P_0 := (x_0, y_0, c_0)$ is perpendicular to the corresponding contour line C.

Closed Sets and Boundary Points

Let $D \subset \mathbb{R}^2$. We say that D is **closed** if every sequence in D that converges in $\mathbb{R}^2$ converges to a point of D itself.

A point $(x_0, y_0) \in \mathbb{R}^2$ is called a **boundary point** of D if there is a sequence in D that converges to (x_0, y_0) and also a sequence in the complement of D that converges to (x_0, y_0).

The set of all boundary points of D in $\mathbb{R}^2$ is called the **boundary** of D (in $\mathbb{R}^2$). We shall denote it by ∂D.

One can show that D is closed $\iff \partial D \subset D$ (Tutorial 9).

Examples: Let $D_1 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ and $D_2 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$. Then D_1 is closed, but D_2 is not. In fact, $\partial D_1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} = \partial D_2$.

Let $D := [a, b] \times [c, d]$. Then D is a closed subset of $\mathbb{R}^2$ and $\partial D = \big\{(x, y) \in D : x \in \{a, b\} \text{ or } y \in \{c, d\}\big\}$, which is the union of the 4 sides of D.

Finding Global Extrema

Let $D \subset \mathbb{R}^2$. We say that D is **bounded** if there is $\alpha > 0$ such that $\|(x, y)\| \leq \alpha$ for all $(x, y) \in D$.

Theorem (Extreme Value Theorem)

Let D be a nonempty closed and bounded subset of $\mathbb{R}^2$, and let $f : D \to \mathbb{R}$ be continuous. Then f is bounded and attains its bounds on D, that is, there are $(x_1, y_1), (x_2, y_2) \in D$ such that

$$\begin{aligned} f(x_1, y_1) &= \min\{f(x, y) : (x, y) \in D\}, \\ f(x_2, y_2) &= \max\{f(x, y) : (x, y) \in D\}. \end{aligned}$$

As in the case of the extreme value theorem for functions of one variable, the proof is quite involved, and we shall skip it.

Critical Points and Global Extrema

Let $D \subset \mathbb{R}^2$, and let $f : D \to \mathbb{R}$. An interior point (x_0, y_0) of D is called a **critical point** of f if either $(\nabla f)(x_0, y_0)$ does not exist, or $(\nabla f)(x_0, y_0)$ exists and is equal to $(0, 0)$.

Proposition

Let D be a nonempty, closed and bounded subset of $\mathbb{R}^2$, and $f : D \to \mathbb{R}$ be continuous. Then the global minimum m as well as the global maximum M of f is attained either at a critical point of f, or at a boundary point of D.

Proof. Let $(x_1, y_1) \in D$ be such that $f(x_1, y_1) = m$. If $(x_1, y_1) \in \partial D$, then we are done. Otherwise (x_1, y_1) is an interior point of D, and f has a local minimum at (x_1, y_1). If $(\nabla f)(x_1, y_1)$ does not exist, then (x_1, y_1) is a critical point of f. If $(\nabla f)(x_1, y_1)$ exists, then necessarily $(\nabla f)(x_1, y_1) = (0, 0)$, and so (x_1, y_1) is a critical point of f. Similarly, for M. $\square$

Finding Global Extrema

Let D be a nonempty closed and bounded subset of $\mathbb{R}^2$, and let $f : D \to \mathbb{R}$ be a continuous function. Here is a procedure for finding the global extrema of f on D.

(i) Determine the critical points of f in D. (Do not wonder whether f has a local extremum or a saddle point.)

(ii) Determine the boundary ∂D of D. Often ∂D consists of closed and bounded 'one dimensional pieces'.

(iii) Restrict the function f to each 'one dimensional piece' of the boundary of D. Use methods of one variable calculus to find the global extrema of f on that piece.

(iv) Compare the values of f at the critical points of f and at the extreme values of f on the one dimensional piece(s) of ∂D. The largest (resp., smallest) among these values is the global maximum (resp., minimum) of f on D.

We shall later give an alternative to (iii) above, called the Orthogonal Gradient Theorem.

Finding Global Extrema: Example

Example:

Let $D := \{(x, y) \in \mathbb{R}^2 : x,\ y \geq 0 \text{ and } x + y \leq 9\}$, and let $f(x, y) := 2 + 2x + 2y - x^2 - y^2$ for $(x, y) \in D$.

(i) (1,1) is the only critical point of f, and $f(1, 1) = 4$.

(ii) The boundary of D has three closed one dimensional pieces, one each on the lines $x = 0$, $y = 0$ and $y = 9 - x$.

(iii) $f(0, y) = 2 + 2y - y^2$ for $y \in [0, 9]$. At the critical point $y = 1$, $f(0, 1) = 3$. Also, $f(0, 0) = 2$ and $f(0, 9) = -61$.
$f(x, 0) = 2 + 2x - x^2$ for $x \in [0, 9]$. At the critical point $x = 1$, $f(1, 0) = 3$. Also, $f(0, 0) = 2$ and $f(9, 0) = -61$.
$f(x, 9 - x) = -2x^2 + 18x - 61$ for $x \in [0, 9]$. At the critical point $x = 9/2$, $f(9/2, 9/2) = -41/2$. Also, $f(0, 9) = -61 = f(9, 0)$.

Finding Global Extrema

(iv) The above values of f are tabulated as follows:

(x, y)	(1,1)	(0,1)	(0,0)	(0,9)	(1,0)	(9,0)	(9/2,9/2)
$f(x, y)$	4	3	2	-61	3	-61	-41/2

Thus, the global maximum of f on D is 4, and it is attained at $(1, 1)$, while the global minimum of f on D is -61, and it is attained at $(0, 9)$ as well as at $(9, 0)$.

Aliter: Clearly, $f(x, y) = 4 - (x - 1)^2 - (y - 1)^2 \leq 4$ for all $(x, y) \in \mathbb{R}^2$. Also, since $(9, 0)$ and $(0, 9)$ are the farthest points of D from the point $(1, 1)$, we see that $f(x, y) \geq 4 - (9 - 1)^2 - (0 - 1)^2 = -61$ for all $(x, y) \in D$.

In fact, $f(1, 1) = 4$ and $f(9, 0) = f(0, 9) = -61$.

If D is a closed and bounded subset of $\mathbb{R}^2$, and $f : D \to \mathbb{R}$ is continuous, then the values of f on the boundary ∂D of D play a special role when we find the global extrema of f on D.

Since the boundary of D often consists of one or more curves, we need to find the extreme values of f, when it is restricted to a curve in D. In this connection the following result is useful.

Proposition (Orthogonal Gradient Theorem)

Let $E \subset \mathbb{R}^2$ and (x_0, y_0) be an interior point of E. Let C be a parametrized curve in E passing through (x_0, y_0) given by $(x(t), y(t))$, where $x, y : [\alpha, \beta] \to E$, and let $t_0 \in (\alpha, \beta)$ with $x(t_0) = x_0$, $y(t_0) = y_0$. Suppose a function $f : E \to \mathbb{R}$, when restricted to C, has a local extremum at (x_0, y_0). If x and y are differentiable at t_0, and f is differentiable at (x_0, y_0), then $(\nabla f)(x_0, y_0) \cdot (x'(t_0), y'(t_0)) = 0$, that is, the gradient of f at (x_0, y_0) is perpendicular to the tangent vector of C at (x_0, y_0).

Proof. Define $\phi : [\alpha, \beta] \to \mathbb{R}$ by $\phi(t) := f(x(t), y(t))$. Then ϕ has a local extremum at t_0, and so $\phi'(t_0) = 0$. But by the chain rule (ii), $\phi'(t_0) = f_x(x_0, y_0)x'(t_0) + f_y(x_0, y_0)y'(t_0)$. $\square$

Example: Let $E := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 2\}$, and $f(x, y) := x\,y$ for $(x, y) \in E$. Let us find the global extrema of f on $D := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$.

- $(\nabla f)(x, y) = (y, x)$ if $x^2 + y^2 < 1$. Hence $(0, 0)$ is the only critical point of f, and $f(0, 0) = 0$.
- The boundary of D is given by $(x(t), y(t))$, where $x(t) := \cos t$, $y(t) := \sin t$ for $t \in [-\pi, \pi]$. Each point of D is an interior point of E.
- $(\nabla f)(x(t), y(t)) \cdot (x'(t), y'(t)) = -\sin^2 t + \cos^2 t = 0$ $\iff t = \pm\pi/4, \pm3\pi/4$, that is, $(x(t), y(t))$ $= (\pm1/\sqrt{2}, \pm1/\sqrt{2})$.
- $f(1/\sqrt{2}, 1/\sqrt{2}) = f(-1/\sqrt{2}, -1/\sqrt{2}) = 1/2$ and $f(-1/\sqrt{2}, 1/\sqrt{2}) = f(1/\sqrt{2}, -1/\sqrt{2}) = -1/2$.

So the maximum of f is $1/2$ and the minimum of f is $-1/2$.

Constrained Extrema

Let $D \subset \mathbb{R}^2$, and let $f, g : D \to \mathbb{R}$. Let us address the question of finding the maximum and the minimum of f on D subject to the constraint $g = 0$. The equation $g(x, y) = 0$, $(x, y) \in D$ often defines a curve C in D implicitly.

For example, let $D := \{(x, y) \in \mathbb{R}^2 : 1/4 \leq x^2 + y^2 \leq 4\}$, and let $g(x, y) := x^2 + y^2 - 1$ for $(x, y) \in D$. Then the curve C defined by $g(x, y) = 0$ is the unit circle in $\mathbb{R}^2$, and it lies in D.

Let (x_0, y_0) be an interior point of D such that $g(x_0, y_0) = 0$. Suppose f has a local extremum at (x_0, y_0) subject to the constraint $g(x, y) = 0$. We would like to show that the gradient vectors $(\nabla f)(x_0, y_0)$ and $(\nabla g)(x_0, y_0)$ are parallel.

First we give a general idea, and then we shall quote a precise result. Suppose f and g are differentiable at (x_0, y_0).

Assume that we are able to solve the equation $g(x, y) = 0$ for y in terms of x near x_0, that is, there a function η defined near x_0 such that $g(x, \eta(x)) = 0$ and $\eta(x_0) = y_0$. If $\eta'(x_0)$ exists, then $g_x(x_0, y_0) + g_y(x_0, y_0)\eta'(x_0) = 0$ by the chain rule (ii).

Consider the function $\phi(x) := f(x, \eta(x))$ for x near x_0. Now ϕ has a local extremum at x_0, and so $\phi'(x_0) = 0$, that is, $f_x(x_0, y_0) + f_y(x_0, y_0)\,\eta'(x_0) = 0$ again by the chain rule (ii).

It follows that $f_x(x_0, y_0)g_y(x_0, y_0) = f_y(x_0, y_0)g_x(x_0, y_0)$, that is, the gradient vectors $(\nabla f)(x_0, y_0) = (f_x(x_0, y_0), f_y(x_0, y_0))$ and $(\nabla g)(x_0, y_0) = (g_x(x_0, y_0), g_y(x_0, y_0))$ are parallel.

In fact, if $g_y(x_0, y_0) \neq 0$, then $(\nabla f)(x_0, y_0) = \lambda_0(\nabla g)(x_0, y_0)$, where $\lambda_0 := f_y(x_0, y_0)/g_y(x_0, y_0)$.

Similarly, if $g(x, y) = 0$ can be solved for x in terms of y near y_0 and if $g_x(x_0, y_0) \neq 0$, then $(\nabla f)(x_0, y_0) = \lambda_0(\nabla g)(x_0, y_0)$, where $\lambda_0 := f_x(x_0, y_0)/g_x(x_0, y_0)$.

Proposition (Lagrange Multiplier Theorem)

Let $D \subset \mathbb{R}^2$, and let (x_0, y_0) be an interior point of D. Suppose $f, g : D \to \mathbb{R}$ have continuous partial derivatives in a neighbourhood of (x_0, y_0). Let $C := \{(x, y) \in D : g(x, y) = 0\}$. Suppose (i) $g(x_0, y_0) = 0$, (ii) $(\nabla g)(x_0, y_0) \neq (0, 0)$, and (iii) the function f, when restricted to C, has a local extremum at (x_0, y_0). Then there is $\lambda_0 \in \mathbb{R}$ such that

$$(\nabla f)(x_0, y_0) = \lambda_0 (\nabla g)(x_0, y_0).$$

The real number λ_0 is called a **Lagrange multiplier**.

We now have the following procedure to determine the global extremum of a function $f : D \to \mathbb{R}$, subject to the constraint $g = 0$. Consider a new variable λ, called an **undetermined multiplier**, and seek simultaneous solutions of the following 3 equations in the 3 unknowns x, y, λ:
$(\nabla f)(x, y) = \lambda (\nabla g)(x, y)$ and $g(x, y) = 0$, $(x, y) \in D$.

Suppose f has a global extremum on the set $\{(x, y) \in D : g(x, y) = 0\}$. (For example, when this set is nonempty, closed and bounded, and f is continuous on it). Then it is also a local extremum of f, and so it is attained either at a simultaneous solution (x_0, y_0) of the above two equations where $\nabla g(x_0, y_0) \neq (0, 0)$, or at a point (x_1, y_1) where $g(x_1, y_1) = 0$ and $(\nabla g)(x_1, y_1) = (0, 0)$.

Examples: (i) Consider the problem of maximizing/minimizing $f(x, y) := x\,y$ subject to $x + y = 2$. For this purpose, we find the global extrema of $\phi(x) := f(x, 2 - x) = x(2 - x)$, $x \in \mathbb{R}$. Since $\phi'(x) = 2(1 - x)$, $x \in \mathbb{R}$, we see that $\phi'(x) > 0$ if $x < 1$ and $\phi'(x) < 0$ if $x > 1$. Also, ϕ is continuous at 1. As in the first derivative test, $\phi(1) = 1$ is the global maximum of ϕ, that is, 1 is the constrained maximum of f. Also, $\phi(x) \to -\infty$ as $x \to \pm\infty$, and so f has no constrained minimum.

(ii) Consider the problem of maximizing/minimizing
$f(x,y) := x\,y$, subject to $x^2+y^2-1=0$.

Let $g(x,y) := x^2+y^2-1$ for $(x,y) \in \mathbb{R}^2$. Note that the set $\{(x,y) \in \mathbb{R}^2 : g(x,y)=0\}$, that is, the unit circle, is nonempty, closed and bounded, and f is continuous on it. Now $(\nabla f)(x,y) = \lambda\,(\nabla g)(x,y)$ and $g(x,y)=0$ means

$$y = 2\lambda x, \quad x = 2\lambda y, \quad \text{and} \quad x^2+y^2-1=0.$$

Then $y\,x = 4\lambda^2 xy$, that is, $4\lambda^2 = 1$, since $x\,y \neq 0$. Thus $\lambda = \pm 1/2$, and the simultaneous solutions of the above equations are given by $(x,y) = \left(\pm 1/\sqrt{2},\ \pm 1/\sqrt{2}\right)$. Also, $(\nabla g)(x,y) \neq (0,0)$ whenever $g(x,y)=0$.

Thus the hypotheses of the Lagrange Multiplier Theorem are satisfied. Hence the maximum of f on the unit circle is $f\left(1/\sqrt{2},\ 1/\sqrt{2}\right) = f(-1/\sqrt{2},\ -1/\sqrt{2}) = 1/2$, while the minimum of f on the unit circle is
$f\left(1/\sqrt{2},\ -1/\sqrt{2}\right) = f\left(-1/\sqrt{2},\ 1/\sqrt{2}\right) = -1/2$.

Sequences in $\mathbb{R}^3$ and Functions of 3 Variables

A sequence in $\mathbb{R}^3$ is denoted by $\big((x_n, y_n, z_n)\big)$, where $x_n, y_n, z_n \in \mathbb{R}$ for $n \in \mathbb{N}$. We say that a sequence $\big((x_n, y_n, z_n)\big)$ in $\mathbb{R}^3$ is **convergent** in $\mathbb{R}^3$ if the sequences $(x_n), (y_n)$ and (z_n) are convergent in $\mathbb{R}$.

Let $D \subset \mathbb{R}^3$ and $f : D \to \mathbb{R}$. We say that f is continuous at $(x_0, y_0, z_0) \in D$ if $f(x_n, y_n, z_n) \to f(x_0, y_0, z_0)$ whenever $(x_n, y_n, z_n) \to (x_0, y_0, z_0)$ in D.

We have already considered the partial derivatives of f, and denoted the gradient of f at an interior point $P_0 := (x_0, y_0, z_0)$ of D by

$$(\nabla f)(P_0) := (f_x(P_0), f_y(P_0), f_z(P_0)).$$

Also, the differentiability of f, and the local and global extrema of f can be defined in a similar manner.

Here is an analogue of the Mean Value Theorem for real-valued functions of three variables.

Let $D \subset \mathbb{R}^3$. Suppose every point of D is an interior point and the line segment joining any two points in D lies in D. Let $(x_0, y_0, z_0), (x_1, y_1, z_1) \in D$, and let $f : D \to \mathbb{R}$.

(Trivariate Mean Value Theorem)

Suppose f has continuous partial derivatives on D. Then there is $(c, d, e) \in D$ lying on the open line segment joining (x_0, y_0, z_0) and (x_1, y_1, z_1) such that $f(x_1, y_1, z_1) - f(x_0, y_0, z_0)$ is equal to

$$(x_1 - x_0)f_x(c, d, e) + (y_1 - y_0)f_y(c, d, e) + (z_1 - z_0)f_z(c, d, e).$$

Proof: Define $\phi : [0, 1] \to \mathbb{R}$ by $\phi(t) := f(x_0 + t(x_1 - x_0), y_0 + t(y_1 - y_0), z_0 + t(z_1 - z_0))$, and use the one variable Lagrange mean value theorem for ϕ. $\square$

In particular, if $\nabla f = 0$ on D, then f is constant on D.

The following analogue of the Orthogonal Gradient Theorem would be useful.

Let $D \subset \mathbb{R}^3$, and let (x_0, y_0, z_0) be an interior point of D. Let C be a parametrized curve in D given by $(x(t), y(t), z(t))$, where $x, y, z : [\alpha, \beta] \to D$. Suppose there is $t_0 \in (\alpha, \beta)$ such that $x(t_0) = x_0$, $y(t_0) = y_0$, $z(t_0) = z_0$, and let a function $f : D \to \mathbb{R}$, when restricted to C, have a local extremum at (x_0, y_0, z_0). If x, y and z are differentiable at t_0, and if f is differentiable at (x_0, y_0, z_0), then

$$(\nabla f)(x_0, y_0, z_0) \cdot (x'(t_0), y'(t_0), z'(t_0)) = 0,$$

that is, the gradient of f at (x_0, y_0, z_0) is perpendicular to the tangent vector of C at the point (x_0, y_0, z_0).

As before, the result follows by the Chain Rule (part (ii)) applied to the function $\phi : [\alpha, \beta] \to \mathbb{R}^3$ given by

$$\phi(t) := f(x(t), y(t), z(t)).$$

Lagrange Multipliers for Functions of 3 Variables

The Lagrange Multiplier Theorem can be adapted to a subset D of $\mathbb{R}^3$ and functions $f, g : D \to \mathbb{R}$. If f and g have continuous partial derivatives, then a global extremum of f on the set $\{(x, y, z) \in D : g(x, y, z) = 0\}$ can be found by comparing the values of f at the simultaneous solutions of the following 4 equations in the 4 unknowns x, y, z, λ: $\nabla f = \lambda \nabla g$ and $g = 0$ at which $\nabla g \neq \mathbf{0}$. We must also consider the values of f at points where $g = 0$ and $\nabla g = \mathbf{0}$.

Further, the result can be adapted to a situation where there are two constraints, namely $g = 0$ and $h = 0$. In this case, we compare the values of f at the simultaneous solutions of the following 5 equations in the 5 unknowns x, y, z, λ, μ: $\nabla f = \lambda \nabla g + \mu \nabla h$ and $g = 0 = h$ at which $\nabla g \neq \mathbf{0}, \nabla h \neq \mathbf{0}$ and $\nabla g \nparallel \nabla h$. We must also consider the values of f at points where $g = 0 = h$, and $\nabla g = \mathbf{0}$ or $\nabla h = \mathbf{0}$ or $\nabla g \parallel \nabla h$.

Multiple constraints

Explanation: Let $D \subset \mathbb{R}^3$ and $f : D \to \mathbb{R}$. Let C denote the curve in D obtained by intersecting the surfaces S_g and S_h defined by $g = 0$ and $h = 0$ respectively. Suppose f, when restricted to C, has a local extremum at a point P_0 of C, and the tangent vector to the curve C at P_0 is nonzero. By the three-dimensional analogue of the orthogonal gradient theorem, this tangent vector is perpendicular to $(\nabla f)(P_0)$

Suppose $(\nabla g)(P_0) \neq \mathbf{0}$, $(\nabla h)(P_0) \neq \mathbf{0}$ and $\nabla g \nparallel \nabla h$. Now the normal lines to the surfaces S_g and S_h at P_0 are perpendicular to the tangent vector to the curve C at P_0, since C lies in S_g as well as in S_h. But the plane in $\mathbb{R}^3$ perpendicular to the tangent vector to the curve C at P_0 is two-dimensional, and so there are $\lambda_0, \mu_0 \in \mathbb{R}$ such that

$$(\nabla f)(P_0) = \lambda_0 (\nabla g)(P_0) + \mu_0 (\nabla h)(P_0).$$

Examples:

(i) Let us find the maximum and the minimum values of the function $f(x,y,z) = x + 2y + 3z$ on the curve of intersection of the plane $x - y + z = 1$ and the cylinder $x^2 + y^2 = 1$.

The continuous function f attains its bounds on the nonempty, closed and bounded subset $\{(x,y,z) : x - y + z = 1 = x^2 + y^2\}$.

The undetermined multipliers $\lambda,\ \mu \in \mathbb{R}$ satisfy

$$(1,2,3) = \lambda\,(1,-1,1) + \mu\,(2x,2y,0),$$

that is, $\lambda + 2\mu x = 1, \quad -\lambda + 2\mu y = 2$ and $\lambda = 3$. This gives

$$\mu x = -1, \quad \mu y = 5/2, \ \text{ and so } \ \mu^2 = (-1)^2 + (5/2)^2.$$

Hence $\mu = \pm\sqrt{29}/2$, $x = \mp 2/\sqrt{29}$ and $y = \pm 5/\sqrt{29}$, and so $z = 1 - x + y = 1 \pm 7/\sqrt{29}$. Also, $(1,-1,1) \nparallel (2x,2y,0)$.

It follows that the constrained maximum of f is $3 + \sqrt{29}$ and the constrained minimum of f is $3 - \sqrt{29}$.

(ii) Let us determine a point on the intersection of the planes $x + y + z = 1$ and $3x + 2y + z = 6$ that is closest to the origin.

Define $f(x, y, z) := x^2 + y^2 + z^2$ for $(x, y, z) \in \mathbb{R}^3$. We wish to find a point $(x, y, z) \in \mathbb{R}^3$ at which f attains its minimum subject to constraints $g(x, y, z) = x + y + z - 1 = 0$ and $h(x, y, z) = 3x + 2y + z - 6 = 0$.

The set $E := \{(x, y, z) \in \mathbb{R}^3 : g(x, y, z) = 0 = h(x, y, z)\}$ is nonempty since $(3, -1, -1) \in E$. It is closed since g and h are continuous, but it is not bounded. Let

$$E_1 := \{(x, y, z) \in E : \|(x, y, z)\| \leq \|(3, -1, -1)\|\}.$$

Then the set E_1 is nonempty, closed and bounded, and so the continuous function attains its minimum on E_1. Further, the minimum of f on E_1 is equal to the minimum of f on E. Thus f attains its minimum on E.

The equations $\nabla f = \lambda \nabla g + \mu \nabla h$ yield

$$x = \frac{\lambda + 3\mu}{2}, \quad y = \frac{\lambda + 2\mu}{2}, \quad z = \frac{\lambda + \mu}{2}.$$

Since $g(x, y, z) = 0 = h(x, y, z)$, we obtain $3\lambda + 6\mu = 2$ and $3\lambda + 7\mu = 6$. This gives $\lambda = -22/3$ and $\mu = 4$. Hence $P_0 := (x_0, y_0, z_0) = (7/3, 1/3, -5/3)$.

We note that $(\nabla g)(P_0) = (1, 1, 1)$ and $(\nabla h)(P_0) = (3, 2, 1)$ are nonzero and nonparallel. Hence f attains its constrained minimum on E at P_0.
We remark that f is not bounded above on the straight line E, and so it cannot have a constrained maximum on E.

ALITER:
$x + y + z = 1$ and $3x + 2y + z = 6 \implies 1 - x - y = z = 6 - 3x - 2y \implies y = 5 - 2x$ and $z = -4 + x$. Hence $x^2 + y^2 + z^2 = x^2 + (5 - 2x)^2 + (-4 + x)^2 = 6x^2 - 28x + 41.$

Therefore our problem reduces to minimizing the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) := 6x^2 - 28x + 41$, $x \in \mathbb{R}$.
Now $f'(x) = 4(3x - 7)$ for $x \in \mathbb{R}$. Clearly,
$x < 7/3 \implies f'(x) < 0 \implies f$ is decreasing on $(-\infty, 7/3)$,
$x > 7/3 \implies f'(x) > 0 \implies f$ is increasing on $(7/3, \infty)$.
Also, f is continuous at $7/3$. As in the First Derivative Test, f has a global minimum at $x = 7/3$.
Further, $x = 7/3 \implies y = 1/3$, $z = -5/3$.
Thus the point on the intersection of the planes that is closest to the origin is $(7/3, 1/3, -5/3)$.

'Lagrange' on the Eiffel Tower

'Lagrange' is among the 72 names inscribed on Eiffel Tower.

MA 105 : Calculus
Division 1, Lecture 17

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Quick review of the previous lecture
- Closed Sets and Boundary Points
- Global Extrema and Critical points
- Illustration of finding global extrema
- Constrained Extrema
- Method of Lagrange Multipliers
- Case of functions of 3 variables and multiple constraints
- Examples

Local Extremum

Let $D \subset \mathbb{R}^2$, and (x_0, y_0) be an interior point of D. We say that a function $f : D \to \mathbb{R}$ has

(i) a **local minimum** at (x_0, y_0) if there is $r > 0$ such that $B((x_0, y_0), r) \subset D$ and $f(x, y) \geq f(x_0, y_0)$ for all $(x, y) \in B((x_0, y_0), r)$.

(ii) a **local maximum** at (x_0, y_0) if there is $r > 0$ such that $B((x_0, y_0), r) \subset D$ and $f(x, y) \leq f(x_0, y_0)$ for all $(x, y) \in B((x_0, y_0), r)$.

(iii) a **local extremum** at (x_0, y_0) if it has a local maximum or a local minimum at (x_0, y_0).

Examples:
Let $f(x, y) := x^2 + y^2$ for $(x, y) \in \mathbb{R}^2$. Then f has a local minimum and $-f$ has a local maximum at $(0, 0)$. In fact, these are the global extrema of f and $-f$.

Proposition (Necessary condition for a local extremum)

Let $D \subset \mathbb{R}^2$, and let (x_0, y_0) be an interior point of D. Suppose $f : D \to \mathbb{R}$ has a local extremum at (x_0, y_0). If $\boldsymbol{u}$ is a unit vector and $(\boldsymbol{D_u} f)(x_0, y_0)$ exists, then $(\boldsymbol{D_u} f)(x_0, y_0) = 0$.

In particular, if $(\nabla f)(x_0, y_0)$ exists, then $(\nabla f)(x_0, y_0) = (0, 0)$.

Proof. Suppose $(\boldsymbol{D_u} f)(x_0, y_0)$ exists. Let E denote the set of all $(x_0 + tu_1, y_0 + tu_2)$ in D, where $t \in \mathbb{R}$. By the Carathéodory lemma for a directional derivative, there is $f_1 : E \to \mathbb{R}$ such that f_1 is continuous at (x_0, y_0), and

$$f(x_0 + tu_1, y_0 + tu_2) - f(x_0, y_0) = t\, f_1(x_0 + tu_1, y_0 + tu_2)$$

whenever $(x_0 + tu_1, y_0 + tu_2) \in D$. Suppose f has a local minimum at (x_0, y_0). Then there is $\delta > 0$ such that $f(x_0 + tu_1, y_0 + tu_2) \geq f(x_0, y_0)$ for all $t \in (-\delta, \delta)$.

Hence $f_1(x_0 + tu_1, y_0 + tu_2) \geq 0$ for all $t \in (0, \delta)$ and $f_1(x_0 + tu_1, y_0 + tu_2) \leq 0$ for all $t \in (-\delta, 0)$.

Thus $0 \leq \lim_{t \to 0^+} f_1(x_0 + tu_1, y_0 + tu_2) = f_1(x_0, y_0)$ as well as $f_1(x_0, y_0) = \lim_{t \to 0^-} f_1(x_0 + tu_1, y_0 + tu_2) \leq 0$ since f_1 is continuous at (x_0, y_0). Hence $(\boldsymbol{D_u} f)(x_0, y_0) = f_1(x_0, y_0) = 0$.

In particular, if $(\nabla f)(x_0, y_0)$ exists, then considering $\boldsymbol{u} := (1, 0)$ and $\boldsymbol{u} := (0, 1)$, we see that $(\nabla f)(x_0, y_0) = (0, 0)$.

A similar argument holds if f has a local maximum at (x_0, y_0).

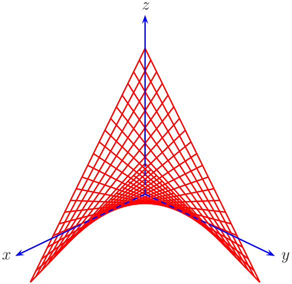
The converse of the above proposition does not hold.

Example:
Let $f(x, y) := x\,y$ for $(x, y) \in \mathbb{R}^2$. Then $(\boldsymbol{D_u} f)(0, 0) = 0$ for every unit vector $\boldsymbol{u}$, but f does not have a local extremum at $(0, 0)$ since $f(0, 0) = 0$, while $f(1/n, 1/n) = 1/n^2 > 0$ and $f(1/n, -1/n) = -1/n^2 < 0$ for all $n \in \mathbb{N}$.

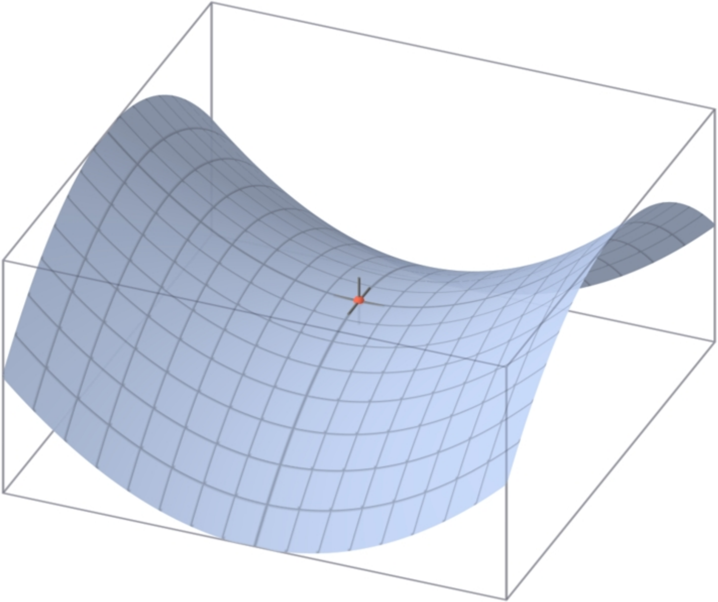
Saddle Point

Let (x_0, y_0) be an interior point of $D \subset \mathbb{R}^2$, and let $f : D \to \mathbb{R}$. Suppose the tangent plane to the surface $z = f(x, y)$ at (x_0, y_0) is the horizontal plane $z = f(x_0, y_0)$. (Thus if $(\nabla f)(x_0, y_0)$ exists, then $(\nabla f)(x_0, y_0) = (0, 0)$.) We say that f has a **saddle point** at (x_0, y_0) if f does not have a local extremum at (x_0, y_0).

Example: $(0, 0)$ is a saddle point of $f(x, y) := x\,y$, $(x, y) \in \mathbb{R}^2$. The surface given by $z = x\,y$ looks as follows.



Picture of a Saddle



Before considering sufficient conditions for a local extremum or for a saddle point, we consider the following useful results.

Let $D \subset \mathbb{R}^2$ be such that every point of D is an interior point, and the line segment joining any two points in D lies in D. Let $(x_0, y_0), (x_1, y_1) \in D$, and consider a function $f : D \to \mathbb{R}$.

(Bivariate Mean Value Theorem)

Suppose f has continuous partial derivatives on D. Then there is $(c, d) \in D$ lying on the open line segment joining (x_0, y_0) and (x_1, y_1) such that

$$f(x_1, y_1) = f(x_0, y_0) + (x_1 - x_0) f_x(c, d) + (y_1 - y_0) f_y(c, d).$$

Proof: Use the one variable MVT for $\phi : [0, 1] \to \mathbb{R}$ defined by $\phi(t) := f(x_0 + t(x_1 - x_0), y_0 + t(y_1 - y_0))$, noting that for $t \in (0, 1)$, $\phi'(t) = (x_1 - x_0) f_x((x_0 + t(x_1 - x_0), y_0 + t(y_1 - y_0)) + (y_1 - y_0) f_y((x_0 + t(x_1 - x_0), y_0 + t(y_1 - y_0))$ by the chain rule. □

To simplify the statements of extensions of the above result, let us introduce a partial differential operator

$$\mathcal{D}_{h,k} := h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}, \quad \text{where } h, k \in \mathbb{R}.$$

It can be applied to a function f which has partial derivatives. Further, by the Mixed Partials Theorem, we obtain

$$\mathcal{D}^2_{h,k}f = \mathcal{D}_{h,k}(\mathcal{D}_{h,k}f) = h^2 f_{xx} + 2hk\ f_{xy} + k^2 f_{yy}$$

if a function f has continuous partial derivatives of order ≤ 2.

(Bivariate Extended Mean Value Theorem)

Suppose f has continuous partial derivatives of order ≤ 2 on D. Let $h := x_1 - x_0$ and $k := y_1 - y_0$. Then there is $(c, d) \in D$ lying on the open line segment joining (x_0, y_0) and (x_1, y_1) with

$$f(x_1, y_1) = f(x_0, y_0) + (\mathcal{D}_{h,k}f)(x_0, y_0) + \frac{1}{2}(\mathcal{D}^2_{h,k}f)(c, d).$$

Thus $f(x_0, y_0) + (\mathcal{D}_{h,k} f)(x_0, y_0)$, or in other words,

$$f(x_0, y_0) + (x_1 - x_0) f_x(x_0, y_0) + (y_1 - y_0) f_y(x_0, y_0)$$

can be considered as an approximation of $f(x_1, y_1)$ with $(1/2)(\mathcal{D}^2_{h,k} f)(c, d)$ as the error.

More generally, we can prove the following result.

(Bivariate Taylor Theorem)

Let $n \in \mathbb{N}$. Suppose f has continuous partial derivatives of order $\leq (n+1)$ on D. Then there is $(c, d) \in D$ lying on the open line segment joining (x_0, y_0) and (x_1, y_1) such that

$$f(x_1, y_1) = \sum_{i=0}^{n} \frac{1}{i!} (\mathcal{D}^i_{h,k} f)(x_0, y_0) + R_n,$$

where $R_n := \dfrac{1}{(n+1)!} (\mathcal{D}^{n+1}_{h,k} f)(c, d)$, $h := x_1 - x_0$, $k := y_1 - y_0$.

The bivariate Taylor theorem can be proved by considering the function $\phi : [0,1] \to \mathbb{R}$ defined by

$$\phi(t) := f(x_0 + t(x_1 - x_0), y_0 + t(y_1 - y_0)),$$

and by using the one variable Taylor theorem for ϕ, as in the case of the bivariate mean value theorem. (Note: The proof involves repeated use of the Mixed Partials Theorem.)

The polynomial

$$P_n(x,y) \ := \ \sum_{i=0}^{n} \frac{1}{i!}(\mathcal{D}^i_{x-x_0,y-y_0} f)(x_0,y_0)$$

of degree $\leq n$ is an 'approximation' of $f(x,y)$ whenever $(x,y) \in D$. It is called the nth **bivariate Taylor polynomial** of f around (x_0,y_0), and

$$R_n(x,y) := \frac{1}{(n+1)!}(\mathcal{D}^{n+1}_{x-x_0,y-y_0} f)(c,d)$$

is called the nth **Lagrange remainder** of f around (x_0,y_0).

Using the Extended Bivariate Mean Value Theorem, we obtain

Proposition (Discriminant Test)

Let $D \subset \mathbb{R}^2$ and let (x_0, y_0) be an interior point of D. Suppose $f : D \to \mathbb{R}$ is such that the first-order and second-order partial derivatives of f exist and are continuous in a neighbourhood of (x_0, y_0), and $(\nabla f)(x_0, y_0) = (0, 0)$. Consider the **discriminant**

$$(\Delta f)(x_0, y_0) := f_{xx}(x_0, y_0) f_{yy}(x_0, y_0) - f_{xy}(x_0, y_0)^2$$

of f at (x_0, y_0).

(i) If $(\Delta f)(x_0, y_0) > 0$ and $f_{xx}(x_0, y_0) < 0$, then f has a local maximum at (x_0, y_0).

(ii) If $(\Delta f)(x_0, y_0) > 0$ and $f_{xx}(x_0, y_0) > 0$, then f has a local minimum at (x_0, y_0).

(iii) If $(\Delta f)(x_0, y_0) < 0$, then f has a saddle point at (x_0, y_0).

The discriminant test is inconclusive if $(\Delta f)(x_0, y_0) = 0$.

Examples: (i) Let $f(x,y) := x^2 + y^2$ for $(x,y) \in \mathbb{R}^2$. Then $(\nabla f)(x,y) = (2x, 2y) = (0,0) \iff (x,y) = (0,0)$. Also, $(\Delta f)(0,0) = 4$ and $f_{xx}(0,0) = 2 > 0$. Hence f has a local minimum at $(0,0)$.

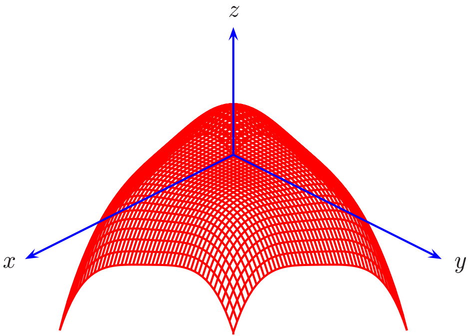
(ii) Let $f(x,y) := -x^2 - y^2$ for $(x,y) \in \mathbb{R}^2$. Then $(\nabla f)(x,y) = (-2x, -2y) = (0,0) \iff (x,y) = (0,0)$. Also, $(\Delta f)(0,0) = 4$ and $f_{xx}(0,0) = -2 < 0$. Hence f has a local maximum at $(0,0)$.

(iii) For $(x,y) \in \mathbb{R}^2$, let $f(x,y) := x\,y$. Then $(\nabla f)(x,y) = (y,x) = (0,0) \iff (x,y) = (0,0)$. Also, $(\Delta f)(0,0) = -1$. Hence f has a saddle point at $(0,0)$.

(iv) For $(x,y) \in \mathbb{R}^2$, let $f_1(x,y) := x^4 + y^4$, $f_2(x,y) := -x^4 - y^4$, and $f_3(x,y) := x^3y^3$.
Then $(\nabla f_i)(0,0) = (0,0)$ and $(\Delta f_i)(0,0) = 0$ for $i = 1,2,3$.
We observe that f_1 has a local minimum, f_2 has a local maximum and f_3 has a saddle point at $(0,0)$.

(v) Let $f(x, y) := 4xy - x^4 - y^4$ for $(x, y) \in \mathbb{R}^2$. Then $(\nabla f)(x, y) = (4(y - x^3), 4(x - y^3)) = (0, 0)$ $\iff (x, y) = (0, 0),\ (1, 1),$ or $(-1, -1)$.

Further, $(\Delta f)(x, y) = (-12x^2)(-12y^2) - 16 = 16(9x^2y^2 - 1)$ for $(x, y) \in \mathbb{R}^2$. In particular, $(\Delta f)(0, 0) = -16 < 0$ and $(\Delta f)(1, 1) = (\Delta f)(-1, -1) = 128 > 0$. Also, $f_{xx}(1, 1) = f_{xx}(-1, -1) = -12 < 0$. Hence f has a saddle point at $(0, 0)$ and local maxima at $(1, 1)$ and $(-1, -1)$.



(vi) Let $f(x, y) := x \sin y$ for $(x, y) \in \mathbb{R}^2$. Then $(\nabla f)(x, y) = (\sin y,\ x \cos y) = (0, 0) \iff (x, y) = (0, k\pi)$ for $k \in \mathbb{Z}$.

Further, $(\Delta f)(x, y) = 0(-x \sin y) - \cos^2 y = -\cos^2 y$ for $(x, y) \in \mathbb{R}^2$.

In particular, $(\Delta f)(0, k\pi) = -1 < 0$ for each $k \in \mathbb{Z}$.

Hence f has a saddle point at $(0, k\pi)$ for each $k \in \mathbb{Z}$.

Double Integral on a Rectangle

The concept of the Riemann integral of a function was motivated by our attempt to find 'the area under a curve'. We now look for a concept which will let us find 'the volume under a surface'. We shall assume that the **volume** of a **cuboid** $[a,b]\times[c,d]\times[p,q]$ is equal to $(b-a)(d-c)(q-p)$.

Let $R := [a,b]\times[c,d]$ be a rectangle in $\mathbb{R}^2$ with $a<b$, $c<d$. Let $f: R \to \mathbb{R}$ be a bounded function. Define

$$m(f) := \inf\{f(x,y) : (x,y) \in R\},\ M(f) := \sup\{f(x,y) : (x,y) \in R\}.$$

Let $n, k \in \mathbb{N}$, and consider a **partition** P of R given by $P := \{(x_i, y_j) : i = 0, 1, \ldots, n \text{ and } j = 0, 1, \ldots, k\}$, where $a := x_0 < x_1 < \cdots < x_n := b,\ \ c = y_0 < y_1 < \cdots < y_k = d$.

The points in P divide the rectangle R into nk nonoverlapping subrectangles $[x_{i-1}, x_i] \times [y_{j-1}, y_j]$, $i = 1, \ldots, n$; $j = 1, \ldots, k$.

Define

$$\begin{aligned} m_{i,j}(f) &:= \inf\{f(x,y) : (x,y) \in [x_{i-1}, x_i] \times [y_{j-1}, y_j]\}, \\ M_{i,j}(f) &:= \sup\{f(x,y) : (x,y) \in [x_{i-1}, x_i] \times [y_{j-1}, y_j]\}. \end{aligned}$$

Clearly, $m(f) \le m_{i,j}(f) \le M_{i,j}(f) \le M(f)$ for $i = 1, \ldots, n;\ j = 1, \ldots, k$.

Define the **lower double sum** and the **upper double sum** of f with respect to P by

$$\begin{aligned} L(P,f) &:= \sum_{i=1}^{n} \sum_{j=1}^{k} m_{i,j}(f)(x_i - x_{i-1})(y_j - y_{j-1}), \\ U(P,f) &:= \sum_{i=1}^{n} \sum_{j=1}^{k} M_{i,j}(f)(x_i - x_{i-1})(y_j - y_{j-1}). \end{aligned}$$

Since $\sum_{i=1}^{n}(x_i - x_{i-1}) = b - a$ and $\sum_{j=1}^{k}(y_j - y_{j-1}) = d - c$, we obtain

$$m(f)(b-a)(d-c) \le L(P,f) \le U(P,f) \le M(f)(b-a)(d-c).$$

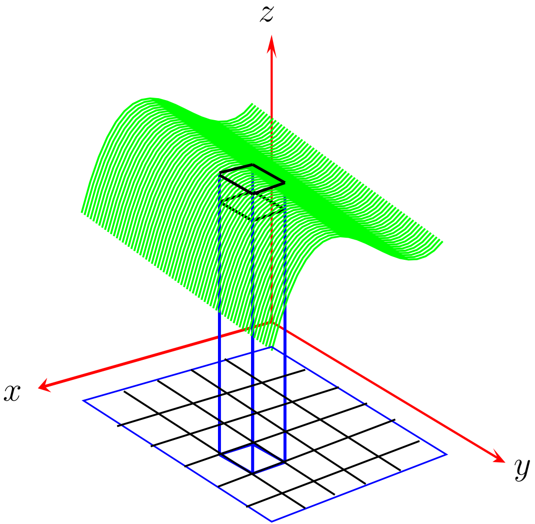


Figure: Summands of lower and upper double sums.

Let P_1 and P_2 be partitions of the rectangle R. A lower double sum increases and a upper double sum decreases when a partition is refined. Let P^* be a common refinement of P_1 and P_2. Then $L(P_1, f) \leq L(P^*, f) \leq U(P^*, f) \leq U(P_2, f)$. Define

$$\begin{aligned} L(f) &:= \sup\{L(P, f) : P \text{ is a partition of } R\}, \\ U(f) &:= \inf\{U(P, f) : P \text{ is a partition of } R\}. \end{aligned}$$

$L(f)$ is called the **lower double integral** of f and $U(f)$ is called the **upper double integral** of f.

Then $L(f) \leq U(f)$ by using the definitions of sup and inf.

Definition

A bounded function $f : R \to \mathbb{R}$ is said to be **(double) integrable** on R if $L(f) = U(f)$.

In this case, the **double integral** of f on R is the common value $U(f) = L(f)$, and it is denoted by

$$\iint_R f \quad \text{or} \quad \iint_R f(x,y)d(x,y).$$

Let $f : R \to \mathbb{R}$ be integrable and nonnegative. The double integral of f on R gives the volume of the solid $E_f := \{(x,y,z) \in \mathbb{R}^3 : (x,y) \in R \text{ and } 0 \le z \le f(x,y)\}$ under the surface $z = f(x,y)$ and above the rectangle R.

Examples:

(i) Let $f(x,y) := 1$ for all $(x,y) \in R$. Then $L(P,f) = (b-a)(d-c) = U(P,f)$ for every partition P of R. Hence f is integrable on R, and its double integral on R is equal to $(b-a)(d-c)$.

(ii) Define the **bivariate Dirichlet function** $f : R \to \mathbb{R}$ by

$$f(x,y) := \begin{cases} 1 & \text{if} \quad x \text{ and } y \text{ are rational numbers,} \\ 0 & \text{if} \quad x \text{ or } y \text{ is an irrational number.} \end{cases}$$

Then f is a bounded function on R. For each partition P of R,

$$m_{i,j}(f) = 0 \text{ and } M_{i,j}(f) = 1, \quad i = 1, \ldots, n;\ j = 1, \ldots, k,$$

and so $L(P, f) = 0$ and $U(P, f) = (b - a)(d - c)$. Thus $L(f) = 0$ and $U(f) = (b - a)(d - c)$. Since $L(f) \neq U(f)$, f is not integrable.

(iii) Let $\phi : [a, b] \to \mathbb{R}$ be bounded, and define $f : R \to \mathbb{R}$ by $f(x, y) := \phi(x)$ for $(x, y) \in R$. Then f is a bounded function. Let $P := \{(x_i, y_j) : i = 0, 1, \ldots, n \text{ and } j = 0, 1, \ldots, k\}$ be any partition of R. Then $P_1 := \{x_0, x_1, \ldots, x_n\}$ is a partition of $[a, b]$, and $m_{i,j}(f) = m_i(\phi)$, $i = 1, \ldots, n$; $j = 1, \ldots, k$, and so

$$\sum_{i=1}^{n} \sum_{j=1}^{n} m_{i,j}(f)(x_i - x_{i-1})(y_j - y_{j-1}) = (d - c) \sum_{i=1}^{n} m_i(\phi)(x_i - x_{i-1}).$$

Thus $L(P, f) = (d - c) L(P_1, \phi)$. Also, $U(P, f) = (d - c) U(P_1, \phi)$.

Further, if $Q := \{x_0, x_1, \ldots, x_n\}$ is a partition of $[a, b]$, then $Q = P_1$, where $P := \{(x_i, y_j) : i = 0, 1, \ldots, n,\ j = 0, 1\}$. So

$L(f) = (d - c)L(\phi)$ and $U(f) = (d - c)U(\phi)$. Hence

f is integrable on $R \iff \phi$ is integrable on $[a, b]$, and in this case, the double integral of f on R is equal to $(d - c)$ times the Riemann integral of ϕ on $[a, b]$.

The following result gives a necessary and sufficient condition for the integrability of a bounded function on R.

Theorem (Riemann condition)

Let $f : [a, b] \times [c, d] \to \mathbb{R}$ be a bounded function. Then f is integrable if and only if for every $\epsilon > 0$, there is a partition P_ϵ of $[a, b] \times [c, d]$ such that $U(P_\epsilon, f) - L(P_\epsilon, f) < \epsilon$.

The proof is very similar to the proof of the result about the Riemann condition in the one variable case.

Double Integration

The Riemann condition can be used to prove many useful results regarding double integration.

(Domain Additivity)

Let $R := [a, b] \times [c, d]$, and let $f : R \to \mathbb{R}$ be a bounded function. Let $s \in (a, b)$, $t \in (c, d)$. Then f is integrable on R if and only if f is integrable on the four subrectangles $[a, s] \times [c, t]$, $[a, s] \times [t, d]$, $[s, b] \times [c, t]$ and $[s, b] \times [t, d]$. In this case, the integral of f on R is the sum of the integrals of f on the four subrectangles.

We make the following conventions:

If $a = b$ or $c = d$, then $\iint_{[a,b]\times[c,d]} f := 0$.

$\iint_{[b,a]\times[c,d]} f := -\iint_{[a,b]\times[c,d]} f$, $\iint_{[a,b]\times[d,c]} f := -\iint_{[a,b]\times[c,d]} f$, $\iint_{[b,a]\times[d,c]} f := \iint_{[a,b]\times[c,d]} f$.

Integrable functions

Let $R := [a, b] \times [c, d]$, and let $f : R \to \mathbb{R}$.

(i) If f is monotonic in each of the two variables, then f is integrable on R.

(ii) If f is bounded on R, and has at most a finite number of discontinuities in R, then f is integrable on R.

Examples:

(i) Let $f(x, y) := [x] + [y]$ for $(x, y) \in R$. Since f is increasing in each variable, f is integrable.

(ii) Let $a, c > 0$ and $r, s \geq 0$. Define $f(x, y) = x^r y^s$ for $(x, y) \in R$. Since f is continuous on R, it is integrable.

(iii) Let $f(0, 0) := 0$ and $f(x, y) := xy/(x^2 + y^2)$ if $(x, y) \in [-1, 1] \times [-1, 1]$ and $(x, y) \neq (0, 0)$. Since f is bounded on R, and it is discontinuous only at $(0, 0)$, f is integrable.

Algebraic and Order Properties

Let $R := [a, b]\times[c, d]$. If $f, g : R \to \mathbb{R}$ are integrable, then

(i) $f + g$ is integrable, and $\iint_R (f + g) = \iint_R f + \iint_R g$,

(ii) αf is integrable, and $\iint_R \alpha f = \alpha \iint_R f$ for all $\alpha \in \mathbb{R}$,

(iii) $f \cdot g$ is integrable,

(iv) If there is $\delta > 0$ such that $|f(x, y)| \geq \delta$ for all $(x, y) \in R$ (so that $1/f$ is bounded), then $1/f$ is integrable,

(v) If $f \leq g$, then $\iint_R f \leq \iint_R g$,

(vi) $|f|$ is integrable, and $|\iint_R f| \leq \iint_R |f|$.

Proof: (v) For any partition P of R, $U(P, f) \leq U(P, g)$, and so $\iint_R f = U(f) \leq U(g) = \iint_R g$.

(vi) $U(P, |f|) - L(P, |f|) \leq U(P, f) - L(P, f)$ for each P.
Also, $-|f| \leq f \leq |f| \implies -\iint_R |f| \leq \iint_R f \leq \iint_R |f|$.

Evaluation of a Double Integral

Suppose a function is integrable on a rectangle R. How can we find its double integral?

To evaluate the Riemann integral of an integrable function f on an interval $[a, b]$, we used a powerful result known as the Fundamental Theorem of Calculus (Part II). If we find a function F on $[a, b]$ whose derivative is equal to f on $[a, b]$, then

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

Later in this course, we shall see some versions of the Fundamental Theorem of Calculus for functions of two/three variables, known as the Green/Stokes theorem.

For the present, we consider an easy and widely used method for evaluating double integrals in which the problem is reduced to a repeated evaluation of Riemann integrals.

MA 105 : Calculus
Division 1, Lecture 18

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Local extremum
- Necessary condition for a local extremum
- Saddle point. Examples
- Bivariate Mean Value Theorem
- Extended Bivariate Mean Value Theorem
- Bivariate Taylor's theorem
- Discriminant Test for functions of two variables
- Examples
- Double integrals on a rectangle
- Riemann condition and domain additivity
- Integrable functions
- Algebraic and order properties of double integrals

Double Integral on a Rectangle

Recall that if $R := [a, b] \times [c, d]$ is a rectangle in $\mathbb{R}^2$ with $a < b$, $c < d$. and $f : R \to \mathbb{R}$ be a bounded function, then for a partition P of R, we defined the **lower double sum** $L(P, f)$ and the **upper double sum** $U(P, f)$ of f with respect to P. Moreover, we defined the **lower double integral** $L(f)$ of f and the **upper double integral** $U(f)$ of f by

$$\begin{aligned} L(f) &:= \sup\{L(P, f) : P \text{ is a partition of } R\}, \\ U(f) &:= \inf\{U(P, f) : P \text{ is a partition of } R\}. \end{aligned}$$

Definition

A bounded function $f : R \to \mathbb{R}$ is said to be **(double) integrable** on R if $L(f) = U(f)$.

We have also seen examples of bounded functions that are integrable and also those that are not integrable.

Riemann condition and Domain additivity

Basic results about integrability over a rectangle are:

Theorem (Riemann condition)

Let $f : [a, b] \times [c, d] \to \mathbb{R}$ be a bounded function. Then f is integrable if and only if for every $\epsilon > 0$, there is a partition P_ϵ of $[a, b] \times [c, d]$ such that $U(P_\epsilon, f) - L(P_\epsilon, f) < \epsilon$.

(Domain Additivity)

Let $R := [a, b]\times[c, d]$, and let $f : R \to \mathbb{R}$ be a bounded function. Let $s \in (a, b)$, $t \in (c, d)$. Then f is integrable on R if and only if f is integrable on the four subrectangles $[a, s]\times[c, t]$, $[a, s]\times[t, d]$, $[s, b]\times[c, t]$ and $[s, b]\times[t, d]$.
In this case, the integral of f on R is the sum of the integrals of f on the four subrectangles.

Let $R := [a, b] \times [c, d]$, and let $f : R \to \mathbb{R}$.

(i) If f is monotonic in each of the two variables, then f is integrable on R.

(ii) If f is bounded on R, and has at most a finite number of discontinuities in R, then f is integrable on R.

Let $R := [a, b] \times [c, d]$. If $f, g : R \to \mathbb{R}$ are integrable, then

(i) $f + g$ is integrable, and $\iint_R (f + g) = \iint_R f + \iint_R g$,

(ii) αf is integrable, and $\iint_R \alpha f = \alpha \iint_R f$ for all $\alpha \in \mathbb{R}$,

(iii) $f \cdot g$ is integrable,

(iv) If there is $\delta > 0$ such that $|f(x, y)| \geq \delta$ for all $(x, y) \in R$ (so that $1/f$ is bounded), then $1/f$ is integrable,

(v) If $f \leq g$, then $\iint_R f \leq \iint_R g$,

(vi) $|f|$ is integrable, and $|\iint_R f| \leq \iint_R |f|$.

Evaluation of a Double Integral

Suppose a function is integrable on a rectangle R. How can we find its double integral?

To evaluate the Riemann integral of an integrable function f on an interval $[a, b]$, we used a powerful result known as the Fundamental Theorem of Calculus (Part II). If we find a function F on $[a, b]$ whose derivative is equal to f on $[a, b]$, then

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

Later in this course, we shall see some versions of the Fundamental Theorem of Calculus for functions of two/three variables, known as the Green/Stokes theorem.

For the present, we consider an easy and widely used method for evaluating double integrals in which the problem is reduced to a repeated evaluation of Riemann integrals.

Theorem (Fubini Theorem on a Rectangle)

Let $R := [a,b]\times[c,d]$, and let $f : R \to \mathbb{R}$ be integrable. Let I denote the double integral of f on R.

(i) If for each fixed $x \in [a,b]$, the Riemann integral $\int_c^d f(x,y)dy$ exists, then the **iterated integral** $\int_a^b \left(\int_c^d f(x,y)dy\right)dx$ exists and is equal to I.

(ii) If for each fixed $y \in [c,d]$, the Riemann integral $\int_a^b f(x,y)dx$ exists, then the **iterated integral** $\int_c^d \left(\int_a^b f(x,y)dx\right)dy$ exists and is equal to I.

(iii) If the hypotheses in both (i) and (ii) above hold, and in particular, if f is continuous on R, then

$$\int_a^b\left(\int_c^d f(x,y)dy\right)dx = I = \int_c^d\left(\int_a^b f(x,y)dx\right)dy.$$

The Fubini theorem can be proved by using the Riemann condition for a double integral and for a Riemann integral.

Special case: Let $\phi : [a, b] \to \mathbb{R}$ and $\psi : [c, d] \to \mathbb{R}$ be Riemann integrable. Define $f : R \to \mathbb{R}$ by $f(x, y) := \phi(x)\psi(y)$, $(x, y) \in R$. Then f is integrable on R, and its double integral is equal to

$$\int_a^b \left(\int_c^d \phi(x)\psi(y)dy \right) dx = \left(\int_a^b \phi(x)dx \right) \left(\int_c^d \psi(y)dy \right).$$

In particular, if $r, s \in \mathbb{R}$ with $r \geq 0$ and $s \geq 0$, then

$$\iint_{[a,b]\times[c,d]} x^r y^s d(x, y) = \left(\frac{b^{r+1} - a^{r+1}}{r + 1} \right) \left(\frac{d^{s+1} - c^{s+1}}{s + 1} \right),$$

provided $0 < a < b$ and $0 < c < d$.

Geometrically, the Fubini theorem says that if f is a nonnegative integrable function on $R := [a, b]\times[c, d]$, then the volume of the solid D under the surface $z = f(x, y)$ and above the rectangle R can be found by the slice method. For $x \in [a, b]$, we find the area $A(x) := \int_c^d f(x, y)dy$, of the cross-section of D perpendicular to the x-axis, or alternatively, for $y \in [c, d]$, we find the area $B(y) := \int_a^b f(x, y)dx$, of the cross-section of D perpendicular to the y-axis. (See Figure.)

$$\text{Vol}\,(D) = \int_a^b A(x)dx = \int_c^d B(y)dy.$$

Examples:

(i) Let $R := [0, 1]\times[0, 1]$, and $f(x, y) := (x + y)^2$, $(x, y) \in R$. Then f is continuous on R. The double integral of f on R is

$$\begin{aligned} \int_0^1\left(\int_0^1 (x+y)^2 dx\right) dy &= \frac{1}{3}\int_0^1 (x+y)^3\Big|_0^1 dy \\ &= \frac{1}{3}\int_0^1 \left((1+y)^3 - y^3\right) dy = \frac{7}{6}. \end{aligned}$$

(ii) Let $R := [0,1]\times[0,1]$, $f(0,0) := 0$, and for $(x,y) \neq (0,0)$, let $f(x,y) := xy(x^2-y^2)/(x^2+y^2)^3$. For $x \in [0,1]$, $x \neq 0$,

$$A(x) := \int_0^1 f(x,y)dy = \int_0^1 \frac{xy(x^2-y^2)}{(x^2+y^2)^3}dy = \frac{x}{2(1+x^2)^2}.$$

(Substitute $x^2+y^2 = u$.) Also, $A(0) = \int_0^1 0dy = 0$. Hence

$$\int_0^1 \left(\int_0^1 f(x,y)dy \right) dx = \int_0^1 A(x)dx = \int_0^1 \frac{x}{2(1+x^2)^2}dx = \frac{1}{8}.$$

By interchanging x and y, $\displaystyle\int_0^1 \left(\int_0^1 f(x,y)dx \right) dy = -\frac{1}{8}$.

Thus the two iterated integrals exist, but they are not equal.

Note that since $f(1/n, 1/2n) = 24n^2/125$ for all $n \in \mathbb{N}$, the function f is not bounded on R, and so it is not integrable on R. Thus Fubini's theorem is not applicable.

Remark

Around 1910, Fichtenholz and Lichtenstein proved the following result.

Theorem

Let $f : [a,b]\times[c,d] \to \mathbb{R}$ be a bounded function. If for each fixed $x \in [a,b]$, the Riemann integral $\int_c^d f(x,y)dy$ exists, and if for each fixed $y \in [c,d]$, the Riemann integral $\int_a^b f(x,y)dx$ exists, then the **iterated integrals** $\int_a^b \left(\int_c^d f(x,y)dy\right)dx$ and $\int_c^d \left(\int_a^b f(x,y)dx\right)dy$ exist and are equal.

The above result is stronger than the Fubini theorem since it does not assume that the function $f : [a,b] \times [c,d] \to \mathbb{R}$ is integrable; it only assumes that f is bounded on $[a,b]\times[c,d]$.

Double Riemann Sum

If we are not able to evaluate the double integral exactly, we attempt to find its approximations.

Given a bounded function $f : R \to \mathbb{R}$, and a partition $P := \{(x_i, y_j) : i = 0, 1, \ldots, n \text{ and } j = 0, 1, \ldots, k\}$, of $R := [a, b] \times [c, d]$, a double sum of the form

$$S(P, f) := \sum_{i=1}^{n} \sum_{j=1}^{k} f(s_i, t_j)(x_i - x_{i-1})(y_j - y_{j-1}),$$

where $(s_i, t_j) \in [x_{i-1}, x_i] \times [y_{j-1}, y_j]$ for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, is called a double Riemann sum for f corresponding to P.

Note: $L(P, f) \le S(P, f) \le U(P, f)$ for any $s_1, \ldots, s_n \in [a, b]$ and $t_1, \ldots, t_k \in [c, d]$.

We define the **mesh** of a partition $P := \{(x_i, y_j) : i = 0, 1, \ldots, n \text{ and } j = 0, 1, \ldots, k\}$ by $\mu(P) := \max\{x_1 - x_0, \ldots, x_n - x_{n-1},\ y_1 - y_0, \ldots, y_k - y_{k-1}\}$.

Theorem: Let f be integrable on R, and let $\epsilon > 0$. Then there is $\delta > 0$ such that $U(P, f) - L(P, f) < \epsilon$ for every partition P satisfying $\mu(P) < \delta$.

Corollary: If f is integrable on R and (P_n) is a sequence of partitions of R with $\mu(P_n) \to 0$, then $U(P_n, f) - L(P_n, f) \to 0$. Further, if $S(P_n, f)$ is a double Riemann sum corresponding to P_n and f, then $S(P_n, f) \to \iint_R f$.

Proof: Given $\epsilon > 0$, find $\delta > 0$ as in the theorem. Then there is $n_0 \in \mathbb{N}$ such that $\mu(P_n) < \delta$ for all $n \geq n_0$, and so $U(P_n, f) - L(P_n, f) < \epsilon$. Thus $U(P_n, f) - L(P_n, f) \to 0$. Since $L(P_n, f) \leq S(P_n, f) \leq U(P_n, f)$, and $L(P_n, f) \leq L(f) = \iint_R f = U(f) \leq U(P_n, f)$, we see that $\left|S(P_n, f) - \iint_R f\right| \leq U(P_n, f) - L(P_n, f) \to 0$. $\square$

Caution:
If we define

$$\mu(P) := \max\{(x_i - x_{i-1})(y_j - y_{j-1}) : 1 \le i \le n;\ 1 \le j \le k\},$$

then we may not obtain $S(P_n, f) \to \iint_R f$ as $\mu(P_n) \to 0$ for every integrable function f on R. An example of a function of this kind is a bit involved.

The preceding corollary allows us to find limits of certain sequences. For example, let

$$s_n := \frac{1}{n^4} \sum_{i=1}^{n} \sum_{j=1}^{n} (i + j)^2 \quad \text{for } n \in \mathbb{N}.$$

Then $s_n = \dfrac{1}{n^2} \displaystyle\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{i}{n} + \frac{j}{n}\right)^2 = S(\mathrm{P}_n, f)$, where $f(x, y) := (x + y)^2$ for $(x, y) \in [0, 1] \times [0, 1]$, and $\mu(\mathrm{P}_n) = 1/n$ for $n \in \mathbb{N}$. Hence $s_n \to \int_0^1 \int_0^1 (x + y)^2 d(x, y) = 7/6$.

Double Integral over a Bounded Set

Let D be a bounded subset of $\mathbb{R}^2$, and let $f : D \to \mathbb{R}$ be a bounded function. Consider a rectangle $R := [a, b] \times [c, d]$ such that $D \subset R$, and define $f^* : R \to \mathbb{R}$ by

$$f^*(x, y) := \begin{cases} f(x, y) & \text{if } (x, y) \in D, \\ 0 & \text{otherwise.} \end{cases}$$

We say that f is **integrable over** D if f^* is integrable on R, and in this case, the **double integral of** f **over** D is defined to be the double integral of f^* on R, that is,

$$\iint_D f(x, y) d(x, y) := \iint_R f^*(x, y) d(x, y).$$

By the domain additivity of double integrals on rectangles, the integrability of f over D and the value of its double integral are independent of the choice of a rectangle R containing D and the corresponding extension f^* of f to R.

Let D be a bounded subset of $\mathbb{R}^2$, and let $f : D \to \mathbb{R}$ be integrable over D. We may also denote the double integral of f over D by $\iint_D f$.

If f is nonnegative, then the **volume** of the solid under the surface given by $z = f(x, y)$ and above the region D is defined to be the double integral of f over D. Thus

$$\mathrm{Vol}\,(E_f) := \iint_D f(x, y)d(x, y),$$

where $E_f := \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in D \text{ and } 0 \le z \le f(x, y)\}$.

Next, if $g : D \to \mathbb{R}$ is integrable, and $f \le g$ on D, then

$$\iint_D \big(g(x, y) - f(x, y)\big)\, d(x, y)$$

is defined to be the **volume between the surfaces** given by $z = f(x, y)$ and $z = g(x, y)$.

The double integral over a bounded subset of $\mathbb{R}^2$ has algebraic and order properties analogous to those of the double integral on a rectangle.

In order to seek conditions under which a bounded function f defined on a bounded subset D of $\mathbb{R}^2$ is integrable over D, we introduce a new concept.

A bounded subset E of $\mathbb{R}^2$ is said to be of (two-dimensional) **content zero** if for every $\epsilon > 0$, there are finitely many rectangles whose union contains E and the sum of whose areas is less than ϵ.

A set of (two-dimensional) content zero is a 'thin' subset of $\mathbb{R}^2$.

Examples:

(i) Every finite subset of $\mathbb{R}^2$ is of content zero.

(ii) The infinite subset $\{(1/n, 1/k) : n, k \in \mathbb{N}\}$ of $\mathbb{R}^2$ is of content zero.

(iii) The subset $\{(x, y) \in [0,1]\times[0,1] : x, y \in \mathbb{Q}\}$ of $\mathbb{R}^2$ is not of content zero. (Consider any ϵ less than 1.)

(iv) Let $\varphi : [a, b] \to \mathbb{R}$ be an integrable function. Then its graph $E := \{(x, \varphi(x)) : x \in [a, b]\}$ is of content zero. To see this, let $\epsilon > 0$. By the Riemann condition, there is a partition $P := \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$ such that $U(P, \varphi) - L(P, \varphi) < \epsilon$. Then $E \subset \bigcup_{i=1}^{n} R_i$, where $R_i := [x_{i-1}, x_i]\times[m_i(\varphi), M_i(\varphi)]$ and

$$\text{Area}\,(R_1) + \cdots + \text{Area}\,(R_n) = U(P, \varphi) - L(P, \varphi) < \epsilon.$$

Similarly, if $\psi : [c, d] \to \mathbb{R}$ is an integrable function, then the set $E := \{(\psi(y), y) : y \in [c, d]\}$ is of content zero.

(v) Let functions x, $y : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b), and let their derivatives be bounded on (a, b). Then the set $E := \{(x(t), y(t)) \in \mathbb{R}^2 : t \in [a, b]\}$ is of two-dimensional content zero. We prove this as follows.

For $n \in \mathbb{N}$, consider the partition of $[a, b]$ into n equal parts, and let s_i be the mid-point of the ith subinterval for $i = 1, \ldots, n$. Let $t \in [a, b]$. Then there is $i \in \{1, \ldots, n\}$ such that $|t - s_i| \leq (b - a)/2n$. By the Mean Value Theorem, there are c_x and c_y between t and s_i such that

$$x(t)-x(s_i) = x'(c_x)(t-s_i) \quad \text{and} \quad y(t)-y(s_i) = x'(c_y)(t-s_i).$$

Let $\alpha \in \mathbb{R}$ with $|x'|, |y'| \leq \alpha$ on (a, b), and $\beta := b - a$. Then $|x(t) - x(s_i)|, |y(t) - y(s_i)| \leq \alpha\beta/2n$, and so $(x(t), y(t))$ lies in a square S_i centered at $(x(s_i), y(s_i))$ and of area $\alpha^2\beta^2/n^2$.

Thus E is contained in the union of the n squares $S_1, \ldots, S_n$, the sum of whose areas is equal to $\alpha^2\beta^2/n$.

Given $\epsilon > 0$, choose $n \in \mathbb{N}$ such that $n > \alpha^2(b - a)^2/\epsilon$. Then the set E is contained in the union of these n squares, the sum of whose areas is less that ϵ. Hence E is of two-dimensional content zero.

Theorem

Let D be a bounded subset of $\mathbb{R}^2$, and $f : D \to \mathbb{R}$ be a bounded function. If the boundary ∂D of D is of (two-dimensional) content zero and if the set of discontinuities of f in D is also of (two-dimensional) content zero, then f is integrable over D.

We omit the proof.

The above theorem says that the two conditions
(i) 'the boundary ∂D of D is of content zero' and
(ii) 'the set of discontinuities of f in D is of content zero'
together are sufficient to guarantee the integrability of f on D. While neither condition can be dropped from the above theorem, the integrability of f does not necessarily imply either condition.

(Find examples to illustrate the above remarks.)

Elementary Regions

Let $\phi_1, \phi_2 : [a, b] \to \mathbb{R}$ be continuous functions on $[a, b]$ such that $\phi_1 \le \phi_2$, and let

$$D := \{(x, y) \in \mathbb{R}^2 : a \le x \le b \text{ and } \phi_1(x) \le y \le \phi_2(x)\}.$$

Alternatively, let $\psi_1, \psi_2 : [c, d] \to \mathbb{R}$ be continuous functions on $[c, d]$ such that $\psi_1 \le \psi_2$, and let

$$D := \{(x, y) \in \mathbb{R}^2 : c \le y \le d \text{ and } \psi_1(y) \le x \le \psi_2(y)\}.$$

In both cases, D is called an **elementary region** in $\mathbb{R}^2$. In the former case, it is said to be of **type** I, and in the latter case, it is said to be of **type** II.

We note that the boundary of an elementary domain in $\mathbb{R}^2$ is of (two-dimensional) content zero. Hence a continuous function defined on such a domain is integrable over it.

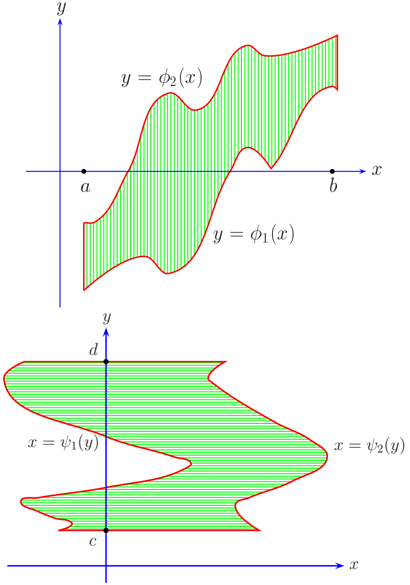


Figure: Elementary regions with boundaries in red colour

Clearly, a rectangle is an elementary region in $\mathbb{R}^2$ of type I as well as of type II.
Also, if $a > 0$, then the disk $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq a^2\}$ is an elementary region in $\mathbb{R}^2$, since

$$D = \left\{(x, y) \in \mathbb{R}^2 : -a \leq x \leq a \text{ and } -\sqrt{a^2 - x^2} \leq y \leq \sqrt{a^2 - x^2}\right\}.$$

Essential feature of an elementary region D in $\mathbb{R}^2$ of type I:
For every $x \in \mathbb{R}$, the vertical cross-section of D at x is a closed and bounded interval.
Essential feature of an elementary region D in $\mathbb{R}^2$ of type II:
For every $y \in \mathbb{R}$, the horizontal cross-section of D at y is a closed and bounded interval.

There do exist bounded subsets of $\mathbb{R}^2$ that are not elementary regions. For example, a star-shaped subset of $\mathbb{R}^2$ or an annulus is not an elementary region, since for some $x \in \mathbb{R}$, the vertical cross-section of D at x is not an interval, and for some $y \in \mathbb{R}$, the horizontal cross-section of D at y is not an interval.

Proposition (Fubini Theorem over Elementary Regions)

Let D be a subset of $\mathbb{R}^2$, and let $f : D \to \mathbb{R}$ be continuous.

(i) If $D := \{(x, y) \in \mathbb{R}^2 : a \le x \le b \text{ and } \phi_1(x) \le y \le \phi_2(x)\}$, where $\phi_1, \phi_2 : [a, b] \to \mathbb{R}$ are continuous, then the iterated integral $\int_a^b \big(\int_{\phi_1(x)}^{\phi_2(x)} f(x, y)dy \big) dx$ exists and equals the double integral $\iint_D f(x, y)d(x, y)$.

(ii) If $D := \{(x, y) \in \mathbb{R}^2 : c \le y \le d \text{ and } \psi_1(y) \le x \le \psi_2(y)\}$, where $\psi_1, \psi_2 : [c, d] \to \mathbb{R}$ are continuous, then the iterated integral $\int_c^d \big(\int_{\psi_1(y)}^{\psi_2(y)} f(x, y)dx \big) dy$ exists and equals the double integral $\iint_D f(x, y)d(x, y)$.

Proof: (i) Let $c := \inf\{\phi_1(x) : x \in [a, b]\}$ and $d := \sup\{\phi_2(x) : x \in [a, b]\}$. Then $D \subset R := [a, b] \times [c, d]$, and the extended function $f^* : R \to \mathbb{R}$ is integrable. Now use the Fubini theorem for f^*. (ii) A similar argument works here.

Examples:

(i) Let $D := \{(x, y) \in \mathbb{R}^2 : 0 \le x \le 1/2 \text{ and } 0 \le y \le x^2\}$ and $f(x, y) := x + y$ for $(x, y) \in D$. Then f is continuous on the elementary region D. By the Fubini theorem,

$$I := \iint_D (x+y)d(x,y) = \int_0^{1/2}\left(\int_0^{x^2}(x+y)dy\right)dx,$$

which is equal to

$$\int_0^{1/2}\left[xy + \frac{y^2}{2}\right]_{y=0}^{y=x^2} dx \;=\; \int_0^{1/2}\left(x^3 + \frac{x^4}{2}\right)dx \;=\; \frac{3}{160}.$$

Alternatively, since
$D = \{(x, y) \in \mathbb{R}^2 : 0 \le y \le 1/4 \text{ and } \sqrt{y} \le x \le 1/2\}$,

$$I = \int_0^{1/4}\left(\int_{\sqrt{y}}^{1/2}(x+y)dx\right)dy = \int_0^{1/4}\left[\frac{x^2}{2} + xy\right]_{x=\sqrt{y}}^{x=1/2} dy = \frac{3}{160}.$$

(ii) Let $D := \{(x, y) \in \mathbb{R}^2 : 0 \le x \le 1 \text{ and } 0 \le y \le 2x\}$ and $f(x, y) := e^{x^2}$ for $(x, y) \in D$. Then f is continuous on the elementary region D. By the Fubini theorem,

$$\iint_D f = \int_0^1 \left(\int_0^{2x} e^{x^2} dy \right) dx = \int_0^1 2x\, e^{x^2} dx = e - 1.$$

Also, since $D = \{(x, y) \in \mathbb{R}^2 : 0 \le y \le 2 \text{ and } y/2 \le x \le 1\}$,

$$\iint_D f = \int_0^2 \left(\int_{y/2}^1 e^{x^2} dx \right) dy.$$

However, the integral $\int_{y/2}^1 e^{x^2} dx$ cannot be evaluated in terms of known functions.

This example shows that an iterated integral may not always be useful in evaluating a double integral.

Constant Function on a Bounded Subset of $\mathbb{R}^2$

Let D be a bounded subset of $\mathbb{R}^2$. Define

$$1_D : D \to \mathbb{R} \quad \text{by} \quad 1_D(x, y) := 1 \text{ for all } (x, y) \in D.$$

If the boundary ∂D of D is of (two-dimensional) content zero, then the continuous function 1_D is integrable on D as we saw. The converse also holds, that is, if the function 1_D is integrable on D, then ∂D is of (two-dimensional) content zero.

Examples: Let $R := [a, b] \times [c, d]$. If $D := R$, then we have seen that 1_D is integrable and its double integral is equal to $(b - a)(d - c)$. But if $D := \{(x, y) \in R : x, y \in \mathbb{Q}\}$, then the function $1_D^* : R \to \mathbb{R}$, obtained by extending the function 1_D as usual, is the bivariate Dirichlet function on R. We have seen that 1_D^* is not integrable on R, that is, 1_D is not integrable over D. Note: In this case $\partial D = [a, b] \times [c, d]$, and so ∂D is not of (two-dimensional) content zero.

Area of a Bounded Subset of $\mathbb{R}^2$

Let $D \subset \mathbb{R}^2$ be bounded. We say that D has an **area** if the function 1_D is integrable over D, and then we define

$$\text{Area}\,(D) := \iint_D 1_D(x,y)d(x,y).$$

Thus a bounded subset D of $\mathbb{R}^2$ has an area $\iff \partial D$ is of content zero.

Important special case: Consider an elementary domain $D := \{(x,y) \in \mathbb{R}^2 : a \le x \le b \text{ and } \phi_1(x) \le y \le \phi_2(x)\}$, where $\phi_1, \phi_2 : [a,b] \to \mathbb{R}$ are continuous. Then ∂D is of content zero, and by the Fubini theorem,

$$\text{Area}\,(D) = \int_a^b \left(\int_{\phi_1(x)}^{\phi_2(x)} 1\,dy \right) dx = \int_a^b \big(\phi_2(x) - \phi_1(x)\big)\,dx,$$

which was our definition of the **area between the curves** $y = \phi_1(x)$ and $y = \phi_2(x)$, $x \in [a,b]$.

The following basic inequality is important.

Let D be a bounded subset of $\mathbb{R}^2$ which has an area. If $f : D \to \mathbb{R}$ is an integrable function, and if $|f| \le \alpha$ on D, then

$$\left| \iint_D f(x,y)d(x,y) \right| \le \iint_D |f(x,y)|d(x,y) \le \alpha \, \text{Area}\,(D).$$

The following result is useful in evaluating a double integral.

Theorem (Domain Additivity)

Let D be a bounded subset of $\mathbb{R}^2$, and let $f : D \to \mathbb{R}$ be a bounded function. Suppose $D = D_1 \cup D_2$, where $D_1 \cap D_2$ is of (two-dimensional) content zero. If f is integrable over D_1 and over D_2, then f is integrable over D and

$$\iint_D f = \iint_{D_1} f + \iint_{D_2} f.$$

MA 105 : Calculus
Division 1, Lecture 19

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Fubini's theorem on a rectangle. Examples
- Remark about a theorem of Fichtenholtz and Lichtenstein
- Double Riemann sums
- Double integrals over bounded sets
- Volume of a solid bounded by two surfaces defined over a planar domain
- Notion of sets of (two-dimensional) content zero. Examples
- Necessary condition for the integrability of a bounded function over a bounded subset
- Elementary regions
- Fubini Theorem over Elementary Regions. Examples
- Area of a bounded subset of R2
- General version of domain additivity

Change of Variables in a Double Integral

The calculation of the double integral $\iint_D f(x,y)d(x,y)$ can often be simplified by reducing it to another double integral $\iint_E g(u,v)d(u,v)$, where the pair (u,v) of new variables is related to the the pair (x,y) of the given variables by a suitable transformation.

In doing so, we look for a function g which is 'simpler' than the given function f such that the domain E of g is also 'simpler' than the given domain D of f.

Let $x^\circ, y^\circ, a_1, b_1, a_2, b_2 \in \mathbb{R}$, and consider the **affine transformation** $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\Psi := (\psi_1, \psi_2)$, where

$$x = \psi_1(u,v) := x^\circ + a_1 u + b_1 v \quad \text{and} \quad y = \psi_2(u,v) := y^\circ + a_2 u + b_2 v.$$

Note that the above equations have a unique solution, that is, the transformation Ψ is one-one $\iff a_1 b_2 - a_2 b_1 \neq 0$.

In this case, the square region $E := [0,1]\times[0,1]$ is transformed to the parallelogram $D := \Psi(E)$ with vertices $\Psi(0,0) = (x^\circ, y^\circ)$, $\Psi(1,0) = (x^\circ + a_1, y^\circ + a_2)$, $\Psi(0,1) = (x^\circ + b_1, y^\circ + b_2)$ and $\Psi(1,1) = (x^\circ + a_1 + b_1, y^\circ + a_2 + b_2)$. Note that Area $(E) = 1$, and it gets scaled to the area $|a_1 b_2 - a_2 b_1|$ of $D = \Psi(E)$.

In general, let $E \subset \mathbb{R}^2$, and consider $\Phi := (\phi_1, \phi_2) : E \to \mathbb{R}^2$. Let $Q_0 := (u_0, v_0)$ be an interior point of E. If $x := \phi_1(u,v)$ and $y := \phi_2(u,v)$ for $(u,v) \in E$, and the functions x and y have partial derivatives at (u_0, v_0), then the **Jacobian** of the transformation Φ at (u_0, v_0) is defined by

$$J(\Phi)(u_0, v_0) := \frac{\partial(x,y)}{\partial(u,v)}(u_0, v_0) := \det \begin{bmatrix} \dfrac{\partial x}{\partial u}(Q_0) & \dfrac{\partial x}{\partial v}(Q_0) \\[2ex] \dfrac{\partial y}{\partial u}(Q_0) & \dfrac{\partial y}{\partial v}(Q_0) \end{bmatrix}.$$

In particular, $J(\Psi)(u,v) = a_1 b_2 - a_2 b_1$, $(u,v) \in \mathbb{R}^2$, for the affine transformation Ψ.

Now suppose E is a square region about Q_0, and the transformation $\Phi := (\phi_1, \phi_2) : E \to \mathbb{R}^2$ is such that the functions ϕ_1 and ϕ_2 are differentiable at Q_0.

Let $P_0 := \Phi(u_0, v_0)$. For $(u, v) \in \mathbb{R}^2$, let

$$\begin{aligned}
\psi_1(u, v) &:= \phi_1(u_0, v_0) + \frac{\partial \phi_1}{\partial u}\Big|_{(u_0,v_0)}(u - u_0) + \frac{\partial \phi_1}{\partial v}\Big|_{(u_0,v_0)}(v - v_0), \\
\psi_2(u, v) &:= \phi_2(u_0, v_0) + \frac{\partial \phi_2}{\partial u}\Big|_{(u_0,v_0)}(u - u_0) + \frac{\partial \phi_2}{\partial v}\Big|_{(u_0,v_0)}(v - v_0).
\end{aligned}$$

By the definition of differentiability of ϕ_1 and ϕ_2 at (u_0, v_0),

$$\phi_1(u, v) = \psi_1(u, v) + \epsilon_1(u, v) \quad \text{and} \quad \phi_2(u, v) = \psi_2(u, v) + \epsilon_2(u, v),$$

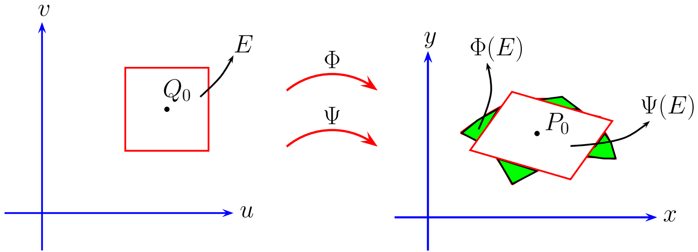
for all $(u, v) \in E$, where $\epsilon_1(u, v)/\|(u - u_0, v - v_0)\| \to 0$ and $\epsilon_2(u, v)/\|(u - u_0, v - v_0)\| \to 0$ as $(u, v) \to (u_0, v_0)$.

Thus the affine transformation $\Psi := (\psi_1, \psi_2) : \mathbb{R}^2 \to \mathbb{R}^2$ maps Q_0 to P_0, and approximates the transformation Φ around Q_0.

Also, by the definition of $\Psi = (\psi_1, \psi_2)$, we obtain

$$J(\Psi) = \left(\frac{\partial \psi_1}{\partial u}\frac{\partial \psi_2}{\partial v} - \frac{\partial \psi_1}{\partial v}\frac{\partial \psi_2}{\partial u}\right) = \left(\frac{\partial \phi_1}{\partial u}\frac{\partial \phi_2}{\partial v} - \frac{\partial \phi_1}{\partial v}\frac{\partial \phi_2}{\partial u}\right)(Q_0) = J(\Phi)(Q_0).$$

Hence Area $(\Psi(E)) = |J(\Psi)|$Area $(E) = |J(\Phi)(Q_0)|$Area (E).



To state our next result, we need the following definition. A subset Ω of $\mathbb{R}^2$ is called **open** if every point of Ω is an interior point of Ω.

Theorem (Change of Variables Formula)

Let D be a closed and bounded subset of $\mathbb{R}^2$ which has an area, and let $f : D \to \mathbb{R}$ be continuous. Suppose Ω is an open subset of $\mathbb{R}^2$ and $\Phi : \Omega \to \mathbb{R}^2$ is a one-one transformation such that $\Phi := (\phi_1, \phi_2)$, where ϕ_1 and ϕ_2 have continuous partial derivatives in Ω and $J(\Phi)(u, v) \neq 0$ for all $(u, v) \in \Omega$. Let $E \subset \Omega$ be such that $\Phi(E) = D$. Then E is a closed and bounded subset of Ω, and E has an area. Moreover, $f \circ \Phi : E \to \mathbb{R}$ is continuous, and

$$\iint_D f(x, y)d(x, y) = \iint_E (f \circ \Phi)(u, v)|J(\Phi)(u, v)|d(u, v).$$

Note: The hypothesis $J(\Phi)(u, v) \neq 0$ for all $(u, v) \in \Omega$ can be weakened by assuming only that $J(\Phi)(u, v) \geq 0$ for all $(u, v) \in \Omega$ or $J(\Phi)(u, v) \leq 0$ for all $(u, v) \in \Omega$, and $J(\Phi)(u, v) = 0$ only for (u, v) in a subset of Ω of (two-dimensional) content zero.

Compare the above result on Change of Variables for double integrals with the result on Integration by Substitution for Riemann integrals proved in Lecture 12:

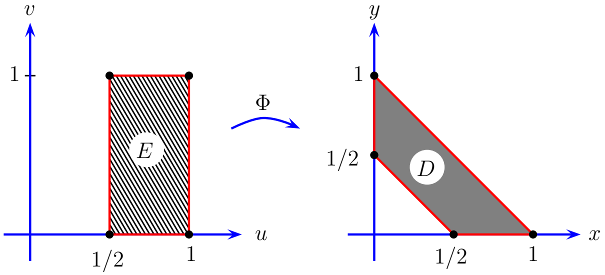
(Integration by substitution)

Let $f : [a, b] \to \mathbb{R}$ be continuous, and let $\phi : [\alpha, \beta] \to \mathbb{R}$ be a continuously differentiable function such that $\phi([\alpha, \beta]) = [a, b]$. If $\phi'(t) \neq 0$ for all $t \in (\alpha, \beta)$, then

$$\int_a^b f(x)dx = \int_\alpha^\beta f(\phi(t))|\phi'(t)|dt.$$

Example

Let $D := \{(x, y) \in \mathbb{R}^2 : x \geq 0,\ y \geq 0 \text{ and } 1 \leq 2(x + y) \leq 2\}$, and define $f : D \to \mathbb{R}$ by $f(x, y) := y/(x + y)$ for $(x, y) \in D$.



Let $u := x+y,\ v := y/(x+y)$, that is, $x := u(1-v),\ y := uv$.

We let $\Omega := \{(u, v) \in \mathbb{R}^2 : u > 0\}$, and define $\Phi : \Omega \to \mathbb{R}^2$ by $\Phi(u, v) = \big(u(1 - v), uv\big)$. Then the map Φ is one-one from Ω onto $\{(x, y) \in \mathbb{R}^2 : x + y > 0\}$. If $\Phi = (\phi_1, \phi_2)$, then the partial derivatives of ϕ_1 and ϕ_2 exist and are continuous.

Also,

$$J(\Phi)(u,v) = \det \begin{bmatrix} 1-v & -u \\ v & u \end{bmatrix} = u \neq 0 \quad \text{for all } (u,v) \in \Omega.$$

Further, if $E := [1/2, 1] \times [0, 1]$, then $\Phi(E) = D$ and $(f \circ \Phi)(u,v) = v$ for all $(u,v) \in E$. Since f is continuous on D, the Change of Variables formula gives

$$\begin{aligned} \iint_D f(x,y)d(x,y) &= \iint_E f(u(1-v), uv)|u|d(u,v) \\ &= \iint_E uv\, d(u,v) \\ &= \left(\int_{1/2}^1 u\, du\right)\left(\int_0^1 v\, dv\right) \\ &= \frac{3}{16}. \end{aligned}$$

Polar Transformation Let $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $\Phi(r,\theta) := (r\cos\theta,\ r\sin\theta)$. If $\Phi := (\phi_1, \phi_2)$, then ϕ_1 and ϕ_2 have continuous partial derivatives in $\mathbb{R}^2$ and

$$J(\Phi)(r,\theta) = \det \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix} = r \quad \text{for all } (r,\theta) \in \mathbb{R}^2.$$

Thus the Jacobian of Φ is nonzero on $\{(r,\theta) \in \mathbb{R}^2 : r \neq 0\}$.
Also, the function $\Phi : \{(r,\theta) \in \mathbb{R}^2 : r > 0 \text{ and } -\pi < \theta \leq \pi\}$ $\to \{(x,y) \in \mathbb{R}^2 : (x,y) \neq (0,0)\}$ is one-one and onto.
In this case, the following result holds.

Proposition

Let D be a closed and bounded subset of $\mathbb{R}^2$ which has an area, and let $f : D \to \mathbb{R}$ be continuous. Suppose the set $E := \{(r,\theta) \in \mathbb{R}^2 : r \geq 0, -\pi \leq \theta \leq \pi, (r\cos\theta, r\sin\theta) \in D\}$, has an area. Then the function $(r,\theta) \longmapsto f(r\cos\theta, r\sin\theta)$ is continuous on E and
$\iint_D f(x,y)d(x,y) = \iint_E f(r\cos\theta, r\sin\theta)\, r\, d(r,\theta)$.

Examples

Let $D := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$. Define

$E := \{(r, \theta) \in \mathbb{R}^2 : r \ge 0, -\pi \le \theta \le \pi \text{ and } (r\cos\theta, r\sin\theta) \in D\}$.

Then $E := [0, 1] \times [-\pi, \pi]$.

(i) Let $f(x, y) := \sqrt{1 - x^2 - y^2}$ for $(x, y) \in D$. Then

$$\begin{aligned}
\iint_D f(x, y)d(x, y) &= \iint_E f(r\cos\theta, r\sin\theta)\, r\, d(r, \theta) \\
&= \int_{-\pi}^{\pi} \left(\int_0^1 \sqrt{1 - r^2}\, r\, dr \right) d\theta \\
&= \int_{-\pi}^{\pi} \frac{1}{2} \left(\int_0^1 \sqrt{s}\, ds \right) d\theta = \frac{2\pi}{3}.
\end{aligned}$$

(ii) Let $f(x, y) := e^{x^2+y^2}$ for $(x, y) \in D$. Then

$$\iint_D f(x, y)d(x, y) = \int_{-\pi}^{\pi} \left(\int_0^1 e^{r^2} r\ dr \right) d\theta = \int_{-\pi}^{\pi} \left(\frac{e-1}{2} \right) d\theta = \pi(e-1).$$

Triple Integrals

The theory of double integrals of functions of two variables readily extends to triple integrals of functions of three variables. No new ideas are needed. We shall, therefore, mention only a few important points.

Consider a **cuboid** $K := [a, b] \times [c, d] \times [p, q]$, where $a < b$, $c < d$, $p < q$. We consider a partition P of K into several subcuboids.

For a bounded function $f : K \to \mathbb{R}$, we define the corresponding lower triple sum $L(P, f)$, the upper triple sum $U(P, f)$, the lower triple integral $L(f)$, and the upper triple integral $U(f)$. As before, f is **(triple) integrable** on K, if $L(f) = U(f)$. In this case, its **triple integral** is denoted by

$$\iiint_K f(x, y, z)d(x, y, z) \quad \text{or} \quad \iiint_K f.$$

If $f := 1$ on K, then f is integrable on K, and its triple integral is equal to the volume of K, whereas if f is the trivariate Dirichlet function on K, then f is not integrable on K.

Further, f is integrable on K if and only if it satisfies the Riemann condition. If f is monotonic in each of the three variables, or if f is continuous on K, then f is integrable.

Fubini's theorem on cuboids

Let f be integrable on K, and let I denote its triple integral. Suppose for each fixed $x \in [a, b]$, the double integral $\iint_{[c,d]\times[p,q]} f(x, y, z)d(y, z)$ exists. Then the **iterated integral** $\int_a^b \left(\iint_{[c,d]\times[p,q]} f(x, y, z)d(y, z)\right)dx$ exists and equals I. Further, if for each fixed $(x, y) \in [a, b] \times [c, d]$, the Riemann integral $\int_p^q f(x, y, z)dz$ exists, then the **iterated integral** $\int_a^b \left[\int_c^d \left(\int_p^q f(x, y, z)dz\right)dy\right]dx$ exists and equals I.

There are other versions of this result with the roles of x, y, z interchanged. They are used in evaluating triple integrals.

Let D be a bounded subset of $\mathbb{R}^3$, and let $f : D \to \mathbb{R}$ be a bounded function. Consider a cuboid $K := [a, b] \times [c, d] \times [p, q]$ such that $D \subset K$, and define $f^* : K \to \mathbb{R}$ by

$$f^*(x, y, z) := \begin{cases} f(x, y, z) & \text{if } (x, y, z) \in D, \\ 0 & \text{otherwise.} \end{cases}$$

We say that f is **integrable** over D if f^* is integrable on K, and in this case, the **triple integral** of f (over D) is defined to be the triple integral of f^* (on K), that is,

$$\iiint_D f(x, y, z)d(x, y, z) := \iiint_K f^*(x, y, z)d(x, y, z).$$

We may also denote the triple integral by $\displaystyle\iiint_D f$.

The triple integral over a bounded subset of $\mathbb{R}^3$ has algebraic and order properties analogous to those of a double integral over a bounded subset of $\mathbb{R}^2$.

In order to seek conditions under which a bounded function f defined on a bounded subset D of $\mathbb{R}^3$ is integrable over D, we introduce the following concept.

A bounded subset E of $\mathbb{R}^3$ is of (three-dimensional) **content zero** if for every $\epsilon > 0$, there are finitely many cuboids whose union contains E and the sum of whose volumes is less than ϵ.

Examples:
(i) Let D_0 be a bounded subset of $\mathbb{R}^2$, and let $\varphi : D_0 \to \mathbb{R}$ be integrable over D_0. Then its graph $E := \{(x, y, \varphi(x, y)) : (x, y) \in D_0\}$ is of (three-dimensional) content zero. To see this, let R be a rectangle containing D_0, and let $\varphi^* : R \to \mathbb{R}$ be the usual extension of φ. Let $\epsilon > 0$. By the Riemann condition for the integrable function φ^*, there is a partition $P := \{(x_i, y_j) : i = 0, 1, \ldots, n \text{ and } j = 0, 1, \ldots, k\}$ of R such that $U(P, \varphi^*) - L(P, \varphi^*) < \epsilon$. Then the nk cuboids $K_{i,j} := [x_{i-1}, x_i] \times [y_{j-1}, y_j] \times [m_{i,j}(\varphi^*), M_{i,j}(\varphi^*)], i = 1, \ldots n$; $j = 1, \ldots, k$, work.

(ii) Let functions $x,\ y,\ z : [a, b]\times[c, d] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on $(a, b)\times(c, d)$, and let their partial derivatives be bounded on $(a, b)\times(c, d)$. Then the set $E := \{(x(u, v), y(u, v), z(u, v)) \in \mathbb{R}^3 : (u, v) \in [a, b]\times[c, d]\}$ is of three-dimensional content zero. We prove this as follows.

For $n \in \mathbb{N}$, consider the partition of $[a, b]\times[c, d]$ into n^2 equal parts, and let (u_i, v_j) be the centroid of the (i, j)th subrectangle for $i, j = 1, \ldots, n$.

Let $(u, v) \in [a, b]\times[c, d]$. Then there are $i, j \in \{1, \ldots, n\}$ such that $|u - u_i| \le (b - a)/2n$ and $|v - v_j| \le (d - c)/2n$. By the Bivariate Mean Value Theorem, there are $(p_x, q_x), (p_y, q_y)$ and (p_z, q_z) on the open line segment joining (u, v) and (u_i, v_j) (and hence belonging to $(a, b)\times(c, d)$) such that

$$\begin{aligned}
x(u,v) - x(u_i,v_j) &= x_u(p_x,q_x)(u-u_i) + x_v(p_x,q_x)(v-v_j),\\
y(u,v) - y(u_i,v_j) &= y_u(p_y,q_y)(u-u_i) + y_v(p_y,q_y)(v-v_j),\\
z(u,v) - z(u_i,v_j) &= z_u(p_z,q_z)(u-u_i) + z_v(p_z,q_z)(v-v_j).
\end{aligned}$$

Let $\alpha \in \mathbb{R}$ with $|x_u|,\ |x_v|,\ |y_u|,\ |y_v| \le \alpha$ on $(a,b)\times(c,d)$, and $\beta := b-a+d-c$. Then

$$|x(u,v)-x(u_i,v_j)|,\ |y(u,v)-y(u_i,v_j)|,\ |z(u,v)-z(u_i,v_j)| \le \frac{\alpha\beta}{2n},$$

and so $(x(u,v),y(u,v),z(u,v))$ lies in a cube $K_{i,j}$ centered at $(x(u_i,v_j),y(u_i,v_j),z(u_i,v_j))$ and of volume $\alpha^3\beta^3/n^3$.

Thus the set S is contained in the union of n^2 cubes $K_{i,j},\ i,j = \ldots n$, the sum of whose volumes is equal to $\alpha^3\beta^3/n$.

Given $\epsilon > 0$, choose $n \in \mathbb{N}$ such that $n > \alpha^3\beta^3/\epsilon$. Then S is contained in the union of these n^2, the sum of whose volumes is less that ϵ. Hence S is of three-dimensional content zero.

The above result can be improved by replacing the rectangle $[a, b] \times [c, d]$ by a bounded convex subset E of $\mathbb{R}^2$ whose interior E° is nonempty.

Further, if E is any bounded convex subset of $\mathbb{R}^2$, and x, y and z are differentiable functions on an open subset of $\mathbb{R}^2$ containing E, and if their first partial derivatives are bounded on E, then we can prove that the set

$$S := \{(x(u, v), y(u, v), z(u, v)) \in \mathbb{R}^3 : (u, v) \in E\}$$

is of three-dimensional content zero.

Remark: The subset $E := [a, b] \times [c, d] \times \{0\}$ of $\mathbb{R}^3$ is of three-dimensional content zero, but the subset $[a, b] \times [c, d]$ of $\mathbb{R}^2$ is not of two-dimensional content zero.

Theorem

Let D be a bounded subset of $\mathbb{R}^3$, and $f : D \to \mathbb{R}$ be a bounded function. Suppose (i) the set of discontinuities of f in D is of (three-dimensional) content zero and (ii) the boundary ∂D of D is of (three-dimensional) content zero. Then f is integrable over D.

We omit the proof.

Suppose D_0 is a subset of $\mathbb{R}^2$ having an area, that is, ∂D_0 is of two-dimensional content zero. Let $\psi_1, \psi_2 : D_0 \to \mathbb{R}$ be continuous, and let $\psi_1 \leq \psi_2$. Consider an **elementary region**

$$D := \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in D_0 \text{ and } \psi_1(x, y) \leq z \leq \psi_2(x, y)\}.$$

Then ∂D is of three-dimensional content zero. Hence if a function is continuous on D, then it is integrable over D.

Proposition (Cavalieri Principle)

Let D be a bounded subset of $\mathbb{R}^3$, and let $f : D \to \mathbb{R}$ be continuous. Suppose $D := \{(x, y, z) \in \mathbb{R}^3 : (x, y) \in D_0$ and $\psi_1(x, y) \le z \le \psi_2(x, y)\}$, where D_0 is a subset of $\mathbb{R}^2$ having an area, $\psi_1, \psi_2 : D_0 \to \mathbb{R}$ are continuous and $\psi_1 \le \psi_2$. Then f is integrable on D, the **iterated integral** exists, and

$$\iiint_D f = \iint_{D_0} \left(\int_{\psi_1(x,y)}^{\psi_2(x,y)} f(x, y, z) dz \right) d(x, y).$$

In particular, if D_0 is an elementary region in $\mathbb{R}^2$ given by $D_0 := \{(x, y) : \mathbb{R}^2 : a \le x \le b \text{ and } \phi_1(x) \le y \le \phi_2(x)\}$, where $\phi_1, \phi_2 : [a, b] \to \mathbb{R}$ are continuous and $\phi_1 \le \phi_2$, then

$$\iiint_D f = \int_a^b \left(\int_{\phi_1(x)}^{\phi_2(x)} \left(\int_{\psi_1(x,y)}^{\psi_2(x,y)} f(x, y, z)\, dz \right) dy \right) dx.$$

The Cavalieri Principle stated above gives a version of Fubini's theorem for an 'elementary domain' D in $\mathbb{R}^3$. Other versions can be obtained similarly.

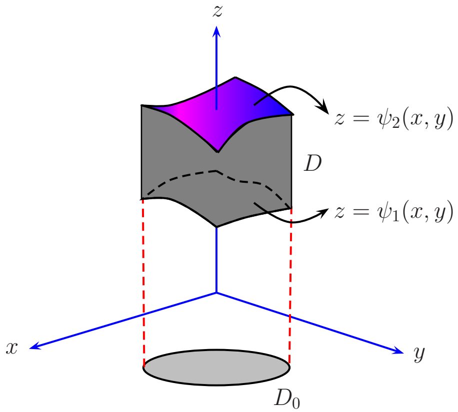


Figure: Illustration of the Cavalieri Principle: a solid between two surfaces defined over D_0.

Constant Function on a Bounded Subset of $\mathbb{R}^3$

Let D be a bounded subset of $\mathbb{R}^3$. Define

$$1_D : D \to \mathbb{R} \quad \text{by} \quad 1_D(x, y, z) := 1 \text{ for all } (x, y, z) \in D.$$

If the boundary ∂D of D is of (three-dimensional) content zero, then the continuous function 1_D is integrable on D. The converse also holds, that is, if the function 1_D is integrable on D, then ∂D is of (three-dimensional) content zero.

Examples:
Let $K := [a, b]\times[c, d]\times[p, q]$. If $D := K$, then 1_D is integrable and its triple integral is equal to $(b - a)(d - c)(q - p)$. But if $D := \{(x, y, z) \in K : x, y, z \in \mathbb{Q}\}$, then the function $1_D^* : K \to \mathbb{R}$, obtained by extending the function 1_D as usual, is the trivariate Dirichlet function on K. Now 1_D^* is not integrable on K, that is, 1_D is not integrable over D. Note: In this case, $\partial D = K$ is not of three-dimensional content zero.

Volume of a Bounded Subset of $\mathbb{R}^3$

Let $D \subset \mathbb{R}^3$ be bounded. We say that D has a **volume** if the function 1_D is integrable over D, and then we define

$$\text{Vol}\,(D) := \iiint_D 1_D(x,y,z)d(x,y,z).$$

Important special case: Consider an elementary domain $D := \{(x,y,z) \in \mathbb{R}^3 : (x,y) \in D_0$ and $\psi_1(x,y) \leq z \leq \psi_2(x,y)\}$, where D_0 is a subset of $\mathbb{R}^2$ having an area, $\psi_1, \psi_2 : D_0 \to \mathbb{R}$ are continuous and $\psi_1 \leq \psi_2$.

By the Cavalier Principle, the volume of D is equal to

$$\iint_{D_0} \left(\int_{\psi_1(x,y)}^{\psi_2(x,y)} dz \right) d(x,y) = \iint_{D_0} \big(\psi_2(x,y) - \psi_1(x,y)\big)\, d(x,y),$$

which was our definition (Lecture 18) of the **volume between the surfaces** $z = \psi_1(x,y)$ and $z = \psi_2(x,y)$, $(x,y) \in D_0$.

For instance, let

$$D := \{(x,y,z) \in \mathbb{R}^3 \colon x^2 + y^2 + 4z^2 \le 6\}.$$

Then $D = \left\{(x,y,z) \in \mathbb{R}^3 : |x| \le \sqrt{6},\ |y| \le \sqrt{6-x^2} \text{ and } |z| \le \sqrt{6-x^2-y^2}/2\right\}$. Hence the volume of the ellipsoidal solid D is equal to

$$8\int_0^{\sqrt{6}} \left(\int_0^{\sqrt{6-x^2}} \left(\int_0^{\sqrt{6-x^2-y^2}\big/2} dz \right) dy \right) dx = 4\sqrt{6}\,\pi.$$

Note: It can be shown that the definitions of the volume of a solid used in the slice method, the washer method and the shell method agree with the general definition of the volume of a bounded subset of $\mathbb{R}^3$ given above.

The following basic inequality is important. Let D be a bounded subset of $\mathbb{R}^3$ which has a volume. If $f : D \to \mathbb{R}$ is an integrable function, and if $|f| \leq \alpha$ on D, then

$$\left| \iiint_D f(x,y,z) d(x,y,z) \right| \leq \iiint_D |f(x,y,z)| d(x,y,z) \leq \alpha \operatorname{Vol}(D).$$

The following result is useful in evaluating a triple integral.

Theorem (Domain Additivity)

Let D be a bounded subset of $\mathbb{R}^3$, and let $f : D \to \mathbb{R}$ be a bounded function. Suppose $D = D_1 \cup D_2$, where $D_1 \cap D_2$ is of (three-dimensional) content zero. If f is integrable over D_1 and over D_2, then f is integrable over D and

$$\iiint_D f = \iiint_{D_1} f + \iiint_{D_2} f.$$

Change of Variables in a Triple Integral

To simplify the calculation of $\iiint_D f(x, y, z)d(x, y, z)$, we may reduce it to another triple integral $\iiint_E g(u, v, w)d(u, v, w)$.

Let $E \subset \mathbb{R}^3$ and $\Phi := (\phi_1, \phi_2, \phi_3) : E \to \mathbb{R}^3$. Suppose $Q_0 := (u_0, v_0, w_0)$ is an interior point of E.
If $x := \phi_1(u, v, w)$, $y := \phi_2(u, v, w)$ and $z := \phi_3(u, v, w)$ for $(u, v, w) \in E$, and the functions x, y and z have partial derivatives at Q_0, then the **Jacobian** of Φ at Q_0 is given by

$$J(\Phi)(Q_0) := \frac{\partial(x, y, z)}{\partial(u, v, w)}(Q_0) := \det \begin{bmatrix} \frac{\partial x}{\partial u}(Q_0) & \frac{\partial x}{\partial v}(Q_0) & \frac{\partial x}{\partial w}(Q_0) \\ \frac{\partial y}{\partial u}(Q_0) & \frac{\partial y}{\partial v}(Q_0) & \frac{\partial y}{\partial w}(Q_0) \\ \frac{\partial z}{\partial u}(Q_0) & \frac{\partial z}{\partial v}(Q_0) & \frac{\partial z}{\partial w}(Q_0) \end{bmatrix}.$$

Theorem (Change of Variables Formula)

Let D be a closed and bounded subset of $\mathbb{R}^3$ which has a volume. Let $f : D \to \mathbb{R}$ be continuous. Suppose Ω is an open subset of $\mathbb{R}^3$ and $\Phi : \Omega \to \mathbb{R}^3$ is a one-one transformation such that $\Phi := (\phi_1, \phi_2, \phi_3)$, where ϕ_1, ϕ_2, ϕ_3 have continuous partial derivatives in Ω and $J(\Phi)(u, v, w) \neq 0$ for all $(u, v, w) \in \Omega$. Let $E \subset \Omega$ be such that $\Phi(E) = D$. Then E is a closed and bounded subset of Ω, and E has a volume. Moreover, $f \circ \Phi : E \to \mathbb{R}$ is continuous, and

$$\iiint_D f = \iiint_E (f \circ \Phi)(u, v, w) |J(\Phi)(u, v, w)| d(u, v, w).$$

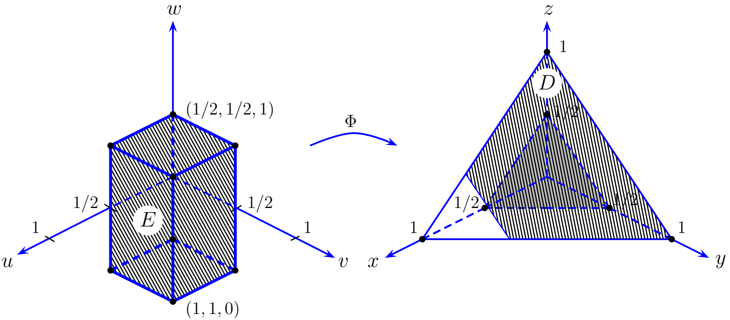
Note: The hypothesis $J(\Phi)(u, v, w) \neq 0$ for all $(u, v, w) \in \Omega$ can be weakened by assuming only that $J(\Phi)(u, v, w) \geq 0$ for all $(u, v, w) \in \Omega$ or $J(\Phi)(u, v, w) \leq 0$ for all $(u, v, w) \in \Omega$, and $J(\Phi)(u, v, w) = 0$ only for (u, v, w) in a subset of Ω of (three-dimensional) content zero.

Example: Let

$$D := \{(x,y,z) \in \mathbb{R}^3 : x,y,z \geq 0,\ x \leq y+z,\ 1 \leq 2(x+y+z) \leq 2\},$$

and define $f : D \to \mathbb{R}$ by $f(x,y,z) := z/(y+z)$.

(Note: $(x,y,z) \in D \implies 0 \leq x \leq x/2 + (y+z)/2 \leq 1/2$.)



$$\text{Let} \quad u := x+y+z, \quad v := \frac{y+z}{x+y+z}, \quad w := \frac{z}{y+z},$$

that is, $x := u(1-v), \quad y := u\,v(1-w), \quad z := u\,v\,w.$

Consider $\Omega := \{(u, v, w) \in \mathbb{R}^3 : u > 0 \text{ and } v > 0\}$, and define $\Phi : \Omega \to \mathbb{R}^3$ by $\Phi(u, v, w) := \big(u(1-v), u\,v(1-w), u\,v\,w\big)$. Now Φ maps Ω onto $\{(x, y, z) \in \mathbb{R}^2 : x + y + z > 0 \text{ and } y + z > 0\}$ and is one-one on Ω. Also, if $\Phi = (\phi_1, \phi_2, \phi_3)$, then the partials of ϕ_1, ϕ_2, and ϕ_3 are continuous, and for all $(u, v, w) \in \Omega$,

$$J(\Phi)(u, v, w) = \det \begin{bmatrix} 1-v & -u & 0 \\ v(1-w) & u(1-w) & -u\,v \\ v\,w & u\,w & u\,v \end{bmatrix} = u^2 v \neq 0.$$

Let $E := [1/2, 1] \times [1/2, 1] \times [0, 1]$. Then $\Phi(E) = D$, and $(f \circ \Phi)(u, v, w) = w$ for $(u, v, w) \in E$. By the CoV formula,

$$\begin{aligned} \iiint_D f &= \iiint_E f\big(u(1-v), u\,v(1-w), u\,v\,w\big)|u^2 v|\,d(u, v, w) \\ &= \iiint_E u^2 v\,w\,d(u, v, w) \\ &= \left(\int_{1/2}^1 u^2\,du\right)\left(\int_{1/2}^1 v\,dv\right)\left(\int_0^1 w\,dw\right) = \frac{7}{128}. \end{aligned}$$

Two important cases involving a change of variables in triple integrals are given by switching to cylindrical coordinates, and to spherical coordinates.

(i) Cylindrical coordinates: Let $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

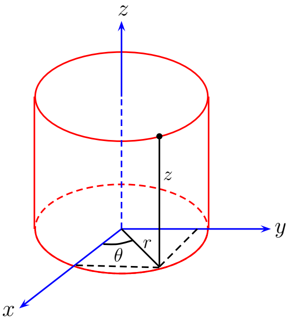
$$\Phi(r, \theta, z) := (r\cos\theta,\ r\sin\theta,\ z) \quad \text{for } (r, \theta, z) \in \mathbb{R}^3.$$

Then for all $(r, \theta, z) \in \mathbb{R}^3$,

$$J(\Phi)(r, \theta, z) = \det \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = r.$$

The Jacobian of Φ is nonzero on $\{(r, \theta, z) \in \mathbb{R}^3 : r \neq 0\}$.

The function $\Phi : \{(r, \theta, z) \in \mathbb{R}^3 : r > 0 \text{ and } -\pi < \theta \leq \pi\} \to \{(x, y, z) \in \mathbb{R}^3 : (x, y) \neq (0, 0)\}$ is one-one and onto.



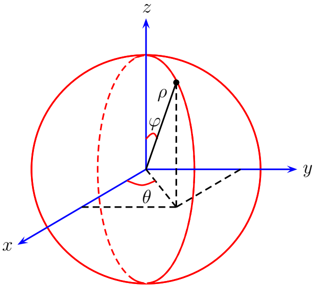


Figure: Illustrations of cylindrical and spherical coordinates.

(ii) Spherical coordinates: Let $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

$$\Phi(\rho, \varphi, \theta) := (\rho \sin\varphi \cos\theta,\ \rho \sin\varphi \sin\theta,\ \rho \cos\varphi) \ \text{ for } (\rho, \varphi, \theta) \in \mathbb{R}^3.$$

Then for all $(\rho, \varphi, \theta) \in \mathbb{R}^3$,

$$\begin{aligned} J(\Phi)(\rho, \varphi, \theta) &= \det \begin{bmatrix} \sin\varphi \cos\theta & \rho \cos\varphi \cos\theta & -\rho \sin\varphi \sin\theta \\ \sin\varphi \sin\theta & \rho \cos\varphi \sin\theta & \rho \sin\varphi \cos\theta \\ \cos\varphi & -\rho \sin\varphi & 0 \end{bmatrix} \\ &= \rho^2 \sin\varphi. \end{aligned}$$

The Jacobian of Φ is nonzero on the set
$\{(\rho, \varphi, \theta) \in \mathbb{R}^3 : \rho \neq 0 \text{ and } \varphi \neq m\pi \text{ for any } m \in \mathbb{Z}\}$.

The function $\Phi : \{(\rho, \varphi, \theta) \in \mathbb{R}^3 : \rho > 0,\ 0 < \varphi < \pi$ and $-\pi < \theta \leq \pi\} \to \{(x, y, z) \in \mathbb{R}^3 : (x, y) \neq (0, 0)\}$ is one-one and onto.

In these cases the following results hold.

Proposition

Let D be a closed and bounded subset of $\mathbb{R}^3$ which has a volume, and let $f : D \to \mathbb{R}$ be continuous.

(i) If $E := \{(r, \theta, z) \in \mathbb{R}^3 : r \geq 0, -\pi \leq \theta \leq \pi$ and $(r\cos\theta,\ r\sin\theta,\ z) \in D\}$, and if E has a volume, then the triple integral of f over D is equal to

$$\iiint_E f(r\cos\theta, r\sin\theta, z)\, r\, d(r, \theta, z).$$

(ii) If $E := \{(\rho, \varphi, \theta) \in \mathbb{R}^3 : \rho \geq 0,\ 0 \leq \varphi \leq \pi,\ -\pi \leq \theta \leq \pi$ and $(\rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta, \rho\cos\varphi) \in D\}$, and if E has a volume, then the triple integral of f over D is equal to

$$\iiint_E f(\rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta, \rho\cos\varphi)\, \rho^2\sin\varphi\ d(\rho, \varphi, \theta).$$

Examples

(i) Let $D := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1 \text{ and } 0 \leq z \leq 1\}$, and define $f : D \to \mathbb{R}$ by $f(x, y, z) := z\sqrt{1 - x^2 - y^2}$. Then

$$\begin{aligned} E &:= \{(r, \theta, z) \in \mathbb{R}^3 : r \geq 0,\ -\pi \leq \theta \leq \pi,\ (r\cos\theta, r\sin\theta, z) \in D\} \\ &= [0, 1]\times[-\pi, \pi]\times[0, 1]. \end{aligned}$$

Hence

$$\begin{aligned} \iiint_D f &= \iiint_E f(r\cos\theta, r\sin\theta, z)\, {\color{red} r}\, d(r, \theta, z) \\ &= \int_0^1 \left[\int_{-\pi}^{\pi} \left(\int_0^1 z\sqrt{1 - r^2}\, r\, dz\right) d\theta\right] dr \\ &= \pi \int_0^1 \sqrt{1 - r^2}\, r\, dr = \frac{\pi}{2}\int_0^1 \sqrt{s}\, ds = \frac{\pi}{3}. \end{aligned}$$

(ii) Let $a > 0$ and $D := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq a^2\}$. Define $f : D \to \mathbb{R}$ by $f(x, y, z) = z^2$. Then

$$\begin{aligned} E &:= \{(\rho, \varphi, \theta) \in \mathbb{R}^3 : \rho \geq 0,\ 0 \leq \varphi \leq \pi,\ -\pi \leq \theta \leq \pi \text{ and} \\ &\qquad (\rho \sin\varphi \cos\theta, \rho \sin\varphi \sin\theta, \rho \cos\varphi) \in D\} \\ &= [0, a] \times [0, \pi] \times [-\pi, \pi]. \end{aligned}$$

Hence the triple integral of f over D is equal to

$$\iiint_E f(\rho \sin\varphi \cos\theta, \rho \sin\varphi \sin\theta, \rho \cos\varphi)\, \rho^2 \sin\varphi \, d(\rho, \varphi, \theta)$$

$$= \int_0^a \left[\int_0^\pi \left(\int_{-\pi}^\pi (\rho^2 \cos^2\varphi)\, \rho^2 \sin\varphi \, d\theta \right) d\varphi \right] d\rho$$

$$= 2\pi \int_0^a \rho^4 \left(\int_0^\pi \cos^2\varphi \sin\varphi \, d\varphi \right) d\rho = \frac{2\pi a^5}{5} \cdot \frac{2}{3} = \frac{4\pi a^5}{15}.$$

MA 105 : Calculus
Division 1, Lecture 20

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Change of Variables in a Double Integral. Example
- Polar transformation, Example
- Triple integrals
- Sets of (three-dimensional) content zero. Examples
- Criterion for the integrability of a function of 3 variables
- Elemenentary regions. Cavalieri Principle
- Volume of a bouded subset of the 3-space
- Domain additivity for triple integrals
- Change of Variables in a Triple Integral. Example
- Cylindrical and Spherical coordinates. Examples.

Theorem (Change of Variables Formula)

Let D be a closed and bounded subset of $\mathbb{R}^3$ which has a volume. Let $f : D \to \mathbb{R}$ be continuous. Suppose Ω is an open subset of $\mathbb{R}^3$ and $\Phi : \Omega \to \mathbb{R}^3$ is a one-one transformation such that $\Phi := (\phi_1, \phi_2, \phi_3)$, where ϕ_1, ϕ_2, ϕ_3 have continuous partial derivatives in Ω and $J(\Phi)(u, v, w) \neq 0$ for all $(u, v, w) \in \Omega$. Let $E \subset \Omega$ be such that $\Phi(E) = D$. Then E is a closed and bounded subset of Ω, and E has a volume. Moreover, $f \circ \Phi : E \to \mathbb{R}$ is continuous, and

$$\iiint_D f = \iiint_E (f \circ \Phi)(u, v, w)|J(\Phi)(u, v, w)|d(u, v, w).$$

Note: The hypothesis $J(\Phi)(u, v, w) \neq 0$ for all $(u, v, w) \in \Omega$ can be weakened by assuming only that $J(\Phi)(u, v, w) \geq 0$ for all $(u, v, w) \in \Omega$ or $J(\Phi)(u, v, w) \leq 0$ for all $(u, v, w) \in \Omega$, and $J(\Phi)(u, v, w) = 0$ only for (u, v, w) in a subset of Ω of (three-dimensional) content zero.

Two important cases involving a change of variables in triple integrals are given by switching to cylindrical coordinates, and to spherical coordinates.

(i) Cylindrical coordinates: Let $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

$$\Phi(r,\theta,z) := (r\cos\theta,\ r\sin\theta,\ z) \quad \text{for } (r,\theta,z) \in \mathbb{R}^3.$$

Then for all $(r,\theta,z) \in \mathbb{R}^3$,

$$J(\Phi)(r,\theta,z) = \det \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} = r.$$

The Jacobian of Φ is nonzero on $\{(r,\theta,z) \in \mathbb{R}^3 : r \neq 0\}$.

The function $\Phi : \{(r,\theta,z) \in \mathbb{R}^3 : r > 0 \text{ and } -\pi < \theta \leq \pi\} \to \{(x,y,z) \in \mathbb{R}^3 : (x,y) \neq (0,0)\}$ is one-one and onto.

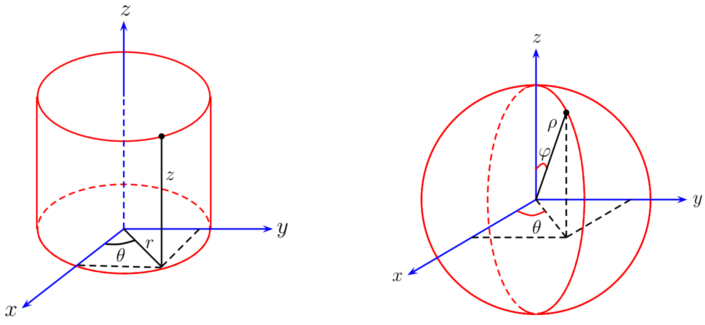


Figure: Illustrations of cylindrical and spherical coordinates.

(ii) Spherical coordinates: Let $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

$$\Phi(\rho, \varphi, \theta) := (\rho \sin\varphi \cos\theta,\ \rho \sin\varphi \sin\theta,\ \rho\cos\varphi) \text{ for } (\rho, \varphi, \theta) \in \mathbb{R}^3.$$

Then for all $(\rho, \varphi, \theta) \in \mathbb{R}^3$,

$$\begin{aligned} J(\Phi)(\rho, \varphi, \theta) &= \det \begin{bmatrix} \sin\varphi\cos\theta & \rho\cos\varphi\cos\theta & -\rho\sin\varphi\sin\theta \\ \sin\varphi\sin\theta & \rho\cos\varphi\sin\theta & \rho\sin\varphi\cos\theta \\ \cos\varphi & -\rho\sin\varphi & 0 \end{bmatrix} \\ &= \rho^2 \sin\varphi. \end{aligned}$$

The Jacobian of Φ is nonzero on the set $\{(\rho, \varphi, \theta) \in \mathbb{R}^3 : \rho \neq 0 \text{ and } \varphi \neq m\pi \text{ for any } m \in \mathbb{Z}\}$.

The function $\Phi : \{(\rho, \varphi, \theta) \in \mathbb{R}^3 : \rho > 0,\ 0 < \varphi < \pi \text{ and } -\pi < \theta \leq \pi\} \to \{(x, y, z) \in \mathbb{R}^3 : (x, y) \neq (0, 0)\}$ is one-one and onto.

In these cases the following results hold.

Proposition

Let D be a closed and bounded subset of $\mathbb{R}^3$ which has a volume, and let $f : D \to \mathbb{R}$ be continuous.

(i) If $E := \{(r, \theta, z) \in \mathbb{R}^3 : r \geq 0, -\pi \leq \theta \leq \pi$ and $(r\cos\theta,\ r\sin\theta,\ z) \in D\}$, and if E has a volume, then the triple integral of f over D is equal to

$$\iiint_E f(r\cos\theta, r\sin\theta, z)\, r\, d(r, \theta, z).$$

(ii) If $E := \{(\rho, \varphi, \theta) \in \mathbb{R}^3 : \rho \geq 0,\ 0 \leq \varphi \leq \pi,\ -\pi \leq \theta \leq \pi$ and $(\rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta, \rho\cos\varphi) \in D\}$, and if E has a volume, then the triple integral of f over D is equal to

$$\iiint_E f(\rho\sin\varphi\cos\theta, \rho\sin\varphi\sin\theta, \rho\cos\varphi)\, \rho^2\sin\varphi\ d(\rho, \varphi, \theta).$$

Vector Algebra

We begin a part of this course known as '**Vector Analysis**'.
It deals with scalar fields and vector fields.
Initially, we recall some concepts in '**Vector Algebra**'.
In Lecture 17, we have defined the **Euclidean space**

$$\mathbb{R}^m := \{\boldsymbol{x} = (x_1, \ldots, x_m) : x_i \in \mathbb{R} \text{ for } i = 1, \ldots, m\},$$

where $m \in \mathbb{N}$. An element of $\mathbb{R}^1 := \mathbb{R}$ is called a **scalar**, and an element of $\mathbb{R}^m$ is called a **vector** if $m \geq 2$.

Let $\boldsymbol{x} := (x_1, \ldots, x_m)$, $\boldsymbol{y} := (y_1, \ldots, y_m) \in \mathbb{R}^m$, and $a \in \mathbb{R}$.
We have already defined the sum $\boldsymbol{x} + \boldsymbol{y}$ and the scalar multiple $a\,\boldsymbol{x}$. Also, we have studied the **dot product** or the **scalar product** of $\boldsymbol{x}$ and $\boldsymbol{y}$ defined by

$$\boldsymbol{x} \cdot \boldsymbol{y} := x_1 y_1 + \cdots + x_m y_m \in \mathbb{R}.$$

Vector Product

Let $m := 3$. Consider $\boldsymbol{x} := (x_1, x_2, x_3)$ and $\boldsymbol{y} := (y_1, y_2, y_3)$. The **cross product** or the **vector product** of $\boldsymbol{x}$ with $\boldsymbol{y}$ is defined by

$$\boldsymbol{x} \times \boldsymbol{y} := (x_2y_3 - x_3y_2,\ x_3y_1 - x_1y_3,\ x_1y_2 - x_2y_1) \in \mathbb{R}^3.$$

$$\text{Let} \quad \boldsymbol{i} := (1, 0, 0), \quad \boldsymbol{j} := (0, 1, 0), \quad \boldsymbol{k} := (0, 0, 1).$$

Then $\boldsymbol{x} := x_1\boldsymbol{i} + x_2\boldsymbol{j} + x_3\boldsymbol{k}$, $\boldsymbol{y} := y_1\boldsymbol{i} + y_2\boldsymbol{j} + y_3\boldsymbol{k}$, and

$$\begin{aligned} \boldsymbol{x} \times \boldsymbol{y} &= (x_2y_3 - x_3y_2)\boldsymbol{i} + (x_3y_1 - x_1y_3)\boldsymbol{j} + (x_1y_2 - x_2y_1)\boldsymbol{k} \\ &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}. \end{aligned}$$

Note: $\boldsymbol{i} \times \boldsymbol{i} = \boldsymbol{j} \times \boldsymbol{j} = \boldsymbol{k} \times \boldsymbol{k} = \boldsymbol{0}$, $\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k}$, $\boldsymbol{j} \times \boldsymbol{k} = \boldsymbol{i}$, $\boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}$.

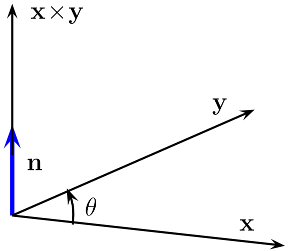
Properties of determinants show that for every $\boldsymbol{z} \in \mathbb{R}^3$,

$$(\boldsymbol{x}+\boldsymbol{y}) \times \boldsymbol{z} = (\boldsymbol{x} \times \boldsymbol{z}) + (\boldsymbol{y} \times \boldsymbol{z}) \quad \text{and} \quad \boldsymbol{y} \times \boldsymbol{x} = -(\boldsymbol{x} \times \boldsymbol{y}).$$

Let $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{y} \neq \mathbf{0}$. It can be shown that

$$\boldsymbol{x} \times \boldsymbol{y} = \|\boldsymbol{x}\| \, \|\boldsymbol{y}\| (\sin\theta) \, \boldsymbol{n},$$

where $\theta \in [0, \pi]$ is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$, and $\boldsymbol{n}$ is the unit vector which is perpendicular to the plane containing $\boldsymbol{x}$ and $\boldsymbol{y}$, and obeys the 'right-hand rule'.



Hence $\|\boldsymbol{x} \times \boldsymbol{y}\| =$ the area of the parallelogram with sides $\boldsymbol{x}$, $\boldsymbol{y}$.

Scalar Triple Product and Vector triple Product

Let $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^3$. Then $\boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z}) \in \mathbb{R}$ and $\boldsymbol{x} \times (\boldsymbol{y} \times \boldsymbol{z}) \in \mathbb{R}^3$ are called the **scalar triple product** and the **vector triple product** of $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ respectively. It is easy to see that if $\boldsymbol{x} := (x_1, x_2, x_3)$, $\boldsymbol{y} := (y_1, y_2, y_3)$, $\boldsymbol{z} := (z_1, z_2, z_3)$, then

$$\boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z}) = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}.$$

Geometrically, $\boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z})$ can be interpreted as the (signed) volume of the parallelopiped defined by the vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$.

One can prove the **Lagrange formula**

$$\boldsymbol{x} \times (\boldsymbol{y} \times \boldsymbol{z}) = (\boldsymbol{x} \cdot \boldsymbol{z})\, \boldsymbol{y} - (\boldsymbol{x} \cdot \boldsymbol{y})\, \boldsymbol{z}$$

by considering each component of the LHS and the RHS.

Scalar Fields and Vector Fields

A **scalar field** is an assignment of a scalar to each point in a region in the space. For example, the temperature at a point on the earth is a scalar field (defined on a subset of $\mathbb{R}^3$).

A **vector field** is an assignment of a vector to each point in a region in the space. For example, the velocity field of a moving fluid is a vector field that associates a velocity vector to each point in the fluid.

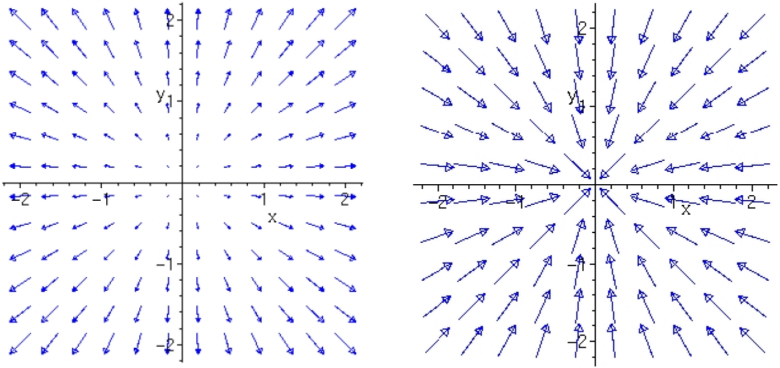
Definition

Let $m \in \mathbb{N}$, and let D be a subset of $\mathbb{R}^m$.

A **scalar field** is a function from D to $\mathbb{R}$.

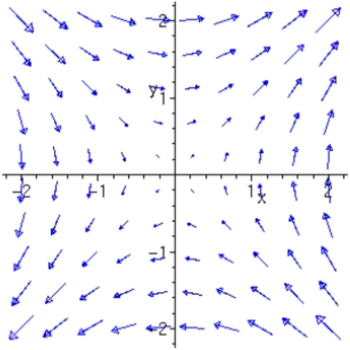
A **vector field** is a function from D to $\mathbb{R}^m$.
If $m = 2$, then it is called a **vector field in the plane**, and if $m = 3$, then it is called a **vector field in the space**.

Depiction of Vector Fields

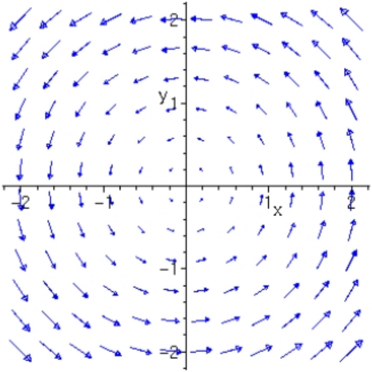


$\boldsymbol{F}(x, y) := (2x, 2y)$ $\qquad$ $\boldsymbol{F}(x, y) := (-x, -y)/\sqrt{x^2 + y^2}$

Depiction of Vector Fields

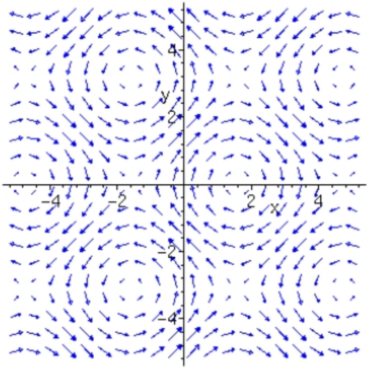


$$\boldsymbol{F}(x, y) := (y, x)$$



$$\boldsymbol{F}(x, y) := (-y, x)$$

Depiction of Vector Fields



$$\boldsymbol{F}(x, y) := (\sin y, \cos x)$$

Smooth Scalar and Vector Fields

Suppose D is an **open** subset of $\mathbb{R}^m$, that is, every point in D is an interior point of D.

A scalar field $f : D \to \mathbb{R}$ is called **smooth** if $\dfrac{\partial f}{\partial x_j}$ exists and is continuous on D for $j = 1, \ldots, m$. The set of all smooth scalar fields on D is denoted by $C^1(D)$. Similarly, the set of all scalar fields on D having continuous partial derivatives of the first and second order is denoted by $C^2(D)$.

Let $\boldsymbol{F} : D \to \mathbb{R}^m$ be a vector field on D, and let

$$\boldsymbol{F}(\boldsymbol{x}) := (F_1(\boldsymbol{x}), \ldots, F_m(\boldsymbol{x})) \quad \text{for } \boldsymbol{x} \in D,$$

where $F_i : D \to \mathbb{R}$ is the ith **component scalar field** on D for $i = 1, \ldots, m$. The vector field $\boldsymbol{F}$ is called **smooth** on D if each F_i is smooth on D, that is, $\dfrac{\partial F_i}{\partial x_j}$ exists and is continuous on D for $i, j = 1, \ldots, m$.

Gradient, Divergence and Curl

Let D be an open subset of $\mathbb{R}^3$.
If f is a smooth scalar field defined on D, then the vector field

$$\operatorname{grad} f := \nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$$

defined on D is called the **gradient field** of f.

If $\boldsymbol{F} := (P, Q, R)$ is a smooth vector field defined on D, then the **divergence field** of $\boldsymbol{F}$ is the scalar field on D defined by

$$\operatorname{div} \boldsymbol{F} := \nabla \cdot \boldsymbol{F} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z},$$

and the **curl field** of $\boldsymbol{F}$ is the vector field on D defined by

$$\operatorname{curl} \boldsymbol{F} := \nabla \times \boldsymbol{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}, \frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}, \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right).$$

As in the case of the cross product $\boldsymbol{x} \times \boldsymbol{y}$, we may write

$$\nabla \times \boldsymbol{F} = \nabla \times (P, Q, R) \;=\; \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}.$$

For any $m \in \mathbb{N}$, we can define the gradient field grad f of a scalar field f defined on an open subset of $\mathbb{R}^m$, and also the divergence field div $\boldsymbol{F}$ of a vector field $\boldsymbol{F}$ defined on an open subset of $\mathbb{R}^m$ in a similar manner. But the curl field curl $\boldsymbol{F}$ is defined only for a vector field $\boldsymbol{F}$ on an open subset of $\mathbb{R}^3$.

If $D \subset \mathbb{R}$ is open, and f is a smooth scalar field on D, then for $(x, y, z) \in \mathbb{R} \times D \times \mathbb{R}$, we can let $\boldsymbol{F}(x, y, z) := (0, f(x), 0)$ and define $\operatorname{curl} f := \operatorname{curl} \boldsymbol{F} = (0, 0, f')$. Also, if $D \subset \mathbb{R}^2$ is open and $\boldsymbol{\Phi} := (P, Q)$ is a smooth vector field on D, then for $(x, y, z) \in D \times \mathbb{R}$, we can let $\boldsymbol{F}(x, y, z) := (P(x, y), Q(x, y), 0)$ and define $\operatorname{curl} \boldsymbol{\Phi} := \operatorname{curl} \boldsymbol{F} = (0, 0, Q_x - P_y)$.

GCD Sequence

Let D be an open subset of $\mathbb{R}^3$. Suppose the first and the second order partial derivatives of $f, P, Q, R : D \to \mathbb{R}$ exist and are continuous on D. By the Mixed Partials Theorem,
(i) $\text{curl}\,(\text{grad}\, f) = \nabla \times (\nabla f) = \mathbf{0}$:

$$\left(\frac{\partial^2 f}{\partial y \partial z} - \frac{\partial^2 f}{\partial z \partial y}, \frac{\partial^2 f}{\partial z \partial x} - \frac{\partial^2 f}{\partial x \partial z}, \frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial x}\right) = (0,0,0),$$

(ii)and if $\boldsymbol{F} := (P, Q, R)$, then $\text{div}\,(\text{curl}\, \boldsymbol{F}) = \nabla \cdot (\nabla \times \boldsymbol{F}) = 0$:

$$\frac{\partial}{\partial x}\left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) + \frac{\partial}{\partial y}\left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) + \frac{\partial}{\partial z}\left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) = 0.$$

Gradient, Curl and Divergence are three **operators**

$$\left\{\begin{matrix}\text{scalar}\\ \text{fields}\end{matrix}\right\} \xrightarrow{\text{grad}} \left\{\begin{matrix}\text{vector}\\ \text{fields}\end{matrix}\right\} \xrightarrow{\text{curl}} \left\{\begin{matrix}\text{vector}\\ \text{fields}\end{matrix}\right\} \xrightarrow{\text{div}} \left\{\begin{matrix}\text{scalar}\\ \text{fields}\end{matrix}\right\}$$

which satisfy $\text{curl}\,(\text{grad}\, f) = \mathbf{0}$ and $\text{div}\,(\text{curl}\, \boldsymbol{F}) = 0$, so that the successive composites are zero.

The above phenomenon raises the following basic questions:

(i) If $\boldsymbol{G}$ is a smooth vector field such that $\operatorname{curl} \boldsymbol{G} = \mathbf{0}$, then must $\boldsymbol{G}$ be a **gradient field**, that is, is there a scalar field f such that $\boldsymbol{G} = \operatorname{grad} f$? (Then f is called a **potential field**.)

(ii) If $\boldsymbol{H}$ is a smooth vector field such that $\operatorname{div} \boldsymbol{H} = 0$, then must $\boldsymbol{H}$ be a **curl field**, that is, is there a vector field $\boldsymbol{F}$ such that $\boldsymbol{H} = \operatorname{curl} \boldsymbol{F}$?

These questions remind us of the following situation. Consider the operator der : $C^1([a, b]) \to C([a, b])$ given by $\operatorname{der}(f) = f'$. If g is a continuous function on $[a, b]$, then is there a continuously differentiable function f on $[a, b]$ such that $g = f'$? The Fundamental theorem of Calculus (Part I) answers this question in the affirmative: If we let $f(x) := \int_a^x g(t)dt$ for $x \in [a, b]$, then $f' = g$ on $[a, b]$.

In view of this, each of the two questions raised above call for a suitable theory of integration, to which we now turn.

Laplacian

Let D be an open subset of $\mathbb{R}^3$, and let f be a smooth scalar field defined on D. Suppose the second order partial derivatives f_{xx}, f_{yy}, f_{zz} exist on D. The **Laplacian field** of f is the scalar field defined on D by

$$\operatorname{div}(\operatorname{grad} f) := \nabla \cdot (\nabla f) = \nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.$$

The Laplacian field is the composition of the operators grad and div:

$$\left\{\begin{matrix}\text{scalar}\\ \text{fields}\end{matrix}\right\} \xrightarrow{\text{grad}} \left\{\begin{matrix}\text{vector}\\ \text{fields}\end{matrix}\right\} \xrightarrow{\text{div}} \left\{\begin{matrix}\text{scalar}\\ \text{fields}\end{matrix}\right\}.$$

The Laplacian plays a very important role in the theory of partial differential equations, and its various applications.

Paths

Let $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$, and let $m \in \mathbb{N}$. A **path** or a **parametrized curve** in $\mathbb{R}^m$ is a continuous function $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$, that is, if $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_m)$, then the function $\gamma_j : [\alpha, \beta] \to \mathbb{R}$ is continuous for each $j = 1, \ldots, m$.

A path $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ is called **closed** if $\boldsymbol{\gamma}(\alpha) = \boldsymbol{\gamma}(\beta)$. A path $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ is called **simple** if $\boldsymbol{\gamma}(t_1) \neq \boldsymbol{\gamma}(t_2)$ for $t_1 < t_2$ in $[\alpha, \beta]$ unless $t_1 = \alpha$ and $t_2 = \beta$.

Let $t \in [\alpha, \beta]$ and suppose $\gamma_1'(t), \ldots, \gamma_m'(t)$ exist. Then

$$\frac{d\boldsymbol{\gamma}}{dt} = \boldsymbol{\gamma}'(t) := (\gamma_1'(t), \ldots, \gamma_m'(t))$$

is called the **tangent vector** to $\boldsymbol{\gamma}$ at t. If $\boldsymbol{\gamma}'(t) \neq \mathbf{0}$, then $\hat{\boldsymbol{t}} := \boldsymbol{\gamma}'(t)/\|\boldsymbol{\gamma}'(t)\|$ is called the **unit tangent vector** to $\boldsymbol{\gamma}$ at t. (We write $\hat{\boldsymbol{t}}$ instead of $\hat{\boldsymbol{t}}_{\boldsymbol{\gamma}}(t)$ for brevity.)

Further, a path $\boldsymbol{\gamma}$ in $\mathbb{R}^m$ is called **smooth** if each $\gamma_j : [\alpha, \beta] \to \mathbb{R}$ is continuously differentiable for $j = 1, \ldots, m$; in case $\boldsymbol{\gamma}$ is a closed curve, that is, if $\boldsymbol{\gamma}(\alpha) = \boldsymbol{\gamma}(\beta)$, we also require $\boldsymbol{\gamma}'(\alpha) = \boldsymbol{\gamma}'(\beta)$. A smooth path is also called a C^1-path.

A smooth path $\boldsymbol{\gamma}$ in $\mathbb{R}^m$ is called **regular** if $\boldsymbol{\gamma}'(t) \neq \mathbf{0}$ for each $t \in [\alpha, \beta]$, that is, if the unit tangent vector to $\boldsymbol{\gamma}$ exists at each $t \in [\alpha, \beta]$.

Examples: For $t \in [-1, 1]$, let $\boldsymbol{\gamma}_1(t) := (t, t^2)$ and $\boldsymbol{\gamma}_2(t) := (t^2, t^3)$. Then $\boldsymbol{\gamma}_1$ is regular, but $\boldsymbol{\gamma}_2$ is not.

A path $\boldsymbol{\gamma}$ in $\mathbb{R}^m$ is called **piecewise smooth** if there are $\alpha := t_0 < t_1 < \cdots < t_n =: \beta$ such that $\boldsymbol{\gamma}$ is smooth on each $[t_{i-1}, t_i]$, $i = 1, \ldots, n$, and it is called **piecewise regular**, if $\boldsymbol{\gamma}$ is regular on each $[t_{i-1}, t_i]$, $i = 1, \ldots, n$.

We shall assume hereafter that all paths are piecewise smooth, unless otherwise stated.

Example Let $a > 0$. Define $\boldsymbol{\gamma}(t) := (a\cos t, a\sin t)$ for $t \in [-\pi, \pi]$. This path is called the **standard parametrized circle in $\mathbb{R}^2$ of radius a.** Since $\boldsymbol{\gamma}(-\pi) = (-1, 0) = \boldsymbol{\gamma}(\pi)$ and $\boldsymbol{\gamma}'(t) := (-a\sin t, a\cos t)$ for $t \in [-\pi, \pi]$, we see that the path $\boldsymbol{\gamma}$ is closed, simple, regular, and its unit tangent vector at $t \in [-\pi, \pi]$ is $\hat{\boldsymbol{t}} := (-\sin t, \cos t)$.

If we let $\widetilde{\boldsymbol{\gamma}}(t) := (a\cos t, -a\sin t)$ for $t \in [-\pi, \pi]$, then it is easy to see that the path $\widetilde{\boldsymbol{\gamma}}$ is also closed, simple, regular, and its unit tangent vector at $t \in [-\pi, \pi]$ is $(-\sin t, -\cos t)$.

Note that $\widetilde{\boldsymbol{\gamma}}([-\pi, \pi]) = \boldsymbol{\gamma}([-\pi, \pi])$, that is, the functions $\widetilde{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}$ have the same range, although they are clearly different paths: one goes around the circle anticlockwise, but the other goes around the circle clockwise.

Further, the path given by $\widetilde{\boldsymbol{\gamma}}(t) := (a\cos 2t, a\sin 2t)$ for $t \in [-\pi, \pi]$ is closed and regular, its unit tangent vector at $t \in [-\pi, \pi]$ is $(-\sin 2t, \cos 2t)$, and has the same range as $\boldsymbol{\gamma}$.

Path-connected and Convex Subsets

A subset D of $\mathbb{R}^m$ is called **path-connected** if for all $\boldsymbol{u}, \boldsymbol{v} \in D$, there is a path $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ such that $\boldsymbol{\gamma}(\alpha) = \boldsymbol{u}$, $\boldsymbol{\gamma}(\beta) = \boldsymbol{v}$ and $\boldsymbol{\gamma}(t) \in D$ for all $t \in (\alpha, \beta)$.

A subset D of $\mathbb{R}^m$ is called **convex** if the line segment joining any two points in D lies in D. A convex set is path-connected since for $\boldsymbol{u}, \boldsymbol{v} \in D$, the straight-line path $\boldsymbol{\gamma} : [0, 1] \to \mathbb{R}^m$ defined by $\boldsymbol{\gamma}(t) := \boldsymbol{u} + t(\boldsymbol{v} - \boldsymbol{u})$ for $t \in [0, 1]$, lies in D.

Examples:

(i) The subset $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ of $\mathbb{R}^2$ is path-connected; in fact it is convex.

(ii) The subset $\{(x, y) \in \mathbb{R}^2 : 1/2 \leq x^2 + y^2 \leq 1\}$ of $\mathbb{R}^2$ is path-connected, but it is not convex.

(iii) The subset $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\} \cup \{(2, 0)\}$ of $\mathbb{R}^2$ is not path-connected.

Arc length

Let $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_m) : [\alpha, \beta] \to \mathbb{R}^m$ be a piecewise smooth parametrized curve. We define the **arc length** of $\boldsymbol{\gamma}$ by

$$\ell(\boldsymbol{\gamma}) := \int_\alpha^\beta \|\boldsymbol{\gamma}'(t)\| dt = \int_\alpha^\beta \sqrt{\gamma_1'(t)^2 + \cdots + \gamma_m'(t)^2} dt.$$

This definition agrees with our earlier definition if $m = 2, 3$.

Example

Let $\boldsymbol{\gamma}$ denote the standard parametrized circle in $\mathbb{R}^2$ of radius a. Then $\ell(\boldsymbol{\gamma}) = \int_{-\pi}^{\pi} \sqrt{a^2 \sin^2 t + a^2 \cos^2 t}\, dt = 2\pi a$.

Next, let $\widetilde{\boldsymbol{\gamma}}(t) := (a \cos 2t, a \sin 2t)$ for $t \in [-\pi, \pi]$. Then $\ell(\widetilde{\boldsymbol{\gamma}}) = \int_{-\pi}^{\pi} \sqrt{4a^2 \sin^2 2t + 4a^2 \cos^2 2t}\, dt = 4\pi a$.

Note that $\widetilde{\boldsymbol{\gamma}}([-\pi, \pi]) = \boldsymbol{\gamma}([-\pi, \pi])$, but $\ell(\widetilde{\boldsymbol{\gamma}}) = 2\ell(\boldsymbol{\gamma})$.

Line Integral of a Scalar Field

Let $m \in \mathbb{N}$. Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a (piecewise smooth) path, and let $C := \boldsymbol{\gamma}([\alpha, \beta])$. Let $f : C \to \mathbb{R}$ be a bounded scalar field. We define the **line integral** of f **along** $\boldsymbol{\gamma}$ by

$$\int_{\boldsymbol{\gamma}} f\, ds := \int_{\alpha}^{\beta} f(\boldsymbol{\gamma}(t)) \|\boldsymbol{\gamma}'(t)\| dt,$$

provided the Riemann integral on the right side exists. In particular, if f is continuous, then $\int_{\boldsymbol{\gamma}} f\, ds$ is well-defined.

Letting $f := 1$ on C, we see that $\int_{\boldsymbol{\gamma}} ds = \ell(\boldsymbol{\gamma})$.

Example: Let $f(x, y) := x^2 + y^2$ for $(x, y) \in \mathbb{R}^2$. If $\boldsymbol{\gamma}$ is the standard parametrized circle of radius a, then

$$\int_{\boldsymbol{\gamma}} f\, ds = \int_{-\pi}^{\pi} (a^2 \cos^2 t + a^2 \sin^2 t) a\, dt = 2\pi a^3.$$

The algebraic and the order properties of the Riemann integral imply similar properties of the line integral of a scalar field.

Next, let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a (piecewise smooth) path, and consider $t_0 \in (\alpha, \beta)$. Define $\boldsymbol{\gamma}_\ell : [\alpha, t_0] \to \mathbb{R}^m$ by $\boldsymbol{\gamma}_\ell(t) := \boldsymbol{\gamma}(t)$ for all $t \in [\alpha, t_0]$ and $\boldsymbol{\gamma}_r : [t_0, \beta] \to \mathbb{R}^m$ by $\boldsymbol{\gamma}_r(t) := \boldsymbol{\gamma}(t)$ for all $t \in [t_0, \beta]$. Then we write $\boldsymbol{\gamma} = \boldsymbol{\gamma}_\ell + \boldsymbol{\gamma}_r$.

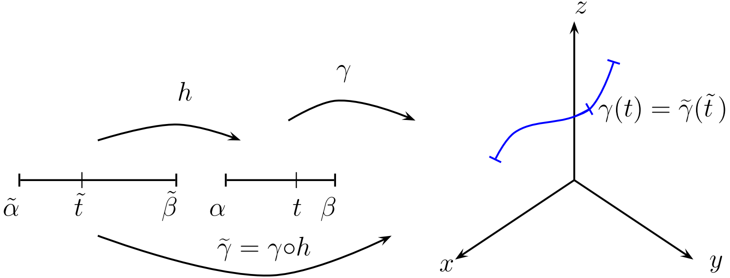
By the Domain Additivity of the Riemann integral, we obtain

$$\int_{\boldsymbol{\gamma}_\ell + \boldsymbol{\gamma}_r} f \, ds = \int_{\boldsymbol{\gamma}_\ell} f \, ds + \int_{\boldsymbol{\gamma}_r} f \, ds$$

for every continuous function $f : \boldsymbol{\gamma}([\alpha, \beta]) \to \mathbb{R}$.

Reparametrization of a Curve

Let $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_m) : [\alpha, \beta] \to \mathbb{R}^m$ be a smooth parametrized curve. Let $h : [\tilde{\alpha}, \tilde{\beta}] \to \mathbb{R}$ be continuously differentiable, let $h'(\tilde{t}) \neq 0$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$, and let $h([\tilde{\alpha}, \tilde{\beta}]) = [\alpha, \beta]$. Then the smooth path $\widetilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ h$ is called a reparametrization of $\boldsymbol{\gamma}$.



Proposition

The line integral of a continuous scalar field along a smooth path is invariant under reparametrization. In particular, the length of a smooth path is invariant under reparametrization.

Proof. Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a smooth path, and let f be a continuous scalar field on $C := \boldsymbol{\gamma}([\alpha, \beta])$.

Let $\widetilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ h$ be a reparametrization of $\boldsymbol{\gamma}$. By the chain rule, $\widetilde{\boldsymbol{\gamma}}'(\tilde{t}) = \big(\gamma_1'(h(\tilde{t}))h'(\tilde{t}), \ldots, \gamma_m'(h(\tilde{t}))h'(\tilde{t})\big)$, which is equal to $\big(\gamma_1'(h(\tilde{t})), \ldots, \gamma_m'(h(\tilde{t})\big)h'(\tilde{t})$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$.

Then our result on Integration by Substitution gives

$$\begin{aligned}
\int_{\widetilde{\boldsymbol{\gamma}}} f\, ds &= \int_{\tilde{\alpha}}^{\tilde{\beta}} f(\widetilde{\boldsymbol{\gamma}}(\tilde{t})) \|\widetilde{\boldsymbol{\gamma}}'(\tilde{t})\| d\tilde{t} \\
&= \int_{\tilde{\alpha}}^{\tilde{\beta}} f(\boldsymbol{\gamma}(h(\tilde{t}))) \|\boldsymbol{\gamma}'(h(\tilde{t}))\| |h'(\tilde{t})| d\tilde{t} \\
&= \int_{\alpha}^{\beta} f(\boldsymbol{\gamma}(t)) \|\boldsymbol{\gamma}'(t)\| dt = \int_{\boldsymbol{\gamma}} f\, ds.
\end{aligned}$$

Considering $f := 1$, we see that the length of a smooth path is invariant under reparametrization. $\square$

Examples: Define $\gamma(t) := (\cos t, \sin t)$ for $t \in [-\pi, \pi]$.
If $\widetilde{\gamma}(\tilde{t}) := (\cos \tilde{t}, -\sin \tilde{t})$ for $\tilde{t} \in [-\pi, \pi]$, then $\widetilde{\gamma}$ is a reparametrization of γ since $\widetilde{\gamma}(\tilde{t}) = \gamma(h(\tilde{t}))$, where $h(\tilde{t}) := -\tilde{t}$ for $\tilde{t} \in [-\pi, \pi]$.

Similarly, if $\widetilde{\gamma}(\tilde{t}) := (\cos 2\tilde{t}, \sin 2\tilde{t})$ for $\tilde{t} \in [-\pi/2, \pi/2]$, then $\widetilde{\gamma}$ is a reparametrization of γ since $\widetilde{\gamma}(\tilde{t}) = \gamma(h(\tilde{t}))$, where $h(\tilde{t}) := 2\tilde{t}$ for $\tilde{t} \in [-\pi/2, \pi/2]$.

Hence in both these cases, $\int_{\widetilde{\gamma}} f \, ds = \int_{\gamma} f \, ds$ for each continuous function on $C := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

On the other hand, if $\widetilde{\gamma}(\tilde{t}) = (\cos 2\tilde{t}, \sin 2\tilde{t})$ for $\tilde{t} \in [-\pi, \pi]$, then $\widetilde{\gamma}$ cannot be a reparametrization of γ since $\ell(\widetilde{\gamma}) = 4\pi \neq 2\pi = \ell(\gamma)$. Note: If we let $h(\tilde{t}) := 2\tilde{t}$ for $\tilde{t} \in [-\pi, \pi]$, then $h([-\pi, \pi]) = [-2\pi, 2\pi] \neq [-\pi, \pi]$.

MA 105 : Calculus
Division 3, Lecture 30

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Vector algebra
- Notions of cross product, scalar triple product and vector triple product
- Scalar fields and vector fields
- Smooth scalar fields and vector fields
- Gradient, divergence and curl
- The GCD sequence
- Laplacian of a smooth scalar field
- Paths or parametrized curves
- Notions of closed/simple/smooth/regular/piecewise smooth/piecewise regular paths. Examples
- Path-connected and convex subsets
- Arc length
- Line integrals of scalar fields along piecewise smooth paths

Line Integral of a Scalar Field

Let $m \in \mathbb{N}$. Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a (piecewise smooth) path, and let $C := \boldsymbol{\gamma}([\alpha, \beta])$. Let $f : C \to \mathbb{R}$ be a bounded scalar field. We define the **line integral** of f **along** $\boldsymbol{\gamma}$ by

$$\int_{\boldsymbol{\gamma}} f\, ds := \int_{\alpha}^{\beta} f(\boldsymbol{\gamma}(t)) \|\boldsymbol{\gamma}'(t)\| dt,$$

provided the Riemann integral on the right side exists. In particular, if f is continuous, then $\int_{\boldsymbol{\gamma}} f\, ds$ is well-defined.

Letting $f := 1$ on C, we see that $\int_{\boldsymbol{\gamma}} ds = \ell(\boldsymbol{\gamma})$.

Example: Let $f(x, y) := x^2 + y^2$ for $(x, y) \in \mathbb{R}^2$. If $\boldsymbol{\gamma}$ is the standard parametrized circle of radius a, then

$$\int_{\boldsymbol{\gamma}} f\, ds = \int_{-\pi}^{\pi} (a^2 \cos^2 t + a^2 \sin^2 t) a\, dt = 2\pi a^3.$$

The algebraic and the order properties of the Riemann integral imply similar properties of the line integral of a scalar field.

Next, let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a (piecewise smooth) path, and consider $t_0 \in (\alpha, \beta)$. Define $\boldsymbol{\gamma}_\ell : [\alpha, t_0] \to \mathbb{R}^m$ by $\boldsymbol{\gamma}_\ell(t) := \boldsymbol{\gamma}(t)$ for all $t \in [\alpha, t_0]$ and $\boldsymbol{\gamma}_r : [t_0, \beta] \to \mathbb{R}^m$ by $\boldsymbol{\gamma}_r(t) := \boldsymbol{\gamma}(t)$ for all $t \in [t_0, \beta]$. Then we write $\boldsymbol{\gamma} = \boldsymbol{\gamma}_\ell + \boldsymbol{\gamma}_r$.

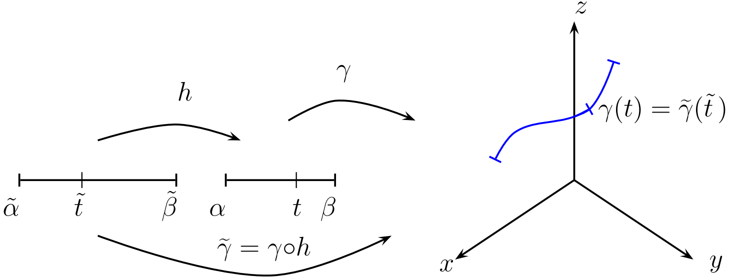
By the Domain Additivity of the Riemann integral, we obtain

$$\int_{\boldsymbol{\gamma}_\ell + \boldsymbol{\gamma}_r} f\, ds = \int_{\boldsymbol{\gamma}_\ell} f\, ds + \int_{\boldsymbol{\gamma}_r} f\, ds$$

for every continuous function $f : \boldsymbol{\gamma}([\alpha, \beta]) \to \mathbb{R}$.

Reparametrization of a Curve

Let $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_m) : [\alpha, \beta] \to \mathbb{R}^m$ be a smooth parametrized curve. Let $h : [\tilde{\alpha}, \tilde{\beta}] \to \mathbb{R}$ be continuously differentiable, let $h'(\tilde{t}) \neq 0$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$, and let $h([\tilde{\alpha}, \tilde{\beta}]) = [\alpha, \beta]$. Then the smooth path $\widetilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ h$ is called a reparametrization of $\boldsymbol{\gamma}$.



Proposition

The line integral of a continuous scalar field along a smooth path is invariant under reparametrization. In particular, the length of a smooth path is invariant under reparametrization.

Proof. Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a smooth path, and let f be a continuous scalar field on $C := \boldsymbol{\gamma}([\alpha, \beta])$.

Let $\widetilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ h$ be a reparametrization of $\boldsymbol{\gamma}$. By the chain rule, $\widetilde{\boldsymbol{\gamma}}'(\tilde{t}) = \big(\gamma_1'(h(\tilde{t}))h'(\tilde{t}), \ldots, \gamma_m'(h(\tilde{t}))h'(\tilde{t})\big)$, which is equal to $\big(\gamma_1'(h(\tilde{t})), \ldots, \gamma_m'(h(\tilde{t})\big) h'(\tilde{t})$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$.

Then our result on Integration by Substitution gives

$$\begin{aligned}
\int_{\widetilde{\boldsymbol{\gamma}}} f\, ds &= \int_{\tilde{\alpha}}^{\tilde{\beta}} f(\widetilde{\boldsymbol{\gamma}}(\tilde{t})) \|\widetilde{\boldsymbol{\gamma}}'(\tilde{t})\| d\tilde{t} \\
&= \int_{\tilde{\alpha}}^{\tilde{\beta}} f(\boldsymbol{\gamma}(h(\tilde{t}))) \|\boldsymbol{\gamma}'(h(\tilde{t}))\| |h'(\tilde{t})| d\tilde{t} \\
&= \int_{\alpha}^{\beta} f(\boldsymbol{\gamma}(t)) \|\boldsymbol{\gamma}'(t)\| dt = \int_{\boldsymbol{\gamma}} f\, ds.
\end{aligned}$$

Considering $f := 1$, we see that the length of a smooth path is invariant under reparametrization. □

Examples: Define $\gamma(t) := (\cos t, \sin t)$ for $t \in [-\pi, \pi]$.
If $\widetilde{\gamma}(\tilde{t}) := (\cos \tilde{t}, -\sin \tilde{t})$ for $\tilde{t} \in [-\pi, \pi]$, then $\widetilde{\gamma}$ is a reparametrization of γ since $\widetilde{\gamma}(\tilde{t}) = \gamma(h(\tilde{t}))$, where $h(\tilde{t}) := -\tilde{t}$ for $\tilde{t} \in [-\pi, \pi]$.

Similarly, if $\widetilde{\gamma}(\tilde{t}) := (\cos 2\tilde{t}, \sin 2\tilde{t})$ for $\tilde{t} \in [-\pi/2, \pi/2]$, then $\widetilde{\gamma}$ is a reparametrization of γ since $\widetilde{\gamma}(\tilde{t}) = \gamma(h(\tilde{t}))$, where $h(\tilde{t}) := 2\tilde{t}$ for $\tilde{t} \in [-\pi/2, \pi/2]$.

Hence in both these cases, $\int_{\widetilde{\gamma}} f\, ds = \int_{\gamma} f\, ds$ for each continuous function on $C := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

On the other hand, if $\widetilde{\gamma}(\tilde{t}) = (\cos 2\tilde{t}, \sin 2\tilde{t})$ for $\tilde{t} \in [-\pi, \pi]$, then $\widetilde{\gamma}$ cannot be a reparametrization of γ since $\ell(\widetilde{\gamma}) = 4\pi \neq 2\pi = \ell(\gamma)$. Note: If we let $h(\tilde{t}) := 2\tilde{t}$ for $\tilde{t} \in [-\pi, \pi]$, then $h([-\pi, \pi]) = [-2\pi, 2\pi] \neq [-\pi, \pi]$.

We now give an intrinsic parametrization of a regular path $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$. For $t \in [\alpha, \beta]$, let

$$s(t) := \int_\alpha^t \|\boldsymbol{\gamma}'(u)\| du.$$

Then $s(\alpha) = 0$, $s(\beta) = \ell(\boldsymbol{\gamma})$, and $s'(t) = \|\boldsymbol{\gamma}'(t)\|$ for all $t \in [\alpha, \beta]$ by the FTC (Part I). Since $\boldsymbol{\gamma}$ is a regular path, $s'(t) > 0$ for all $t \in [\alpha, \beta]$, and so $t \longmapsto s(t)$ is a strictlty increasing differentiable function from $[\alpha, \beta]$ onto $[0, \ell(\boldsymbol{\gamma})]$. Let $h : [0, \ell(\boldsymbol{\gamma})] \to [\alpha, \beta]$ denote its inverse. Then h is differentiable and its derivative does not vanish on $[0, \ell(\boldsymbol{\gamma})]$. Define $\widetilde{\boldsymbol{\gamma}}(s) := \boldsymbol{\gamma}(h(s))$ for $s \in [0, \ell(\boldsymbol{\gamma})]$.

Then $\widetilde{\boldsymbol{\gamma}}$ is a reparametrization of the regular path $\boldsymbol{\gamma}$; it is known as the **arc length parametrization**, and s is called the **arc length parameter**.

Example: Let $\boldsymbol{\gamma}$ denote the standard parametrized circle in $\mathbb{R}^2$ of radius a. Then

$$s(t) = \int_{-\pi}^{t} \|(-a\sin t, a\cos t)\| dt = a(t+\pi) \quad \text{for } t \in [-\pi, \pi].$$

The inverse of the function $t \longmapsto a(t+\pi)$ is given by the function $s \longmapsto (s/a) - \pi$ for $s \in [0, 2\pi a]$. Hence the arc length parametrization of $\boldsymbol{\gamma}$ is given by

$$\widetilde{\boldsymbol{\gamma}}(s) = \left(a\cos\left(\frac{s}{a} - \pi\right), \, a\sin\left(\frac{s}{a} - \pi\right)\right) = \left(-a\cos\frac{s}{a}, \, a\sin\frac{s}{a}\right)$$

for $s \in [0, 2\pi a]$.

In view of the equality $\dfrac{ds}{dt} = s'(t) = \|\boldsymbol{\gamma}'(t)\|$ for $t \in [\alpha, \beta]$ stated earlier, we introduce the following differential notation.

For a regular path $\boldsymbol{\gamma}$, we let $ds := \|\boldsymbol{\gamma}'(t)\| dt$. (Compare our definition of $\int_{\boldsymbol{\gamma}} f \, ds := \int_{\alpha}^{\beta} f(\boldsymbol{\gamma}(t))\|\boldsymbol{\gamma}'(t)\| dt$.) We note that

$$\boldsymbol{\gamma}'(t) = \frac{d\boldsymbol{\gamma}}{dt} = \frac{d\boldsymbol{\gamma}}{ds}\frac{ds}{dt} = \frac{d\boldsymbol{\gamma}}{ds}\|\boldsymbol{\gamma}'(t)\|, \quad \text{and so} \quad \hat{\boldsymbol{t}} = \frac{\boldsymbol{\gamma}'(t)}{\|\boldsymbol{\gamma}'(t)\|} = \frac{d\boldsymbol{\gamma}}{ds}.$$

Geometric Curve

Let $[\alpha, \beta]$ be an interval in $\mathbb{R}$, and let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a (piecewise smooth) path. Then $C := \boldsymbol{\gamma}([\alpha, \beta])$ is called a **geometric curve**; $\boldsymbol{\gamma}$ is said to give a **parametrization** of C.

Suppose $\boldsymbol{\gamma}$ is simple. In view of the reparametrization result for continuous scalar fields on paths proved above, we define

$$\int_C f\, ds := \int_{\boldsymbol{\gamma}} f\, ds = \int_\alpha^\beta f(\boldsymbol{\gamma}(t))\|\boldsymbol{\gamma}'(t)\|dt$$

for a continuous scalar field f on C. In particular, we write

$$\ell(C) := \ell(\boldsymbol{\gamma}) = \int_\alpha^\beta \|\boldsymbol{\gamma}'(t)\|dt.$$

Line Integral of a Vector Field

Let $m \in \mathbb{N}$. Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a (piecewise smooth) path, and let $C := \boldsymbol{\gamma}([\alpha, \beta])$. Let $\boldsymbol{F} : C \to \mathbb{R}^m$ be a bounded vector field. We define the **line integral** of $\boldsymbol{F}$ **along** $\boldsymbol{\gamma}$ by

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} := \int_{\alpha}^{\beta} \boldsymbol{F}(\boldsymbol{\gamma}(t)) \cdot \boldsymbol{\gamma}'(t) dt,$$

provided the Riemann integral on the right side exists. In particular, if $\boldsymbol{F}$ is continuous, that is, if $\boldsymbol{F} = (F_1, \ldots, F_m)$ and each $F_i : C \to \mathbb{R}$ is continuous, then $\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s}$ is well-defined.

In analogy with our differential notation $ds := \|\boldsymbol{\gamma}'(t)\| dt$, we let $d\boldsymbol{s} := \boldsymbol{\gamma}'(t) dt$.

Thus if $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_m)$, then $d\boldsymbol{s} := (\gamma_1'(t) dt, \ldots, \gamma_m'(t) dt)$.

Let $m = 3$, and $\boldsymbol{\gamma}(t) := (x(t), y(t), z(t))$ for $t \in [\alpha, \beta]$. Then

$$\boldsymbol{\gamma}'(t) := (x'(t), y'(t), z'(t)) \quad \text{for } t \in [\alpha, \beta].$$

Let $\boldsymbol{F} := (P, Q, R)$. Then we write the line integral

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\alpha}^{\beta} \left(P \frac{dx}{dt} + Q \frac{dy}{dt} + R \frac{dz}{dt} \right) dt$$

as $\displaystyle\int_{\boldsymbol{\gamma}} P\,dx + Q\,dy + R\,dz$. If the path $\boldsymbol{\gamma}$ is regular, then

$$d\boldsymbol{s} = \boldsymbol{\gamma}'(t)dt = \hat{\boldsymbol{t}}\, \|\boldsymbol{\gamma}'(t)\| dt = \hat{\boldsymbol{t}}\, ds,$$

and so we may write the line integral $\displaystyle\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s}$ as $\displaystyle\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot \hat{\boldsymbol{t}}\, ds$.

The algebraic properties of the Riemann integral imply similar properties of the line integral of a vector field. Also, as for the line integral of a scalar field, $\int_{\boldsymbol{\gamma}_\ell + \boldsymbol{\gamma}_r} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}_\ell} \boldsymbol{F} \cdot d\boldsymbol{s} + \int_{\boldsymbol{\gamma}_r} \boldsymbol{F} \cdot d\boldsymbol{s}$.

Examples:

(i) Let $\gamma(t) := (\cos t, \sin t, \cos t \sin t) \in \mathbb{R}^3$ for $t \in [-\pi, \pi]$, and $\boldsymbol{F}(x, y, z) := (y, z, x)$ for $(x, y, z) \in \mathbb{R}^3$. Noting that $\cos t \sin t = (\sin 2t)/2$ for $t \in [-\pi, \pi]$, we obtain

$$\int_\gamma \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}} y\, dx + z\, dy + x\, dz$$

$$= \int_{-\pi}^{\pi} \Big((\sin t)(-\sin t) + \frac{1}{2}(\sin 2t)(\cos t) + (\cos t)(\cos 2t) \Big)\, dt$$

$$= -\pi + 0 + 0 = -\pi.$$

(ii) Let $\gamma(t) := (t, 2t^2)$ for $t \in [1, 2]$, and $\boldsymbol{F}(x, y) := (x\, y, y^2)$ for $(x, y) \in \mathbb{R}^2$. Then

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_1^2 (2t^3, 4t^4) \cdot (1, 4t) dt = \int_1^2 (2t^3 + 16t^5) dt = \frac{351}{2}.$$

On the other hand, if $\widetilde{\gamma}(t) := (-t, 2t^2)$ for $t \in [-2, -1]$, then

$$\int_{\widetilde{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{-2}^{-1} (-2t^3, 4t^4) \cdot (-1, 4t) dt = \int_{-2}^{-1} (2t^3 + 16t^5) dt = -\frac{351}{2}.$$

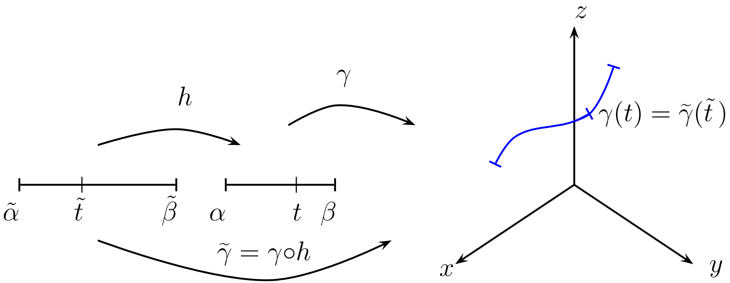
In the example (ii) above, the path $\widetilde{\gamma}$ is a reparametrization of the path γ, and the line integral of the vector field $\boldsymbol{F}$ along $\widetilde{\gamma}$ is the negative of the line integral of $\boldsymbol{F}$ along γ.

Thus line integrals of a vector field are not, in general, invariant under reparametrization. However, they can differ only in sign, as we shall see in the next proposition.

Proposition

The line integral of a continuous vector field along a smooth path is invariant under a reparametrization only up to its sign.

Proof. Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a smooth path, and let $\boldsymbol{F}$ be a continuous vector field on $C := \boldsymbol{\gamma}([\alpha, \beta])$. Suppose $\widetilde{\boldsymbol{\gamma}} := \boldsymbol{\gamma} \circ h$ is a reparametrization of $\boldsymbol{\gamma}$.



As before, by the chain rule,
$\widetilde{\boldsymbol{\gamma}}'(\tilde{t}) = \big(\gamma_1'(h(\tilde{t})), \ldots, \gamma_m'(h(\tilde{t})\big)\, h'(\tilde{t})$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$, and so

$$\begin{aligned}\int_{\widetilde{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} &= \int_{\tilde{\alpha}}^{\tilde{\beta}} \boldsymbol{F}(\widetilde{\gamma}(\tilde{t})) \cdot \widetilde{\gamma}'(\tilde{t})\, d\tilde{t} \\ &= \int_{\tilde{\alpha}}^{\tilde{\beta}} \boldsymbol{F}(\gamma(h(\tilde{t}))) \cdot \gamma'(h(\tilde{t}))\, h'(\tilde{t})\, d\tilde{t}.\end{aligned}$$

Since $h'(\tilde{t}) \neq 0$ for $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$ and h' is continuous on $[\tilde{\alpha}, \tilde{\beta}]$, either $h'(\tilde{t}) > 0$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$, or $h'(\tilde{t}) < 0$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$.

In the former case, $h'(\tilde{t}) = |h'(\tilde{t})|$, and so our result on Integration by Substitution shows that $\int_{\widetilde{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\gamma} \boldsymbol{F} \cdot d\boldsymbol{s}$.

In the latter case, $h'(\tilde{t}) = -|h'(\tilde{t})|$, and so our result on Integration by Substitution shows that

$$\int_{\widetilde{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = -\int_{\gamma} \boldsymbol{F} \cdot d\boldsymbol{s}. \qquad \square$$

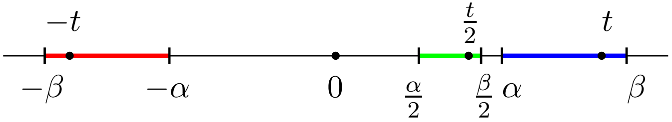
Examples Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a path, and let $\boldsymbol{F} : \boldsymbol{\gamma}([\alpha, \beta]) \to \mathbb{R}^m$ be a continuous vector field.

(i) Let $\tilde{\alpha} := -\beta$, $\tilde{\beta} := -\alpha$, and define $h(\tilde{t}) := -\tilde{t}$, $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$. Then $\widetilde{\boldsymbol{\gamma}}(\tilde{t}) := \boldsymbol{\gamma}(-\tilde{t})$, $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$. Since $h'(\tilde{t}) = -1 < 0$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$, we obtain

$$\int_{\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s} = -\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s}.$$

(ii) Let $\tilde{\alpha} := \alpha/2$, $\tilde{\beta} := \beta/2$, and define $h(\tilde{t}) := 2\tilde{t}$, $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$. Then $\widetilde{\boldsymbol{\gamma}}(\tilde{t}) := \boldsymbol{\gamma}(h(\tilde{t})) = \boldsymbol{\gamma}(2\tilde{t})$ for $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$. Since $h'(\tilde{t}) = 2 > 0$ for all $\tilde{t} \in [\tilde{\alpha}, \tilde{\beta}]$, we obtain

$$\int_{\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s}.$$



Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a path. Define $\tilde{\boldsymbol{\gamma}} : [-\beta, -\alpha] \to \mathbb{R}^m$ by $\widetilde{\boldsymbol{\gamma}}(\tilde{t}) := \boldsymbol{\gamma}(-\tilde{t})$ for $\tilde{t} \in [-\beta, -\alpha]$. Then the path $\tilde{\boldsymbol{\gamma}}$ is called the **negative** of $\boldsymbol{\gamma}$, and it is denoted by $-\boldsymbol{\gamma}$. Clearly, $-(-\boldsymbol{\gamma}) = \boldsymbol{\gamma}$.

Note that $(-\boldsymbol{\gamma})([-\beta, -\alpha]) = \boldsymbol{\gamma}([\alpha, \beta])$.

Also, $(-\boldsymbol{\gamma})(-\beta) = \boldsymbol{\gamma}(\beta)$ and $(-\boldsymbol{\gamma})(-\alpha) = \boldsymbol{\gamma}(\alpha)$, that is, the initial point of $-\boldsymbol{\gamma}$ is the final point of $\boldsymbol{\gamma}$, and the final point of $-\boldsymbol{\gamma}$ is the initial point of $\boldsymbol{\gamma}$.

Note: If $h(\tilde{t}) = -\tilde{t}$ for all $\tilde{t} \in [-\beta, -\alpha]$, then $h' = -1$.
If $\boldsymbol{F}$ is a continuous vector field on $\boldsymbol{\gamma}([\alpha, \beta])$, then we obtain

$$\int_{-\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = -\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s}.$$

Physical interpretation of the line integral of a vector field:

Suppose a vector field $\boldsymbol{F}$ is defined on $D \subset \mathbb{R}^m$, and $\boldsymbol{\gamma}$ is a regular path. If $C := \gamma([\alpha, \beta]) \subset D$, then $\boldsymbol{F} \cdot \hat{\boldsymbol{t}}$ is the component of $\boldsymbol{F}$ in the direction of the unit tangent vector $\hat{\boldsymbol{t}}$.

If $\boldsymbol{F}$ is a force field in the domain D, then the line integral

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot \hat{\boldsymbol{t}}\, ds$$

represents the work done by the force $\boldsymbol{F}$ along the path $\boldsymbol{\gamma}$.

Similarly, if $\boldsymbol{F}$ is the velocity field of a moving fluid in the domain D, then the line integral $\displaystyle\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s}$ represents the flow of the fluid along the path $\boldsymbol{\gamma}$; in case $\boldsymbol{\gamma}$ is a closed path, the line integral represents the circulation of the fluid along $\boldsymbol{\gamma}$.

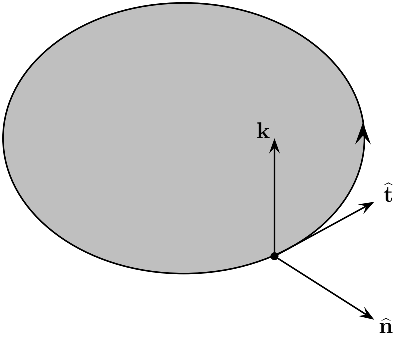
Flux Integral

There is a variant of the circulation known as the flux of a moving fluid. We shall now consider its version in $\mathbb{R}^2$.

Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^2$ be a simple closed regular path. Let D be an open subset of $\mathbb{R}^2$ containing $C := \boldsymbol{\gamma}([\alpha, \beta])$. Let $\hat{\boldsymbol{n}}$ denote the (continuous) **outward unit normal** to $\boldsymbol{\gamma}$. If $\boldsymbol{F}$ is a continuous vector field on D, then

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot \hat{\boldsymbol{n}} \, ds$$

is called the **flux integral** of $\boldsymbol{F}$ along $\boldsymbol{\gamma}$. It represents the flux of a velocity field $\boldsymbol{F}$ along a simple closed regular path $\boldsymbol{\gamma}$.



Suppose $\gamma(t)$ moves in the anti-clockwise direction (as seen from high above) as the parameter t goes from α to β. Then

$$\hat{\boldsymbol{n}} = \hat{\boldsymbol{t}} \times \boldsymbol{k} = \frac{d\boldsymbol{\gamma}}{ds} \times \boldsymbol{k} = \left(\frac{dx}{ds}\,\boldsymbol{i} + \frac{dy}{ds}\,\boldsymbol{j} \right) \times \boldsymbol{k} = \frac{dy}{ds}\,\boldsymbol{i} - \frac{dx}{ds}\,\boldsymbol{j}.$$

Hence if $\boldsymbol{F} := P\,\boldsymbol{i} + Q\,\boldsymbol{j}$, then the flux integral of $\boldsymbol{F}$ along $\boldsymbol{\gamma}$ is

$$\int_{\gamma} \boldsymbol{F} \cdot \hat{\boldsymbol{n}}\, ds = \int_{\gamma} \left(P\,\frac{dy}{ds} - Q\,\frac{dx}{ds} \right) ds = \int_{\gamma} P\,dy - Q\,dx.$$

Example

Let γ denote the standard unit circle in $\mathbb{R}^2$ of radius 1, and let $\boldsymbol{F} := (P, Q)$, where $P(x, y) := x - y$ and $Q(x, y) := x$ for $(x, y) \in \mathbb{R}^2$.

Then the circulation of $\boldsymbol{F}$ along γ is equal to $\int_\gamma \boldsymbol{F} \cdot d\boldsymbol{s}$, that is,

$$\int_\gamma P\,dx + Q\,dy = \int_{-\pi}^{\pi} \big((\cos t - \sin t)(-\sin t) + (\cos t)(\cos t)\big)\,dt = 2\pi.$$

Also, the flux of $\boldsymbol{F}$ along γ is equal to $\int_\gamma \boldsymbol{F} \cdot \hat{\boldsymbol{n}}\,ds$, that is,

$$\int_\gamma P\,dy - Q\,dx = \int_{-\pi}^{\pi} \big((\cos t - \sin t)(\cos t) - (\cos t)(-\sin t)\big)\,dt = \pi.$$

Path Independence of Line Integrals

Let $m \in \mathbb{N}$, $D \subset \mathbb{R}^m$, and consider a path $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$. We say that $\boldsymbol{\gamma}$ **lies in** D if $\boldsymbol{\gamma}([\alpha, \beta]) \subset D$. Also, we say that $\boldsymbol{\gamma}(\alpha)$ is the **initial point** and $\boldsymbol{\gamma}(\beta)$ is the **final point** of $\boldsymbol{\gamma}$.

Let $\boldsymbol{F} : D \to \mathbb{R}^m$ be a continuous vector field. We say that line integrals of $\boldsymbol{F}$ are **path-independent** in D if

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s}$$

for all (piecewise smooth) paths $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ lying in D which have the same initial point as well as the same final point.

Note: The path independence of line integrals of $\boldsymbol{F}$ in D does not mean that $\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s}$ for any smooth paths $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ in D. It only says that this equality holds if the smooth paths $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ lie in D, and both have the same initial point, and both have the same final point.

Examples:

(i) Let $\boldsymbol{F}(x,y) := (x+y, x)$ for $(x,y) \in \mathbb{R}^2$. Consider a smooth path $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^2$. If $\boldsymbol{\gamma}(t) := (x(t), y(t))$ for $t \in [\alpha, \beta]$, then

$$\begin{aligned}\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} &= \int_\alpha^\beta (x(t)+y(t), x(t)) \cdot (x'(t), y'(t))dt \\ &= \int_\alpha^\beta \big(x(t)x'(t) + (xy)'(t)\big)\, dt \\ &= \frac{1}{2}x^2(\beta) - \frac{1}{2}x^2(\alpha) + x(\beta)y(\beta) - x(\alpha)y(\alpha).\end{aligned}$$

Thus if smooth paths $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ have the same initial point, and also the same final point, then $\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s}$. Hence line integrals of $\boldsymbol{F}$ are path-independent in $\mathbb{R}^2$.

(ii) Let $\boldsymbol{F}(x,y) := (x-y, x)$ for $(x,y) \in \mathbb{R}^2$. Consider $\boldsymbol{\gamma}(t) := (t, 0)$ for $t \in [-1, 1]$, and $\widetilde{\boldsymbol{\gamma}}(t) := (-\cos t, \sin t)$ for $t \in [0, \pi]$. The initial point of both $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ is $(-1, 0)$

and the final point of both γ and $\widetilde{\gamma}$ is $(1, 0)$. However,

$$\int_\gamma \boldsymbol{F} \cdot d\boldsymbol{s} \;=\; \int_{-1}^{1} (t - 0, t) \cdot (1, 0) dt = \int_{-1}^{1} t \, dt = 0,$$

whereas

$$\begin{aligned}
\int_{\widetilde{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} &= \int_0^\pi (-\cos t - \sin t, -\cos t) \cdot (\sin t, \cos t) dt \\
&= \int_0^\pi (-\cos t \sin t - \sin^2 t - \cos^2 t) dt \\
&= -\frac{1}{2} \int_0^\pi \sin 2t dt - \int_0^\pi dt = 0 - \pi = -\pi.
\end{aligned}$$

Hence line integrals of $\boldsymbol{F}$ are not path-independent in $\mathbb{R}^2$.

We now give a criterion for the path-independence of line integrals of a vector field.

Proposition

Let $\boldsymbol{F}$ be a continuous vector field on $D \subset \mathbb{R}^m$. Then line integrals of $\boldsymbol{F}$ are path-independent in D $\iff$ the line integral of $\boldsymbol{F}$ along every closed path that lies in D is zero.

$\implies$) Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a closed path that lies in D. Then $\boldsymbol{\gamma}(\alpha) = \boldsymbol{\gamma}(\beta) \in D$. Define $\boldsymbol{\gamma}_0(t) := \boldsymbol{\gamma}(\alpha)$ for all $t \in [\alpha, \beta]$. Then both $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_0$ lie in D, and they have the same initial point $\boldsymbol{\gamma}(\alpha)$ as well as the same final point $\boldsymbol{\gamma}(\alpha)$. By the path-independence of $\boldsymbol{F}$ in D,

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}_0} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\alpha}^{\beta} \boldsymbol{F}(\boldsymbol{\gamma}(\alpha)) \cdot \boldsymbol{0}\, dt = 0.$$

$\impliedby$) Let $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ be paths lying in D which have the same initial point, and also the same final point. Then $\boldsymbol{\gamma}_c := \boldsymbol{\gamma} + (-\widetilde{\boldsymbol{\gamma}})$ is a closed path in D, and so

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} - \int_{\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} + \int_{-\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}_c} \boldsymbol{F} \cdot d\boldsymbol{s} = 0. \quad \square$$

FTC for Line Integrals

Consider the following versions of Part I and Part II of the Fundamental Theorem of Calculus for Riemann integrals of scalar fields on an open interval in $\mathbb{R}$.

Part I: Let I be an open interval in $\mathbb{R}$, and let g be a continuous scalar field on I. If $a \in I$, and we let $f(x) := \int_a^x g(t)dt$ for $x \in I$, then $f' = g$ on I.

Part II: Let I be an open interval in $\mathbb{R}$, and let f be a smooth scalar field on I. Then

$$\int_a^b f'(x)dx = f(b) - f(a) \quad \text{for all } a, b \in I.$$

We shall now attempt to find analogues of the above results for line integrals of vector fields. In this process, we shall obtain another important criterion for the path-independence of line integrals of a vector field.

Theorem (FTC for Line Integrals: Part I)

Let D be an open subset of $\mathbb{R}^m$, and let $\boldsymbol{G}$ be a continuous vector field on D. Suppose line integrals of $\boldsymbol{G}$ are path-independent in D. If D is path-connected and for a fixed $\boldsymbol{x}_0$ in D, we define

$$f(\boldsymbol{x}) := \int_{\gamma_{\boldsymbol{x}}} \boldsymbol{G} \cdot d\boldsymbol{s}, \quad \text{where } \gamma_{\boldsymbol{x}} \text{ is a path in } D \text{ from } \boldsymbol{x}_0 \text{ to } \boldsymbol{x},$$

then $\nabla f = \boldsymbol{G}$ on D. Consequently, $\boldsymbol{G}$ is a gradient field on D.

Proof: Suppose D is path-connected. Then for every $\boldsymbol{x} \in D$, there is a (piecewise smooth) path $\gamma_{\boldsymbol{x}}$ with initial point $\boldsymbol{x}_0$ and final point $\boldsymbol{x}$. Further, as line integrals of $\boldsymbol{G}$ are path-independent on D, the function $f : D \to \mathbb{R}$ mentioned above is well-defined.

For simplicity, let $m = 2$, and consider $\boldsymbol{x} := (x_1, x_2) \in D$.

MA 105 : Calculus
Division 1, Lecture 22

Prof. Sudhir R. Ghorpade
IIT Bombay

Proposition (Criterion for Path Independence)

Let $\boldsymbol{F}$ be a continuous vector field on $D \subset \mathbb{R}^m$. Then line integrals of $\boldsymbol{F}$ are path-independent in $D \iff$ the line integral of $\boldsymbol{F}$ along every closed path that lies in D is zero.

$\Longrightarrow$) Let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a closed path that lies in D. Then $\boldsymbol{\gamma}(\alpha) = \boldsymbol{\gamma}(\beta) \in D$. Define $\boldsymbol{\gamma}_0(t) := \boldsymbol{\gamma}(\alpha)$ for all $t \in [\alpha, \beta]$. Then both $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_0$ lie in D, and they have the same initial point $\boldsymbol{\gamma}(\alpha)$ as well as the same final point $\boldsymbol{\gamma}(\alpha)$. By the path-independence of $\boldsymbol{F}$ in D,

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}_0} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\alpha}^{\beta} \boldsymbol{F}(\boldsymbol{\gamma}(\alpha)) \cdot \mathbf{0}\, dt = 0.$$

$\Longleftarrow$) Let $\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\gamma}}$ be paths lying in D which have the same initial point, and also the same final point. Then $\boldsymbol{\gamma}_c := \boldsymbol{\gamma} + (-\widetilde{\boldsymbol{\gamma}})$ is a closed path in D, and so

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} - \int_{\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} + \int_{-\widetilde{\boldsymbol{\gamma}}} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\boldsymbol{\gamma}_c} \boldsymbol{F} \cdot d\boldsymbol{s} = 0. \qquad \square$$

FTC for Line Integrals

Consider the following versions of Part I and Part II of the Fundamental Theorem of Calculus for Riemann integrals of scalar fields on an open interval in $\mathbb{R}$.

Part I: Let I be an open interval in $\mathbb{R}$, and let g be a continuous scalar field on I. If $a \in I$, and we let $f(x) := \int_a^x g(t)dt$ for $x \in I$, then $f' = g$ on I.

Part II: Let I be an open interval in $\mathbb{R}$, and let f be a smooth scalar field on I. Then

$$\int_a^b f'(x)dx = f(b) - f(a) \quad \text{for all } a, b \in I.$$

We shall now attempt to find analogues of the above results for line integrals of vector fields. In this process, we shall obtain another important criterion for the path-independence of line integrals of a vector field.

Theorem (FTC for Line Integrals: Part I)

Let D be an open subset of $\mathbb{R}^m$, and let $\boldsymbol{G}$ be a continuous vector field on D. Suppose line integrals of $\boldsymbol{G}$ are path-independent in D. If D is path-connected and for a fixed $\boldsymbol{x}_0$ in D, we define

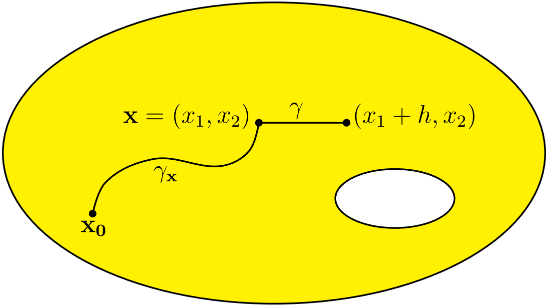
$$f(\boldsymbol{x}) := \int_{\gamma_{\boldsymbol{x}}} \boldsymbol{G} \cdot d\boldsymbol{s}, \quad \text{where } \gamma_{\boldsymbol{x}} \text{ is a path in } D \text{ from } \boldsymbol{x}_0 \text{ to } \boldsymbol{x},$$

then $\nabla f = \boldsymbol{G}$ on D. Consequently, $\boldsymbol{G}$ is a gradient field on D.

Proof: Suppose D is path-connected. Then for every $\boldsymbol{x} \in D$, there is a (piecewise smooth) path $\gamma_{\boldsymbol{x}}$ with initial point $\boldsymbol{x}_0$ and final point $\boldsymbol{x}$. Further, as line integrals of $\boldsymbol{G}$ are path-independent on D, the function $f : D \to \mathbb{R}$ mentioned above is well-defined.

For simplicity, let $m = 2$, and consider $\boldsymbol{x} := (x_1, x_2) \in D$.

Since the set D is open in $\mathbb{R}^2$, the line segment joining (x_1, x_2) to $(x_1 + h, x_2)$ lies in D for all $h \in \mathbb{R}$ if $|h|$ is sufficiently small.



Let $\gamma(t) := (x_1 + th, x_2)$ for $t \in [0, 1]$. Now $\boldsymbol{\gamma_x}$ followed by $\boldsymbol{\gamma}$ is a (piecewise smooth) path from $\boldsymbol{x}_0$ to $(x_1 + h, x_2)$. Hence

$$f(x_1+h, x_2) - f(x_1, x_2) = \int_{\boldsymbol{\gamma}} \boldsymbol{G} \cdot d\boldsymbol{s} = \int_0^1 \boldsymbol{G}(x_1+th, x_2) \cdot (h, 0) dt.$$

Let $\boldsymbol{G} := (G_1, G_2)$. It follows that for all $h \neq 0$ and small $|h|$,

$$\frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h} = \int_0^1 G_1(x_1 + th, x_2)dt.$$

Also, since G_1 is continuous at (x_1, x_2),

$$\int_0^1 G_1(x_1+th, x_2)dt \to \int_0^1 G_1(x_1, x_2)dt = G_1(x_1, x_2) \text{ as } h \to 0.$$

Thus the partial of f at any $(x_1, x_2) \in D$ with respect to the first variable is equal to $G_1(x_1, x_2)$. Similarly, the partial of f at (x_1, x_2) with respect to the second variable equals $G_2(x_1, x_2)$. Hence $\nabla f = (f_{x_1}, f_{x_2}) = (G_1, G_2) = \boldsymbol{G}$.

Now if D is any open subset of $\mathbb{R}^2$, then by considering components of D, each of which is open and path-connected, we see that $\nabla f = \boldsymbol{G}$. $\square$

Theorem (FTC for Line Integrals: Part II)

Let D be an open subset of $\mathbb{R}^m$, and let f be a smooth scalar field on D. Let $\boldsymbol{\gamma}$ be a path that lies in D, A be the initial point of $\boldsymbol{\gamma}$, and B be the final point of $\boldsymbol{\gamma}$. Then

$$\int_{\boldsymbol{\gamma}} (\nabla f)\cdot d\boldsymbol{s} = f(B) - f(A).$$

Consequently, line integrals of a continuous gradient field on D are path-independent in D.

Proof: By part (ii) of the chain rule,

$$\int_{\boldsymbol{\gamma}} (\nabla f)\cdot d\boldsymbol{s} = \int_{\alpha}^{\beta} (\nabla f)(\boldsymbol{\gamma}(t))\cdot \boldsymbol{\gamma}'(t)dt = \int_{\alpha}^{\beta} (f\circ\boldsymbol{\gamma})'(t)dt.$$

So by Part II of the FTC for the function $f\circ\boldsymbol{\gamma} : [\alpha,\beta] \to \mathbb{R}$,

$$\int_{\boldsymbol{\gamma}} \nabla f\cdot d\boldsymbol{s} = f(\boldsymbol{\gamma}(\beta)) - f(\boldsymbol{\gamma}(\alpha)) = f(B) - f(A). \quad \square$$

The previous result shows that if $\boldsymbol{G}$ is a continuous gradient field on an open subset of $\mathbb{R}^m$, then the evaluation of a line integral $\int_\gamma \boldsymbol{G} \cdot d\boldsymbol{s}$ can be simplified on replacing γ by a suitable path having the same initial point and the same final point.

If we actually know a scalar field f such that $\nabla f = \boldsymbol{G}$, then the evaluation of $\int_\gamma \boldsymbol{G} \cdot d\boldsymbol{s}$ is immediate even when the path γ is very complicated. This process is similar to the evaluation of a Riemann integral by finding an antiderivative of its integrand.

Example: Let $\boldsymbol{G}(x, y, z) := (y\,z, z\,x, x\,y)$ for $(x, y, z) \in \mathbb{R}^3$, and let $\gamma(t) := (\cos^4 t, \sin^4 t, \tan^4 t)$, $t \in [0, \pi/4]$. If we let $f(x, y, z) := x\,y\,z$ for $(x, y, z) \in \mathbb{R}^3$, then $\nabla f = \boldsymbol{G}$ on $\mathbb{R}^3$, and

$$\int_\gamma \boldsymbol{G} \cdot d\boldsymbol{s} = f(\gamma(\pi/4)) - f(\gamma(0)) = f\left(\frac{1}{4}, \frac{1}{4}, 1\right) - f(1, 0, 0) = \frac{1}{16}.$$

Parts I and II of the FTC for line integrals show that line integrals of a continuous vector field $\boldsymbol{G}$ on an open subset of $\mathbb{R}^m$ are path-independent if and only if $\boldsymbol{G}$ is a gradient field.

In view of the results proved above, it is important to know whether a continuous vector field $\boldsymbol{F}$ is indeed a gradient field. First we prove a necessary condition when $\boldsymbol{F}$ is smooth.

Proposition (Cross-Derivative Test)

Let $D \subset \mathbb{R}^m$ be open, and let $\boldsymbol{F} := (F_1, \ldots, F_m) : D \to \mathbb{R}^m$ be a smooth vector field. Suppose $\boldsymbol{F}$ is a gradient field. Then $\dfrac{\partial F_i}{\partial x_j} = \dfrac{\partial F_j}{\partial x_i}$ for all $i, j = 1, \ldots, m$.

Proof: Let $\boldsymbol{F} = \nabla f$ on D, where $f : D \to \mathbb{R}$ is smooth. Since the vector field $\boldsymbol{F}$ is also smooth, the partial derivatives of f of orders 1 and 2 are continuous on D. By the Mixed Partials Theorem,

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial F_j}{\partial x_i} \quad \text{for } i, j = 1, \ldots, m. \quad \square$$

The above test says that 'cross partials' are equal.

Special Cases of the Cross-Derivative Test

$m := 2$, $\boldsymbol{F} := (P, Q)$: $\quad \dfrac{\partial P}{\partial y} = \dfrac{\partial Q}{\partial x}.$

$m := 3$, $\boldsymbol{F} := (P, Q, R)$: $\dfrac{\partial P}{\partial y} = \dfrac{\partial Q}{\partial x}, \ \dfrac{\partial Q}{\partial z} = \dfrac{\partial R}{\partial y}, \ \dfrac{\partial R}{\partial x} = \dfrac{\partial P}{\partial z},$

that is, $\operatorname{curl} \boldsymbol{F} = \operatorname{curl}(\operatorname{grad} f) = \mathbf{0}$.

Examples:

(i) Let $D := \mathbb{R}^2$, and let $\boldsymbol{F} := (P, Q)$, where $P(x, y) := -y$ and $Q(x, y) := x$ for $(x, y) \in D$. Then $P_y = -1$ and $Q_x = 1$. By the cross-derivative test, $\boldsymbol{F}$ is not a gradient field on D.

(ii) Even if $\boldsymbol{F}$ satisfies the cross-derivative test on an open subset D of $\mathbb{R}^m$, $\boldsymbol{F}$ may not be a gradient field. For instance, let $D := \{(x, y) \in \mathbb{R}^2 : (x, y) \neq (0, 0)\}$, and $\boldsymbol{F} := (P, Q)$, where $P := -y/(x^2 + y^2)$ and $Q := x/(x^2 + y^2)$, $(x, y) \in D$. Then $P_y = (y^2 - x^2)/(x^2 + y^2)^2 = Q_x$. But $\boldsymbol{F}$ is not a gradient field on D: If $\boldsymbol{\gamma}$ is the standard unit circle in D, then

$$\int_{\gamma} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{-\pi}^{\pi} (-\sin t, \cos t) \cdot (-\sin t, \cos t) dt = 2\pi \neq 0.$$

Next, let $D := \{(x, y, z) \in \mathbb{R}^3 : (x, y) \neq (0, 0)\}$, and let $\boldsymbol{F} := (P, Q, R)$, where $P := -y/(x^2 + y^2)$, $Q := x/(x^2 + y^2)$ and $R := 1$ for $(x, y, z) \in D$. As above, $P_y = Q_x$, and so $\text{curl}\,(\boldsymbol{F}) = (R_y - Q_z, P_z - R_x, Q_x - P_y) = \mathbf{0}$. But $\boldsymbol{F}$ is not a gradient field on D: If $\gamma(t) := (\cos t, \sin t, 0)$, $t \in [-\pi, \pi]$, then

$$\int_{\gamma} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{-\pi}^{\pi} (-\sin t, \cos t, 1) \cdot (-\sin t, \cos t, 0) dt = 2\pi \neq 0.$$

Remark

If $D := \mathbb{R}^m$ and if $\boldsymbol{F}$ satisfies the cross-derivative test on $\mathbb{R}^m$, then it can be proved that $\boldsymbol{F}$ is indeed a gradient field on D. In fact, this also holds if D is an open simply connected subset of $\mathbb{R}^m$ and $\boldsymbol{F}$ satisfies the cross-derivative test on D. Roughly speaking, D is simply connected if it does not have any holes!

Let $D \subset \mathbb{R}^2$. Then D is called **simply connected** if every simple closed path lying in D encloses points of D only. If D is bounded, then it is simply connected if and only if the complement $\mathbb{R}^2 \setminus D$ of D in $\mathbb{R}^2$ is path-connected.
For example, a disk or a rectangle in $\mathbb{R}^2$ is simply connected, but an annulus $\{(x, y) \in \mathbb{R}^2 : 1 < x^2 + y^2 < 4\}$ is not.

Let $D \subset \mathbb{R}^3$. Then D is called **simply connected** if every simple closed path lying in D is the 'edge' of a bounded surface lying entirely in D. For example, a ball or a cuboid in $\mathbb{R}^3$ is simply connected, and so is the solid body $D := \{(x, y, z) \in \mathbb{R}^3 : 1 < x^2 + y^2 + z^2 < 4\}$, but the set $\{(x, y, z) \in \mathbb{R}^3 : (x, y) \neq (0, 0)\}$ is not.

Let a smooth vector field $\boldsymbol{F}$ satisfy the cross-derivative test on an open simply connected subset D of $\mathbb{R}^2$ or $\mathbb{R}^3$. The theorems of Green and Stokes will show that $\int_\gamma \boldsymbol{F} \cdot d\boldsymbol{s} = 0$ for every simple closed path lying in D, and hence for every closed path lying in D. This implies that $\boldsymbol{F}$ is a gradient field on D.

As a special case, let $D := I \times J \times K$, where I, J, K are open intervals in $\mathbb{R}$. Suppose a smooth vector field $\boldsymbol{F} := (P, Q, R)$ on D satisfies the cross-derivative test, that is, $\text{curl}\,(\boldsymbol{F}) = \mathbf{0}$ on D. Then $\boldsymbol{F}$ is a gradient field. How to find a scalar field f on D such that $\boldsymbol{F} = \nabla f$, that is, $(P, Q, R) = (f_x, f_y, f_z)$? This can be accomplished by repeated integration as follows. Fix $x_0 \in I$.

Solve $f_x = P$ by integrating with respect to x: Let

$$f(x, y, z) := \int_{x_0}^{x} P(u, y, z)du + g(y, z) \quad \text{for } (x, y, z) \in D,$$

where g is an arbitrary function of $(y, z) \in J \times K$. Use this expression for f in the equation $f_y = Q$ to obtain an equation for g_y. Solve it by integrating with respect to y. Its solution will involve an arbitrary function h of $z \in K$. Use this expression for g in the equation $f_z = R$ to obtain an equation for h'. Solve it by integrating with respect to z. Its solution will involve an arbitrary constant. This yields f with $\nabla f = \boldsymbol{F}$.

Further, if $\boldsymbol{F} = \nabla \tilde{f}$ for a scalar field $\tilde{f}$ on D as well, then $\nabla(\tilde{f} - f) = 0$ on D. By the Trivariate Mean Value Theorem, $\tilde{f} - f = c$, that is, $\tilde{f} = f + c$, where c is a constant.

Example:

Let $\boldsymbol{F} := (P, Q, R)$ where $P := x^2 + y\,z$, $Q := y^2 + z\,x$, $R := z^2 + x\,y$ for $(x, y, z) \in \mathbb{R}^3$. It is easy to check that $\operatorname{curl}(\boldsymbol{F}) = \mathbf{0}$. Let f be a scalar field such that $\nabla f = \boldsymbol{F}$. Then

$f_x = P = x^2 + y\,z \implies f(x, y, z) = x^3/3 + x\,y\,z + g(y, z)$.
$f_y = Q = y^2 + z\,x \implies x\,z + g_y = y^2 + z\,x$, and so $g_y = y^2$.
Thus $g(y, z) = y^3/3 + h(z)$, and as a consequence,

$$f(x, y, z) = \frac{x^3 + y^3}{3} + x\,y\,z + h(z).$$

$f_z = R = z^2 + x\,y \implies x\,y + h' = z^2 + x\,y$, and so $h' = z^2$.
Thus $h(z) = z^3/3 + c$, and as a consequence,

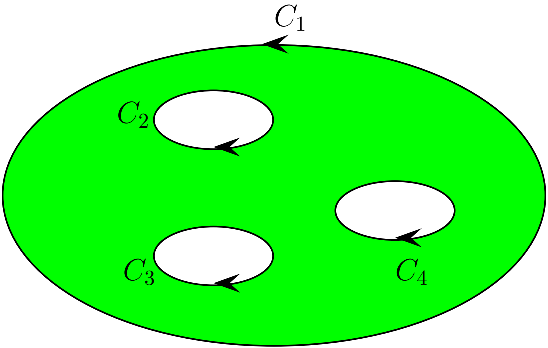
$$f(x, y, z) = \frac{x^3 + y^3 + z^3}{3} + x\,y\,z + c, \quad \text{where } c \text{ is a constant.}$$

Let γ be a simple closed piecewise smooth path in $\mathbb{R}^2$, and let $C := \gamma([\alpha, \beta])$ denote the geometric curve in $\mathbb{R}^2$ defined by γ. Suppose $\boldsymbol{F}$ is a continuous vector field defined on C.

Since the line integral of a continuous vector field along a smooth path is invariant under a reparametrization only up to its sign, $\int_C \boldsymbol{F} \cdot d\boldsymbol{s}$ is determined only up to its sign. To determine it completely, we specify the sense or the direction in which the geometric curve C is traversed. For example, we may specify that as we travel along the curve C with $\boldsymbol{k}$ as our upright direction, the region in $\mathbb{R}^2$ enclosed by C lies on our left. In this case, the curve C is traversed **anticlockwise** (as viewed from high above the curve).

Suppose D is a bounded region in $\mathbb{R}^2$ whose boundary ∂D consists of a finite number of nonintersecting simple closed piecewise smooth geometric curves $C_1, \ldots, C_k$.

Suppose the curve C_1 is the outer boundary of D.



If the outer boundary curve C_1 of D is traversed anticlockwise, but the inner boundary curves $C_2, \ldots, C_k$ are traversed clockwise, then we say that the boundary ∂D of the region D is **positively oriented**. In this case, the region D is on our left as we travel along any part of its boundary (with $\boldsymbol{k}$ as our upright direction).

The Green Theorem

We have defined line integrals in terms of Riemann integrals. An important result of Green relates line integrals along the boundary of a domain to double integrals over the domain. It can be viewed as a two-dimensional analogue of Part II of the Fundamental Theorem of Calculus for Riemann integrals and for line integrals.

Theorem

Let D be a closed and bounded subset of $\mathbb{R}^2$ such that the boundary ∂D of D consists of a finite number of nonintersecting simple closed piecewise smooth curves, and suppose it is positively oriented. Let $\boldsymbol{F} := (P, Q)$ be a smooth vector field on an open subset Ω of $\mathbb{R}^2$ containing D. Then

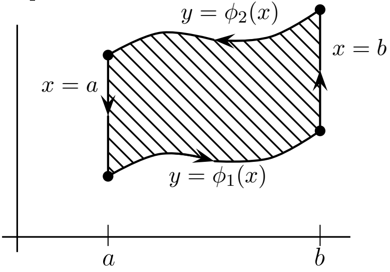
$$\int_{\partial D} P\,dx + Q\,dy = \iint_D (Q_x - P_y)\,d(x,y).$$

Proof. The LHS is well-defined since P and Q are continuous on ∂D, while the RHS is well-defined since Q_x and P_y are continuous on D and since ∂D is of two-dimensional content zero (as shown in Lecture 25).

First suppose D is an elementary region of type I, and

$$D := \{(x, y) \in \mathbb{R}^2 : a \le x \le b \text{ and } \phi_1(x) \le y \le \phi_2(x)\},$$

where $\phi_1, \phi_2 : [a, b] \to \mathbb{R}$ are continuous and $\phi_1 \le \phi_2$.



Clearly,

$$\int_{\partial D} P\, dx \;=\; \int_a^b P(x, \phi_1(x))dx + 0 + \int_b^a P(x, \phi_2(x))dx + 0$$

which is equal to $\int_a^b \big(P(x,\phi_1(x)) - P(x,\phi_2(x))\big)\,dx$.

On the other hand, the Fubini theorem and the FTC show that

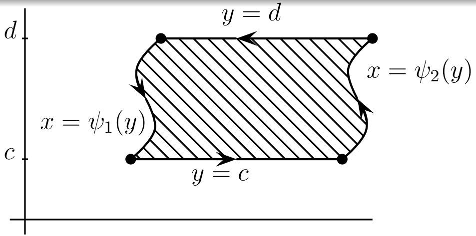
$$\begin{aligned}\iint_D P_y\, d(x,y) &= \int_a^b \left(\int_{\phi_1(x)}^{\phi_2(x)} P_y\, dy\right) dx \\ &= \int_a^b \big(P(x,\phi_2(x)) - P(x,\phi_1(x))\big)\,dx.\end{aligned}$$

Thus $\displaystyle\int_{\partial D} P\ dx = -\iint_D P_y\, d(x,y).$

Next, suppose D is an elementary region of type II, and

$$D := \{(x,y)\in\mathbb{R}^2 : c \le y \le d \text{ and } \psi_1(y) \le x \le \psi_2(y)\},$$

where $\psi_1,\psi_2 : [c,d] \to \mathbb{R}$ are continuous and $\psi_1 \le \psi_2$.



Clearly,

$$\int_{\partial D} Q\,dy \;=\; \int_c^d Q(\psi_2(y), y)dy + 0 + \int_d^c Q(\psi_1(y), y)dy + 0$$

which is equal to $\int_c^d \big(Q(\psi_2(y), y) - Q(\psi_1(y), y)\big)\,dy$.

On the other hand, the Fubini theorem and the FTC show that

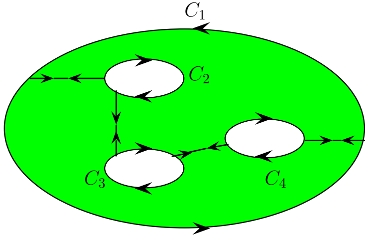
$$\begin{aligned}\iint_D Q_x\,d(x,y) &= \int_c^d \left(\int_{\psi_1(y)}^{\psi_2(y)} Q_x\,dx\right)dy \\ &= \int_c^d \big(Q(\psi_2(y), y) - Q(\psi_1(y), y)\big)\,dy.\end{aligned}$$

Thus $\int_{\partial D} Q\, dy = \iint_D Q_x\, d(x, y)$.

If D is an elementary region of type I as well as of type II, we may combine the two equalities proved above, and obtain

$$\int_{\partial D} P\ dx + Q\ dy = \iint_D (Q_x - P_y) d(x, y).$$

This proof applies when D is a rectangle or a disk. It can be modified to treat the case where D is simply connected, so that ∂D has only the outer component. Further, the general case of a **multiply connected** domain D can be treated by introducing cuts as shown below.



Computational Uses of the Green Theorem

(i) Calculation of a double integral

Example: Finding the area of a subset of $\mathbb{R}^2$

Let D be a closed and bounded subset of $\mathbb{R}^2$ whose positively oriented boundary ∂D consists of a finite number of nonintersecting simple piecewise smooth closed curves. Then

$$\text{Area}\,(D) := \iint_D 1_D d(x,y).$$

Let P, Q be smooth functions on an open subset containing D such that $Q_x - P_y = 1$ on D. By the Green theorem,

$$\text{Area}\,(D) = \iint_D (Q_x - P_y) d(x,y) = \int_{\partial D} P\ dx + Q\ dy.$$

For instance, for $(x,y) \in D$, we may let

$$P(x,y) := -\frac{y}{2} \quad \text{and} \quad Q(x,y) := \frac{x}{2}.$$

Then clearly $Q_x - P_y = (1/2) + (1/2) = 1$ on D, and so

$$\text{Area}\,(D) = \frac{1}{2}\int_{\partial D} x\ dy - y\ dx.$$

We note that Area (D) is also equal to

$$\int_{\partial D} x\,dy \quad \text{or} \quad -\int_{\partial D} y\,dx$$

if we let $Q(x,y) := x,\ P := 0$, or $P(x,y) := -y,\ Q := 0$.

Suppose the positively oriented boundary ∂D of D is parametrized by $(x(t), y(t)),\ t \in [\alpha, \beta]$. Then

$$\text{Area}\,(D) = \frac{1}{2}\int_{\alpha}^{\beta} \big(x(t)y'(t) - y(t)x'(t)\big)\,dt = \frac{1}{2}\int_{\alpha}^{\beta} W(x,y)(t)dt,$$

where $W(x,y) := \det \begin{bmatrix} x & y \\ x' & y' \end{bmatrix}$ is called the **Wronskian** of x, y.

As a special case, suppose ∂D is given by a polar equation

$$r = p(\theta) \quad \text{for } \theta \in [\alpha, \beta].$$

Let $x(\theta) := p(\theta)\cos\theta$ and $y(\theta) := p(\theta)\sin\theta$ for $\theta \in [\alpha, \beta]$. Then

$$W(x,y)(\theta) = \det \begin{bmatrix} p(\theta)\cos\theta & p(\theta)\sin\theta \\ p'(\theta)\cos\theta - p(\theta)\sin\theta & p'(\theta)\sin\theta + p(\theta)\cos\theta \end{bmatrix},$$

which is equal to $p^2(\theta)$, and so

$$\text{Area}\,(D) = \frac{1}{2}\int_\alpha^\beta p^2(\theta)d\theta,$$

as we had defined earlier. For instance, the area enclosed by the cardioid given by $r = a(1-\cos\theta)$, $\theta \in [-\pi, \pi]$, is equal to

$$\frac{1}{2}\int_{-\pi}^{\pi} a^2(1 - 2\cos\theta + \cos^2\theta)d\theta = \frac{3\pi a^2}{2}.$$

(ii) Calculation of the line integral along an oriented boundary

Example:

Let $D := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1 \text{ and } y \geq 0\}$, so that its positively oriented boundary ∂D consists of the line segment from $(-1, 0)$ to $(1, 0)$ followed by the semicircle in the upper half-plane having radius 1 and centre $(0, 0)$. Let $P(x, y) := y^2$ and $Q(x, y) := 3xy$ for $(x, y) \in \mathbb{R}^2$.

To calculate $\int_{\partial D} P\,dx + Q\,dy$, we may use the Green theorem and obtain

$$\begin{aligned}\int_{\partial D} y^2 dx + 3xy\,dy &= \iint_D \left[\frac{\partial}{\partial x}(3x\,y) - \frac{\partial}{\partial y}(y^2)\right]\, d(x, y) \\ &= \int_{-1}^{1}\left(\int_0^{\sqrt{1-x^2}} y\ dy\right)dx = \int_{-1}^{1} \frac{1-x^2}{2}dx = \frac{2}{3}.\end{aligned}$$

Consequences of the Green Theorem

Proposition

Let $\boldsymbol{F} := (P, Q)$ be a smooth vector field on an open subset containing a closed and bounded subset D of $\mathbb{R}^2$ such that $Q_x = P_y$ on D. If ∂D consists of a finite number of nonintersecting simple closed piecewise smooth curves, and if ∂D is positively oriented, then

$$\int_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\partial D} P \ dx + Q \ dy = 0.$$

Proof

By the Green theorem,

$$\int_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_D (Q_x - P_y) d(x, y) = \iint_D 0 \, d(x, y) = 0. \quad \square$$

MA 105 : Calculus
Division 1, Lecture 23

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Path independence of line integrals
- Examples. Criterion for path independence
- Fundamental Theorem of Calculus for Line Integrals
- Cross Derivative Test. Special cases
- Sufficiency of cross derivative test for continuous vector fields on simply connected domains
- Positive orientation for the boundary of a planar region
- Green's theorem
- Computational uses of Green's theorem
- Consequences of Green's theorem

Consequences of the Green Theorem

Proposition

Let $\boldsymbol{F} := (P, Q)$ be a smooth vector field on an open subset containing a closed and bounded subset D of $\mathbb{R}^2$ such that $Q_x = P_y$ on D. If ∂D consists of a finite number of nonintersecting simple closed piecewise smooth curves, and if ∂D is positively oriented, then

$$\int_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\partial D} P\ dx + Q\ dy = 0.$$

Proof

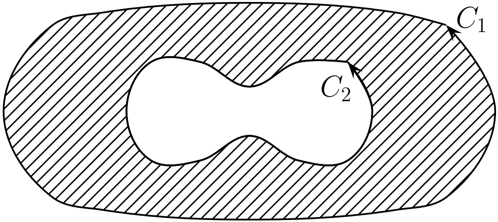
By the Green theorem,

$$\int_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_D (Q_x - P_y) d(x, y) = \iint_D 0\, d(x, y) = 0. \quad \square$$

Invariance of Some Line Integrals

Suppose C_1 and C_2 are simple closed piecewise smooth curves such that C_2 lies in the interior of C_1, that is, C_1 encloses C_2, and suppose both are oriented counterclockwise. Let P, Q be smooth scalar fields satisfying $Q_x = P_y$ on an open set containing C_1, C_2 and the region between them. Then

$$\int_{C_1} P\ dx + Q\ dy = \int_{C_2} P\ dx + Q\ dy.$$



This result follows by applying the Green theorem to the subset D of $\mathbb{R}^2$ consisting of C_1, C_2 and the region between them, and by noting that $\partial D = C_1 + (-C_2)$.

This is known as the deformation principle for line integrals along paths in $\mathbb{R}^2$.

Example: (Gauss Law in $\mathbb{R}^2$): Let C be a piecewise smooth simple closed geometric curve in $\mathbb{R}^2$, and let D denote the union of C and the region enclosed by C, so that $\partial D = C$. Suppose $(0,0) \notin C$, and C is positively oriented. Then

$$\int_C \frac{\tilde{\boldsymbol{r}}}{r^2} \cdot d\boldsymbol{s} = \begin{cases} 0 & \text{if } (0,0) \notin D, \\ 2\pi & \text{if } (0,0) \in D, \end{cases}$$

where $\tilde{\boldsymbol{r}}(x,y) := (-y,x)$ and $r(x,y) := \|(x,y)\|$, $(x,y) \in \mathbb{R}^2$.

Proof. Let $\boldsymbol{F} := \tilde{\boldsymbol{r}}/r^2$, that is, $\boldsymbol{F} := (P,Q)$, where $P = -y/r^2 = -y/(x^2+y^2)$ and $Q := x/r^2 = x/(x^2+y^2)$ for $\boldsymbol{r} \neq \boldsymbol{0}$. Then $Q_x - P_y = 0$ on $D \setminus \{(0,0)\}$ as we saw earlier.

First, let $(0,0) \notin D$. Then the vector field $\boldsymbol{F}$ is smooth on D. Hence $\int_C \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_D (Q_x - P_y) d(x,y) = 0$ by the Green theorem.

Next, let $(0,0) \in D$. We have seen earlier that $\int_{C_1} \boldsymbol{F} \cdot d\boldsymbol{s} = 2\pi$, where C_1 denotes the standard parametrized unit circle in $\mathbb{R}^2$ (which is positively oriented). Since $(0,0)$ is an interior point of D, there is $\epsilon > 0$ such that the closed disk of radius ϵ and center $(0,0)$ lies inside D. Let C_ϵ denote the standard parametrized circle of radius ϵ (which is also positively oriented). Then

$$\int_C \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{C_\epsilon} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{C_1} \boldsymbol{F} \cdot d\boldsymbol{s} = 2\pi$$

by the deformation principle for line integrals, applied first to the curves C and C_ϵ, and then to the curves C_ϵ and C_1. $\square$

Proposition

Let Ω be a simply connected open subset of $\mathbb{R}^2$, and let $\boldsymbol{F} := (P, Q)$ be a smooth vector field such that $Q_x = P_y$ on Ω. Then $\boldsymbol{F}$ is a gradient field on Ω.

Proof. Let C denote a simple closed smooth curve in Ω. Since Ω is simply connected, C encloses points of Ω only. Let D denote the set of points on and inside C. By the Green theorem, $\int_C \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_D (Q_x - P_y) d(x, y) = 0$. This also holds if a smooth closed curve C in Ω intersects itself, since we can break up C into several simple closed smooth curves. As a result, line integrals of $\boldsymbol{F}$ are path-independent in D. $\square$

Recall the differential notation for a regular smooth curve γ:

$ds = \|\gamma'(t)\| dt$, $d\boldsymbol{s} = \gamma'(t) dt$, $\hat{\boldsymbol{t}} = \gamma'(t)/\|\gamma'(t)\|$, $d\boldsymbol{s} = \hat{\boldsymbol{t}}\, ds$, $\hat{\boldsymbol{n}} = \hat{\boldsymbol{t}} \times \boldsymbol{k}$. If $\boldsymbol{F} := (P, Q)$, then $\boldsymbol{F} \cdot \hat{\boldsymbol{n}}\; ds = P\, dy - Q\, dx$.

Alternative Formulations of the Green Formula

We can express the conclusion of the Green theorem, namely,

$$\int_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\partial D} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) d(x,y).$$

as follows.

(i) Circulation-Curl Form or Tangential Form:

$$\int_{\partial D} \boldsymbol{F} \cdot \hat{\boldsymbol{t}}\,ds = \iint_D (\operatorname{curl} \boldsymbol{F}) \cdot \boldsymbol{k}\,d(x,y). \quad [\text{Let } \boldsymbol{F} := (P, Q, 0).]$$

(The Stokes theorem is a 3-dimensional curved version of this.)

(ii) Flux-Divergence Form or Normal Form:

$$\int_{\partial D} \boldsymbol{F} \cdot \hat{\boldsymbol{n}}\,ds = \iint_D (\operatorname{div} \boldsymbol{F}) d(x,y). \quad [\text{Let } \boldsymbol{F} := (-Q, P, 0).]$$

(The Gauss theorem is a 3-dimensional analogue of this.)

Surfaces in $\mathbb{R}^3$

We now move up one dimension, and pass from paths and line integrals to surfaces and surface integrals.

Just as a curve in $\mathbb{R}^2$ is given implicitly by an equation $F(x, y) = 0$, or parametrically by $(x(t), y(t))$ for $t \in [\alpha, \beta]$, a surface in $\mathbb{R}^3$ is given implicitly by an equation $F(x, y, z) = 0$, or parametrically by $(x(u, v), y(u, v), z(u, v))$ for $(u, v) \in E$, where E is a subset of $\mathbb{R}^2$. Here u and v are independent parameters, that is, they do not depend on each other.

For example, the **unit sphere** in $\mathbb{R}^3$ is given implicitly by the equation $x^2 + y^2 + z^2 - 1 = 0$, and also parametrically by $\big((\sin u)(\cos v), (\sin u)(\sin v), \cos u\big)$ for $(u, v) \in [0, \pi] \times [-\pi, \pi]$.

We have discussed the tangent planes and normal lines for implicitly defined surfaces in Lecture 21. Now we shall treat parametrically defined surfaces.

Parametrized Surfaces

A **parametrized surface** in $\mathbb{R}^3$ is a continuous map $\Phi : E \to \mathbb{R}^3$, where E is a path-connected subset of $\mathbb{R}^2$ having an area. Let $\Phi(u, v) := (x(u, v), y(u, v), z(u, v))$ for $(u, v) \in E$. The parametrized surface Φ is called **smooth** if the functions x, y, z have continuous partial derivatives on an open subset of $\mathbb{R}^2$ containing E.

Given a parametrized surface, we may eliminate the parameters u, v to obtain an equation involving only x, y, z, which defines the surface implicitly. The converse process is generally harder.

Examples

(i) Consider a continuous function $f : E \to \mathbb{R}$, and define $\Phi : E \to \mathbb{R}^3$ by $\Phi(x, y) := (x, y, f(x, y))$ for $(x, y) \in E$. This parametrized surface is called the **graph** of f. It is implicitly given by the equation $F(x, y, z) := z - f(x, y) = 0$.

(ii) Let $a > 0$ and $E := [-\pi, \pi] \times [0, h]$, where $h > 0$.

Define $\Phi : E \to \mathbb{R}^3$ by $\Phi(\theta, z) := (a\cos\theta, a\sin\theta, z)$ for $(\theta, z) \in E$. This parametrized surface is called the right circular **cylinder** of radius a and height h. It is implicitly given by the equation $F(x, y, z) := x^2 + y^2 - a^2 = 0$ with $0 \le z \le h$.

Define $\Phi : E \to \mathbb{R}^3$ by $\Phi(\theta, z) := (z\cos\theta, z\sin\theta, z)$ for $(\theta, z) \in E$. This parametrized surface is called the right circular **cone** of height h. It is implicitly given by the equation $F(x, y, z) := x^2 + y^2 - z^2 = 0$ with $0 \le z \le h$.

(iii) Let $a > 0$ and $E := [0, \pi] \times [-\pi, \pi]$. Define $\Phi : E \to \mathbb{R}^3$ by $\Phi(\varphi, \theta) := (a\sin\varphi\cos\theta, a\sin\varphi\sin\theta, a\cos\varphi)$ for $(\varphi, \theta) \in E$. This parametrized surface is called the **standard sphere of radius** a. It is implicitly given by the equation $F(x, y, z) := x^2 + y^2 + z^2 - a^2 = 0$.

Normal Vector to a Parametrized Surface

Let $\Phi : E \to \mathbb{R}^3$ be a smooth parametrized surface. Let $(u_0, v_0) \in E$, and let $P_0 := \Phi(u_0, v_0)$.

Suppose $\boldsymbol{\gamma}(t) := (u(t), v(t))$, $t \in [\alpha, \beta]$, is a smooth path that lies in E and passes through (u_0, v_0), and let $t_0 \in [\alpha, \beta]$ be such that $\boldsymbol{\gamma}(t_0) = (u_0, v_0)$.

Now the path $\Phi \circ \boldsymbol{\gamma}$ lies in $\Phi(E)$, and by the Chain Rule, the tangent vector to this path at P_0 is

$$(\Phi \circ \boldsymbol{\gamma})'(t_0) = \Phi_u(u_0, v_0)\ u'(t_0) + \Phi_v(u_0, v_0)\ v'(t_0).$$

Because $\Phi_u(u_0, v_0) \times \Phi_v(u_0, v_0)$ is perpendicular to both $\Phi_u(u_0, v_0)$ and $\Phi_v(u_0, v_0)$, it is also perpendicular to the tangent vector $(\Phi \circ \boldsymbol{\gamma})'(t_0)$. Since this holds for any such path $\boldsymbol{\gamma}$ that lies in E, $(\Phi_u \times \Phi_v)(u_0, v_0) := \Phi_u(u_0, v_0) \times \Phi_v(u_0, v_0)$ is called a **normal vector** to Φ at P_0. It is also called the **fundamental product** for Φ at $(u_0, v_0) \in E$.

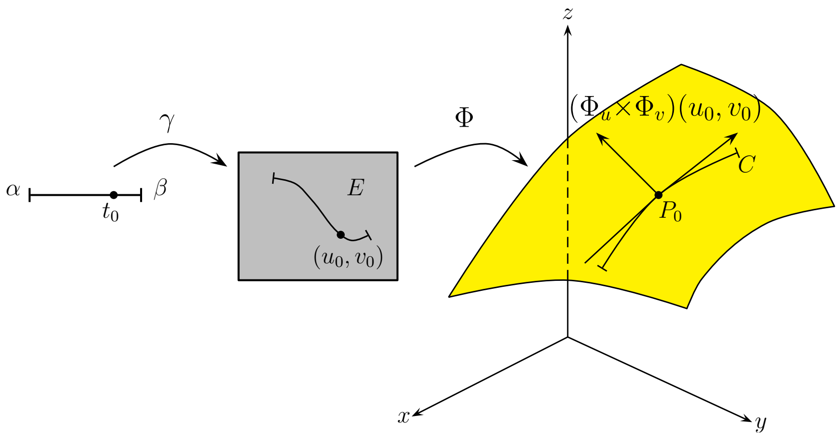


Figure: Normal vector to a parametrized surface at P_0

The Fundamental Product

How do we find the fundamental product of a parametrized surface $\Phi : E \to \mathbb{R}^3$ at $(u, v) \in E$?
Let $\Phi(u, v) := (x(u, v), y(u, v), z(u, v))$ for $(u, v) \in E$. Then

$$\begin{aligned}
\Phi_u \times \Phi_v &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} \\
&= \left(\begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix}, \begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix}, \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix} \right) \\
&= \left(\frac{\partial(y, z)}{\partial(u, v)}, \frac{\partial(z, x)}{\partial(u, v)}, \frac{\partial(x, y)}{\partial(u, v)} \right),
\end{aligned}$$

where $\dfrac{\partial(y, z)}{\partial(u, v)}$ is the Jacobian of the transformation $(u, v) \longmapsto (y(u, v), z(u, v))$ etc., as defined in Lecture 19.

We observe that

$$\|(\Phi_u \times \Phi_v)(u, v)\| = \sqrt{\left(\frac{\partial(y,z)}{\partial(u,v)}\right)^2 + \left(\frac{\partial(z,x)}{\partial(u,v)}\right)^2 + \left(\frac{\partial(x,y)}{\partial(u,v)}\right)^2}.$$

A smooth parametrized surface Φ is said to be **regular** if $(\Phi_u \times \Phi_v)(u, v) \neq \mathbf{0}$ for every $(u, v) \in E$.
In this case,

$$\hat{\boldsymbol{n}}\,(u, v) := \frac{(\Phi_u \times \Phi_v)(u, v)}{\|(\Phi_u \times \Phi_v)(u, v)\|}$$

is called a **unit normal vector** to Φ at $\Phi(u, v)$, $(u, v) \in E$.

Let $(u_0, v_0) \in E$, and let $P_0 := \Phi(u_0, v_0)$. If $P_0 := (x_0, y_0, z_0)$, then the **tangent plane** to Φ at P_0 is given by the equation

$$(\Phi_u \times \Phi_v)(u_0, v_0) \cdot (x - x_0, y - y_0, z - z_0) = 0,$$

that is, by the equation $a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$,

where $(a, b, c) := (\Phi_u \times \Phi_v)\,(u_0, v_0)$.

Examples

(i) Graph of a function

Let E be a path-connected subset of $\mathbb{R}^2$ having an area, and let $f : E \to \mathbb{R}$ be a smooth function.
Let $\varPhi(x, y) = (x, y, f(x, y))$ for $(x, y) \in E$. Then

$$\varPhi_x \times \varPhi_y = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 1 & 0 & f_x \\ 0 & 1 & f_y \end{vmatrix} = (-f_x, -f_y, 1) \quad \text{on } E.$$

Since $\|\varPhi_x \times \varPhi_y\| = \sqrt{f_x^2 + f_y^2 + 1} \neq 0$ on E, $\varPhi$ is a regular surface. The tangent plane to $\varPhi$ at $(x_0, y_0, f(x_0, y_0))$ is given by

$$z - f(x_0, y_0) = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

(ii) Cylinder

Let $\Phi(\theta, z) := (a\cos\theta, a\sin\theta, z)$ for $(\theta, z) \in [-\pi, \pi] \times [0, h]$, where $a > 0$. Then

$$\Phi_\theta \times \Phi_z = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ -a\sin\theta & a\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = (a\cos\theta, a\sin\theta, 0).$$

Since $\|\Phi_\theta \times \Phi_z\| = a \neq 0$ on $(\theta, z) \in [-\pi, \pi] \times [0, h]$, Φ is a regular surface. The tangent plane to Φ at $(a\cos\theta_0, a\sin\theta_0, z_0)$ is given by

$$(\cos\theta_0)x + (\sin\theta_0)y = a.$$

(iii) Sphere

Let $\Phi(\varphi, \theta) := (a\sin\varphi\cos\theta, a\sin\varphi\sin\theta, a\cos\varphi)$ for $(\varphi, \theta) \in [0, \pi] \times [-\pi, \pi]$, where $a > 0$. Then

$$\Phi_\varphi \times \Phi_\theta = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a\cos\varphi\cos\theta & a\cos\varphi\sin\theta & -a\sin\varphi \\ -a\sin\varphi\sin\theta & a\sin\varphi\cos\theta & 0 \end{vmatrix}.$$

Thus

$$\begin{aligned}\Phi_\varphi \times \Phi_\theta &= (a^2\sin^2\varphi\cos\theta, a^2\sin^2\varphi\sin\theta, a^2\sin\varphi\cos\varphi) \\ &= (a\sin\varphi)\,\Phi(\varphi,\theta).\end{aligned}$$

Since $\|\Phi(\varphi,\theta)\| = a$ for $(\varphi,\theta) \in [0,\pi]\times[-\pi,\pi]$, we obtain

$$\|\Phi_\varphi \times \Phi_\theta\| = a^2\sin\varphi \neq 0 \quad \text{if } (\varphi,\theta) \in (0,\pi)\times[-\pi,\pi],$$

and then the tangent plane to Φ at $\Phi(\varphi_0,\theta_0)$ is given by

$$(\sin\varphi_0\cos\theta_0)x + (\sin\varphi_0\sin\theta_0)y + (\cos\varphi_0)z = a.$$

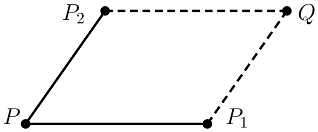
Using the parametrization $\Phi(x,y) := \left(x, y, \sqrt{a^2-x^2-y^2}\right)$ for $(x,y)\in\mathbb{R}^2$ satisfying $x^2+y^2\le a^2$, we obtain the unit normal vector at the **north pole** $P_0 := (0,0,a)$ to be $(0,0,1)$, and the equation of the tangent plane at P_0 to be $z=a$. Similarly we may treat the **south pole** $Q_0 := (0,0,-a)$.

Surface Area

Let $\Phi : E \to \mathbb{R}^3$ be a smooth parametrized surface, where E is a path-connected subset of $\mathbb{R}^2$ having an area. Let $(u, v) \in E$. For $h, k \in \mathbb{R}$ with $|h|, |k|$ small, consider

$$P := \Phi(u, v), \quad P_1 := \Phi(u + h, v) \approx \Phi(u, v) + h\, \Phi_u(u, v),$$

$$P_2 := \Phi(u, v+k) \approx \Phi(u, v)+k\, \Phi_v(u, v), \quad Q := \Phi(u+h, v+k).$$



Area of the parallelogram with sides PP_1 and PP_2

$$= \|(P_1 - P) \times (P_2 - P)\| \approx \|\Phi_u(u, v) \times \Phi_v(u, v)\|\, |h||k|.$$

In view of this approximation, we define

$$\text{Area}\,(\Phi) := \iint_E \|(\Phi_u \times \Phi_v)(u, v)\| d(u, v).$$

Since the subset E of $\mathbb{R}^2$ has an area, the two-dimensional content of ∂E is zero. Also, the function $\|\Phi_u \times \Phi_v\|$ is continuous on E. Hence the integral in the definition of Area (Φ) is well-defined. In analogy with the differential notation $ds = \|\gamma'(t)\|dt$, we introduce the following differential notation:

$$dS = \|\Phi_u \times \Phi_v\| d(u, v).$$

Examples

(i) Graph of a function

Let a subset E of $\mathbb{R}^2$ have an area, $f : E \to \mathbb{R}$ be a smooth function, and $\Phi(x, y) = (x, y, f(x, y))$ for $(x, y) \in E$. Then

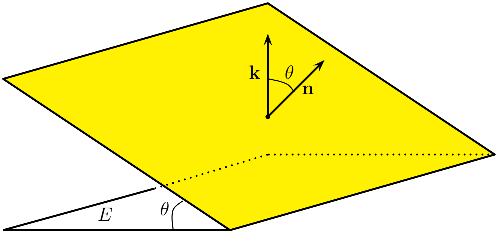
$$\begin{aligned} \text{Area}\,(\Phi) &= \iint_E \|(-f_x, -f_y, 1)\| d(x, y) \\ &= \iint_E \sqrt{1 + f_x^2 + f_y^2}\, d(x, y). \end{aligned}$$

Let $\theta(x,y) \in [0,\pi]$ be the angle between $\Phi_x \times \Phi_y = (-f_x, -f_y, 1)$ at (x,y) and $\boldsymbol{k} = (0,0,1)$. Since $(\Phi_x \times \Phi_y) \cdot \boldsymbol{k} = 1 \neq 0$, we see that $\theta(x,y) \neq \pi/2$ and

$$\cos\theta(x,y) = \frac{1}{\sqrt{1+f_x^2+f_y^2}}, \qquad \text{Area}\,(\Phi) = \iint_E \frac{d(x,y)}{\cos\theta(x,y)}.$$

In case $\theta(x,y) := \theta$, a constant for all $(x,y) \in E$, then $\theta \neq \pi/2$ and we obtain

$$\text{Area}\,(\Phi) = \frac{\text{Area}\,(E)}{\cos\theta} \quad \textcolor{red}{\text{(Area Cosine Principle)}}.$$



(ii) Let $E := [-\pi, \pi] \times [0, h]$, $\Phi(\theta, z) := (a\cos\theta, a\sin\theta, z)$, and $\Psi(\theta, z) := (a\cos 2\theta, a\sin 2\theta, z)$ for $(\theta, z) \in E$. Then

$$\begin{aligned} \text{Area}\,(\Phi) &= \iint_E \|\Phi_\theta \times \Phi_z\| d(\theta, z) = \iint_E a\, d(\theta, z) = 2\pi a\, h, \\ \text{Area}\,(\Psi) &= \iint_E \|\Psi_\theta \times \Psi_z\| d(\theta, z) = \iint_E 2a\, d(\theta, z) = 4\pi a\, h. \end{aligned}$$

We note that $\Psi(E) = \Phi(E)$, but $\text{Area}\,(\Psi) = 2\,\text{Area}\,(\Phi)$.

(iii) Let $E := [0, \pi] \times [-\pi, \pi]$, and
$\Phi(\varphi, \theta) = (a\sin\varphi\cos\theta, a\sin\varphi\sin\theta, a\cos\varphi)$ for $(\varphi, \theta) \in E$.
Then

$$\begin{aligned} \text{Area}\,(\Phi) &= \iint_E \|\Phi_\varphi \times \Phi_\theta\| d(\varphi, \theta) = \iint_E a^2 \sin\varphi\, d(\varphi, \theta) \\ &= \int_{-\pi}^{\pi} \left(\int_0^{\pi} a^2 \sin\varphi\, d\varphi \right) d\theta = 4\pi a^2. \end{aligned}$$

(iv) Let C be a smooth curve in $\mathbb{R}^2$ given by $\boldsymbol{\gamma}(t) := (x(t), y(t))$, $t \in [\alpha, \beta]$. If C lies on or above the x-axis, and C is revolved about the x-axis, then it generates a surface parametrized by

$$\Phi(t,\theta) := (x(t),\ y(t)\cos\theta,\ y(t)\sin\theta) \quad \text{for } (t,\theta) \in E,$$

where $E := [\alpha, \beta] \times [-\pi, \pi]$. For all $(t, \theta) \in E$,

$$\begin{aligned}(\Phi_t \times \Phi_\theta)(t,\theta) &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x'(t) & y'(t)\cos\theta & y'(t)\sin\theta \\ 0 & -y(t)\sin\theta & y(t)\cos\theta \end{vmatrix} \\ &= (y(t)y'(t), -x'(t)y(t)\cos\theta, -x'(t)y(t)\sin\theta).\end{aligned}$$

By the Fubini theorem, we obtain

$$\begin{aligned}\text{Area}\,(\Phi) &= \iint_E \sqrt{y(t)^2 y'(t)^2 + x'(t)^2 y(t)^2}\, d(t,\theta) \\ &= 2\pi \int_\alpha^\beta y(t)\sqrt{x'(t)^2 + y'(t)^2} dt, \text{ as defined in Lecture 11.}\end{aligned}$$

Surface Integral of a Scalar Field

Let E be a subset of $\mathbb{R}^2$ having an area, and let $\varPhi : E \to \mathbb{R}^3$ be a smooth parametrized surface. Let $S := \varPhi(E)$, and $f : S \to \mathbb{R}$ be a bounded scalar field. The **surface integral** of f **across** $\varPhi$ is defined by

$$\iint_{\varPhi} f\, dS := \iint_E f(\varPhi(u,v)) \|(\varPhi_u \times \varPhi_v)(u,v)\| d(u,v),$$

provided the double integral on the right side exists.
In particular, if f is continuous, then $\iint_{\varPhi} f\, dS$ is well-defined.

Leting $f := 1$ on S, we see that $\iint_{\varPhi} dS = \text{Area}\,(\varPhi)$.

The algebraic and the order properties of the double integral imply similar properties of the surface integral of a scalar field.

MA 105 : Calculus
Division 1, Lecture 24

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Invariance of some line integrals
- Alternative formulations of Green's formula
- Surfaces in 3-space: Implicitly defined and parametrized surface
- Normal vector to a smooth parametrized surface
- Fundamental product and Jacobians
- Unit normal vector to a regular parametrized surface
- Tangent plane to a smooth parametrized surface
- Examples
- Surface area of a parametrized surface
- Area Cosine Principle
- Surface integral of a scalar field

Surface Integral of a Scalar Field

Let E be a subset of $\mathbb{R}^2$ having an area, and let $\Phi : E \to \mathbb{R}^3$ be a smooth parametrized surface. Let $S := \Phi(E)$, and $f : S \to \mathbb{R}$ be a bounded scalar field. The **surface integral** of f **across** Φ is defined by

$$\iint_{\Phi} f\, dS := \iint_{E} f(\Phi(u,v)) \|(\Phi_u \times \Phi_v)(u,v)\| d(u,v),$$

provided the double integral on the right side exists.
In particular, if f is continuous, then $\iint_{\Phi} f\, dS$ is well-defined.

Leting $f := 1$ on S, we see that $\iint_{\Phi} dS = \text{Area}\,(\Phi)$.

The algebraic and the order properties of the double integral imply similar properties of the surface integral of a scalar field.

Example

Let $E := [0, \pi/2] \times [-\pi, \pi]$, and for $(\varphi, \theta) \in E$, let $\Phi(\varphi, \theta) := (a \sin\varphi \cos\theta, a \sin\varphi \sin\theta, a \cos\varphi)$. Then

$$\begin{aligned} \text{Area}\,(\Phi) &= \iint_E \|\Phi_\varphi \times \Phi_\theta\| d(\varphi, \theta) = \iint_E a^2 \sin\varphi \, d(\varphi, \theta) \\ &= \int_{-\pi}^{\pi} \left(\int_0^{\pi/2} a^2 \sin\varphi \, d\varphi \right) d\theta = 2\pi a^2. \end{aligned}$$

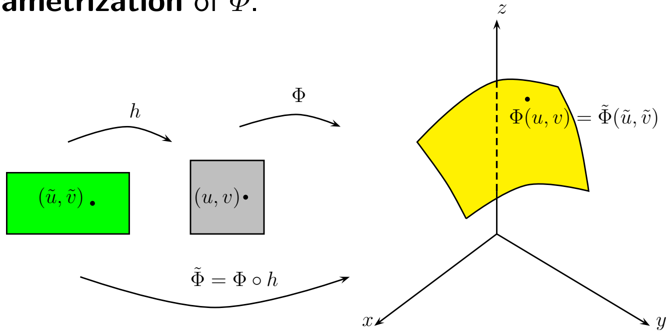
Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(x, y, z) := z$ for $(x, y, z) \in \mathbb{R}^3$. Then

$$\begin{aligned} \iint_\Phi f \, dS &= \int_{-\pi}^{\pi} \left(\int_0^{\pi/2} (a \cos\varphi)(a^2 \sin\varphi) d\varphi \right) d\theta \\ &= \pi a^3 \int_0^{\pi/2} \sin 2\varphi \, d\varphi = \pi a^3. \end{aligned}$$

Reparametrization of a Surface

Let E be a path-connected subset of $\mathbb{R}^2$ having an area, and let $\Phi : E \to \mathbb{R}^3$ be a smooth parametrized surface.

Let $\widetilde{E}$ be a path-connected subset of $\mathbb{R}^2$ having an area, and let $h : \widetilde{E} \to \mathbb{R}^2$ be a continuously differentiable and one-one function such that $h(\widetilde{E}) = E$ and its Jacobian $J(h)$ does not vanish on $\widetilde{E}$. Then the smooth surface $\tilde{\Phi} := \Phi \circ h$ is called a **reparametrization** of Φ.



Proposition

The surface integral of a continuous scalar field over a smooth surface is invariant under reparametrization. In particular, the area of a smooth surface is invariant under reparametrization.

Proof. Let E be a path-connected subset of $\mathbb{R}^2$ having an area. Let $\varPhi : E \to \mathbb{R}^3$ be a smooth parametrized surface, and let f be a continuous scalar field on $S := \varPhi(E)$.

Let $\tilde{\varPhi} := \varPhi \circ h$ be a reparametrization of $\varPhi$. Using the chain rule, it can be checked that

$$(\widetilde{\varPhi}_{\tilde{u}} \times \widetilde{\varPhi}_{\tilde{v}})(\tilde{u}, \tilde{v}) = (\varPhi_u \times \varPhi_v)(h(\tilde{u}, \tilde{v}))J(h)(\tilde{u}, \tilde{v})$$

for all $(\tilde{u}, \tilde{v}) \in \widetilde{E}$.

Then our result on the change of variables formula for double integration gives

$$\begin{aligned}
\iint_{\widetilde{\Phi}} f\, dS &= \iint_{\widetilde{E}} f(\widetilde{\Phi}(\tilde{u},\tilde{v})) \|(\widetilde{\Phi}_{\tilde{u}} \times \widetilde{\Phi}_{\tilde{v}})(\tilde{u},\tilde{v})\| d(\tilde{u},\tilde{v}) \\
&= \iint_{\widetilde{E}} f(\Phi \circ h(\tilde{u},\tilde{v})) \|(\Phi_u \times \Phi_v)(h(\tilde{u},\tilde{v}))\| |J(h)(\tilde{u},\tilde{v})| d(\tilde{u},\tilde{v}) \\
&= \iint_{E} f(\Phi(u,v)) \|(\Phi_u \times \Phi_v)(u,v)\| d(u,v) \\
&= \iint_{\Phi} f\, dS.
\end{aligned}$$

Considering $f := 1$, we see that the area of a smooth surface is invariant under reparametrization. $\square$

Examples: Let $E := [0, \pi] \times [-\pi, \pi]$, and define
$\Phi(\varphi, \theta) := (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$ for $(\varphi, \theta) \in E$.

If $\widetilde{E} := [-\pi, \pi] \times [0, \pi]$, and we define
$\widetilde{\Phi}(\theta, \varphi) := (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$ for $(\theta, \varphi) \in \widetilde{E}$,
then $\widetilde{\Phi}$ is a reparametrization of Φ since $\widetilde{\Phi}(\theta, \varphi) = \Phi(h(\theta, \varphi))$,
where $h : \widetilde{E} \to E$ is given by $h(\theta, \varphi) := (\varphi, \theta)$ with $J(h) = -1$.

Similarly, if $\widetilde{E} := [0, \pi/2] \times [-\pi/2, \pi/2]$, and we define
$\widetilde{\Phi}(\varphi, \theta) := (\sin 2\varphi \cos 2\theta, \sin 2\varphi \sin 2\theta, \cos 2\varphi)$ for $(\varphi, \theta) \in \widetilde{E}$,
then $\widetilde{\Phi}$ is a reparametrization of Φ since $\widetilde{\Phi}(\varphi, \theta) = \Phi(h(\varphi, \theta))$,
where $h : \widetilde{E} \to E$ is given by $h(\varphi, \theta) := (2\varphi, 2\theta)$ with $J(h) = 4$.

In both these cases,

$$\iint_{\widetilde{\Phi}} f \, dS = \iint_{\Phi} f \, dS$$

for each continuous scalar field f defined on the unit sphere
$S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$.

Geometric Surface

Let E be a subset of $\mathbb{R}^2$ having an area, and let $\varPhi : E \to \mathbb{R}^3$ be a parametrized surface. Then $S := \varPhi(E)$ is called a **geometric surface**, and $\varPhi$ is called a **parametrization** of S.

Suppose $\varPhi$ is smooth and one-one except possibly on a subset of E of two-dimensional content zero. In view of the reparametrization result for continuous scalar fields on parametrized surfaces proved above, we define

$$\iint_S f \, dS := \iint_{\varPhi} f \, dS = \iint_E f(\varPhi(u,v)) \|(\varPhi_u \times \varPhi_v)(u,v)\| d(u,v)$$

for a continuous scalar field f on S, and in particular,

$$\text{Area}\,(S) := \text{Area}\,(\varPhi) = \iint_E \|(\varPhi_u \times \varPhi_v)(u,v)\| d(u,v).$$

Surface Integral of a Vector Field

Let E be a subset of $\mathbb{R}^2$ having an area, and $\Phi : E \to \mathbb{R}^3$ be a smooth parametrized surface. Let $\boldsymbol{F}$ be a bounded vector field on $\Phi(E)$. We define the **surface integral** of $\boldsymbol{F}$ **across** Φ by

$$\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S} := \iint_{E} \boldsymbol{F}(\Phi(u,v)) \cdot (\Phi_u \times \Phi_v)(u,v)\, d(u,v),$$

provided the double integral on the right exists. In particular, if $\boldsymbol{F}$ is continuous, then $\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S}$ is well-defined.

In analogy with our differential notation $dS := \|\Phi_u \times \Phi_v\| d(u,v)$, we introduce the following differential notation:

$$d\boldsymbol{S} := (\Phi_u \times \Phi_v) d(u,v).$$

Let $\Phi(u, v) := (x(u, v), y(u, v), z(u, v))$ for $(u, v) \in E$. Then

$$\Phi_u \times \Phi_v = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} = \left(\frac{\partial(y, z)}{\partial(u, v)}, \frac{\partial(z, x)}{\partial(u, v)}, \frac{\partial(x, y)}{\partial(u, v)} \right).$$

If $\boldsymbol{F} := (P, Q, R)$, then we write the surface integral

$$\iint_\Phi \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_E \left(P \frac{\partial(y, z)}{\partial(u, v)} + Q \frac{\partial(z, x)}{\partial(u, v)} + R \frac{\partial(x, y)}{\partial(u, v)} \right) d(u, v)$$

as $\displaystyle\iint_\Phi P\, dy \wedge dz + Q\, dz \wedge dx + R\, dx \wedge dy$,

with the differential notation $dy \wedge dz := \dfrac{\partial(y, z)}{\partial(u, v)} d(u, v)$, etc.

If the surface Φ is regular, then

$$d\boldsymbol{S} = (\Phi_u \times \Phi_v) d(u, v) = \hat{\boldsymbol{n}} \, \|\Phi_u \times \Phi_v\| d(u, v) = \hat{\boldsymbol{n}}\, dS,$$

and we write the surface integral $\displaystyle\iint_\Phi \boldsymbol{F} \cdot d\boldsymbol{S}$ as $\displaystyle\iint_\Phi \boldsymbol{F} \cdot \hat{\boldsymbol{n}}\, dS$.

The algebraic properties of the double integral imply similar properties of the surface integral of a vector field.

Physical interpretation:

Let $\boldsymbol{F}$ be a vector field on a subset D of $\mathbb{R}^3$. Let $\Phi : E \to \mathbb{R}^3$ be a regular surface such that $\Phi(E) \subset D$. Then $\boldsymbol{F} \cdot \hat{\boldsymbol{n}}$ is the component of $\boldsymbol{F}$ in the direction of the unit normal $\hat{\boldsymbol{n}}$.

If $\boldsymbol{F}$ is an electric field, a magnetic field, or a velocity field, then the surface integral

$$\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{\Phi} \boldsymbol{F} \cdot \hat{\boldsymbol{n}}\, dS$$

represents the flux, that is, the rate of flow (per unit area) of the field $\boldsymbol{F}$ across the surface Φ. As a result, this integral is known as a **flux integral**.

Line Integrals of Scalar Fields and Vector Fields

Recall: Let $m \in \mathbb{N}$, and let $\boldsymbol{\gamma} : [\alpha, \beta] \to \mathbb{R}^m$ be a (piecewise smooth) path.

(i) Let $f : \boldsymbol{\gamma}([\alpha, \beta]) \to \mathbb{R}$ be a bounded scalar field. Then the **line integral** of f **along** $\boldsymbol{\gamma}$ is defined by

$$\int_{\boldsymbol{\gamma}} f\, ds := \int_{\alpha}^{\beta} f(\boldsymbol{\gamma}(t)) \|\boldsymbol{\gamma}'(t)\| dt,$$

provided the Riemann integral on the right side exists.

(ii) Let $\boldsymbol{F} : \boldsymbol{\gamma}([\alpha, \beta]) \to \mathbb{R}^m$ be a bounded vector field. Then the **line integral** of $\boldsymbol{F}$ **along** $\boldsymbol{\gamma}$ is defined by

$$\int_{\boldsymbol{\gamma}} \boldsymbol{F} \cdot d\boldsymbol{s} := \int_{\alpha}^{\beta} \boldsymbol{F}(\boldsymbol{\gamma}(t)) \cdot \boldsymbol{\gamma}'(t) dt,$$

provided the Riemann integral on the right side exists.

Surface Integrals of Scalar Fields and Vector Fields

Recall: Let E be a subset of $\mathbb{R}^2$ having an area, and let $\Phi : E \to \mathbb{R}^3$ be a smooth parametrized surface.

(i) Let $f : \Phi(E) \to \mathbb{R}$ be a bounded scalar field. Then the **surface integral** of f **across** Φ is defined by

$$\iint_{\Phi} f \, dS := \iint_{E} f(\Phi(u, v)) \|(\Phi_u \times \Phi_v)(u, v)\| d(u, v),$$

provided the double integral on the right exists.

(ii) Let $\boldsymbol{F} : \Phi(E) \to \mathbb{R}^3$ be a bounded vector field. Then the **surface integral** of $\boldsymbol{F}$ **across** Φ is defined by

$$\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S} := \iint_{E} \boldsymbol{F}(\Phi(u, v)) \cdot (\Phi_u \times \Phi_v)(u, v) \, d(u, v),$$

provided the double integral on the right exists.

Examples

(i) Let a subset E of $\mathbb{R}^2$ have an area, and let $f : E \to \mathbb{R}$ be a smooth function. Let the smooth parametrized surface $\Phi : E \to \mathbb{R}^3$ represent the graph of f, and let $\boldsymbol{F} : \Phi(E) \to \mathbb{R}^3$ be a continuous vector field. If $\boldsymbol{F} := (P, Q, R)$, then

$$\iint_\Phi \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_E (-P\, f_x - Q\, f_y + R) d(x, y)$$

since $d\boldsymbol{S} = (\Phi_x \times \Phi_y) d(x, y) = (-f_x, -f_y, 1) d(x, y)$.

For instance, let $E := [0, 1] \times [0, 1]$, $f(x, y) := x + y + 1$ for $(x, y) \in E$. If $\boldsymbol{F}(x, y, z) := (x^2, y^2, z)$ for $(x, y, z) \in \mathbb{R}^3$, then

$$\begin{aligned}\iint_\Phi \boldsymbol{F} \cdot d\boldsymbol{S} &= \iint_E \big(- x^2 - y^2 + (x + y + 1)\big) d(x, y) \\ &= \int_0^1 \left(\int_0^1 (x + y + 1 - x^2 - y^2) dy \right) dx \\ &= = \frac{1}{2} + \frac{1}{2} + 1 - \frac{1}{3} - \frac{1}{3} = \frac{4}{3}.\end{aligned}$$

(ii) Let $E := [-\pi, \pi] \times [0, h]$, and $\Phi(\theta, z) := (a\cos\theta, a\sin\theta, z)$ for $(\theta, z) \in E$. If $\boldsymbol{F}(x, y, z) := (y, z, x)$ for $(x, y, z) \in \mathbb{R}^3$, then

$$\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{E} (a^2 \cos\theta \sin\theta + z\, a \sin\theta + 0) d(\theta, z) = 0,$$

since $d\boldsymbol{S} = (\Phi_\theta \times \Phi_z) d(\theta, z) = (a\cos\theta, a\sin\theta, 0) d(\theta, z)$.

(iii) Let $E := [0, \pi] \times [-\pi, \pi]$, and let $\Phi(\varphi, \theta) := (a\sin\varphi\cos\theta, a\sin\varphi\sin\theta, a\cos\varphi)$ for $(\varphi, \theta) \in E$. If $\boldsymbol{F}(x, y, z) := (x, y, z)$ for $(x, y, z) \in \mathbb{R}^3$, then

$$\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{E} \Phi(\varphi, \theta) \cdot (a\sin\varphi) \Phi(\varphi, \theta) d(\varphi, \theta) = 4a^3\pi,$$

since $d\boldsymbol{S} = (\Phi_\varphi \times \Phi_\theta) d(\varphi, \theta) = a\sin\varphi\, \Phi(\varphi, \theta) d(\varphi, \theta)$.

On the other hand, let $\widetilde{E} := [-\pi, \pi] \times [0, \pi]$, and let $\widetilde{\Phi}(\theta, \varphi) := (a \sin\varphi \cos\theta, a \sin\varphi \sin\theta, a \cos\varphi)$ for $(\theta, \varphi) \in \widetilde{E}$. If $\boldsymbol{F}(x, y, z) := (x, y, z)$ for $(x, y, z) \in \mathbb{R}^3$ as above, then

$$\iint_{\widetilde{\Phi}} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{\widetilde{E}} \widetilde{\Phi}(\theta, \varphi) \cdot (-a \sin\varphi)\, \widetilde{\Phi}(\theta, \varphi) d(\theta, \varphi) = -4a^3\pi,$$

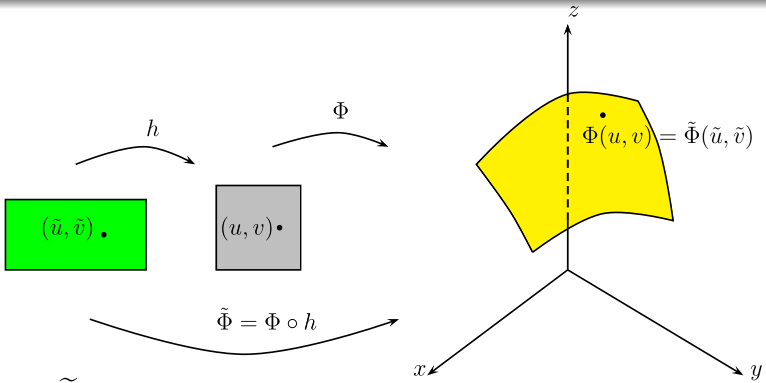
since $d\boldsymbol{S} = (\widetilde{\Phi}_\theta \times \widetilde{\Phi}_\varphi) d(\theta, \varphi) = -a \sin\varphi\, \widetilde{\Phi}(\theta, \varphi) d(\theta, \varphi)$.

Proposition

The surface integral of a continuous vector field across a smooth parametrized surface is invariant under a reparametrization only up to its sign.

Proof. Let $\Phi : E \to \mathbb{R}^3$ be a smooth parametrized surface, and let $\boldsymbol{F}$ be a continuous vector field on $S := \Phi(E)$. Suppose $\widetilde{\Phi} := \Phi \circ h$ be a reparametrization of Φ, where E and $\widetilde{E}$ are path-connected. By the chain rule, for all $(\tilde{u}, \tilde{v}) \in \widetilde{E}$,

$$(\widetilde{\Phi}_{\tilde{u}} \times \widetilde{\Phi}_{\tilde{v}})(\tilde{u}, \tilde{v}) = (\Phi_u \times \Phi_v)(h(\tilde{u}, \tilde{v}))\, J(h)(\tilde{u}, \tilde{v}).$$



Since $\widetilde{E}$ is path-connected, and $J(h)(\tilde{u}, \tilde{v}) \neq 0$ for all $(\tilde{u}, \tilde{v}) \in \widetilde{E}$, either $J(h)(\tilde{u}, \tilde{v}) > 0$ for all $(\tilde{u}, \tilde{v}) \in \widetilde{E}$ or $J(h)(\tilde{u}, \tilde{v}) < 0$ or all $(\tilde{u}, \tilde{v}) \in \widetilde{E}$. By the change of variables formula (which involves $|J(h)(\tilde{u}, \tilde{v})|$), we obtain $\iint_{\widetilde{\Phi}} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S}$ in the former case, while $\iint_{\widetilde{\Phi}} \boldsymbol{F} \cdot d\boldsymbol{S} = -\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S}$ in the latter. $\square$

Examples

Let $\widetilde{E} := \{(v, u) : (u, v) \in E\}$, and define $h(\tilde{u}, \tilde{v}) := (\tilde{v}, \tilde{u})$ for $(\tilde{u}, \tilde{v}) \in \widetilde{E}$. Then we see that $\widetilde{\Phi}(\tilde{u}, \tilde{v}) := \Phi(h(\tilde{u}, \tilde{v})) = \Phi(\tilde{v}, \tilde{u})$ for $(\tilde{u}, \tilde{v}) \in \widetilde{E}$. Since $J(h)(\tilde{u}, \tilde{v}) = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1 < 0$ for all $(\tilde{u}, \tilde{v}) \in \widetilde{E}$, we obtain

$$\iint_{\widetilde{\Phi}} \boldsymbol{F} \cdot d\boldsymbol{S} = -\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S}.$$

On the other hand, let $\widetilde{E} := \{(u/2, v/2) : (u, v) \in E\}$, and define $h(\tilde{u}, \tilde{v}) := (2\tilde{u}, 2\tilde{v})$ for $(\tilde{u}, \tilde{v}) \in \widetilde{E}$. Then we see that $\widetilde{\Phi}(\tilde{u}, \tilde{v}) := \Phi(h(\tilde{u}, \tilde{v})) = \Phi(2\tilde{u}, 2\tilde{v})$ for $(\tilde{u}, \tilde{v}) \in \widetilde{E}$. Since $J(h)(\tilde{u}, \tilde{v}) = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0$ for all $(\tilde{u}, \tilde{v}) \in \widetilde{E}$, we obtain

$$\iint_{\widetilde{\Phi}} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S}.$$

Opposite of a Parametrized Surface

Let $\Phi : E \to \mathbb{R}^3$ be a smooth parametrized surface, and let $\widetilde{E} := \{(v, u) : (u, v) \in E\}$. Define

$$\widetilde{\Phi}(\tilde{u}, \tilde{v}) := \Phi(\tilde{v}, \tilde{u}) \quad \text{for } (\tilde{u}, \tilde{v}) \in \widetilde{E}.$$

Then the parametrized surface $\widetilde{\Phi}$ is called the **opposite** of Φ, and it will be denoted by $\Phi^{\,\mathrm{op}}$. Clearly, $(\Phi^{\,\mathrm{op}})^{\,\mathrm{op}} = \Phi$.
Note that $\Phi^{\,\mathrm{op}}(\widetilde{E}) = \Phi(E)$.
Also, by the chain rule, for all $(\tilde{u}, \tilde{v}) \in \widetilde{E}$,

$$(\Phi_{\tilde{u}}^{\mathrm{op}} \times \Phi_{\tilde{v}}^{\mathrm{op}})(\tilde{u}, \tilde{v}) = \Phi_v(\tilde{v}, \tilde{u}) \times \Phi_u(\tilde{v}, \tilde{u}) = -(\Phi_u \times \Phi_v)(\tilde{v}, \tilde{u}).$$

Note: If $h(\tilde{u}, \tilde{v}) = (\tilde{v}, \tilde{u})$ for $(\tilde{u}, \tilde{v}) \in \widetilde{E}$, then $J(h)(\tilde{u}, \tilde{v}) = -1$.

If $\boldsymbol{F}$ is a continuous vector field on $\Phi(E)$, then we obtain

$$\iint_{\Phi^{\,\mathrm{op}}} \boldsymbol{F} \cdot d\boldsymbol{S} = -\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S}.$$

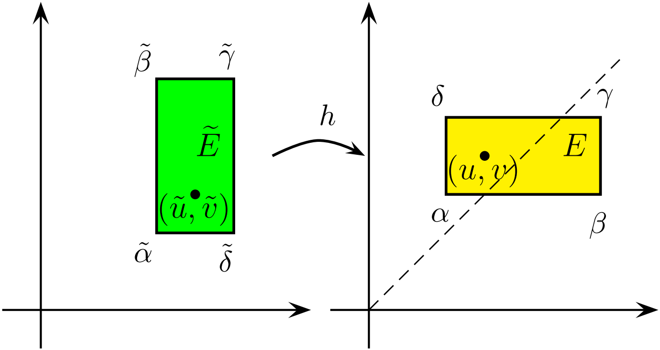


Figure: Parameter domain $\widetilde{E}$ of the opposite Φ^{op} of a parametrized surface Φ is obtained by reflecting the parameter domain E of Φ in the dashed line $u = v$.
Note that the anti-clockwise direction of the corners $\alpha, \beta, \gamma, \delta$ of E is changed to the clockwise direction of the corresponding corners $\tilde\alpha, \tilde\beta, \tilde\gamma, \tilde\delta$ of $\widetilde{E}$.

Orientable Surfaces

Consider a regular parametrized surface $\Phi : E \to \mathbb{R}^3$, so that $(\Phi_u \times \Phi_v)(u, v) \neq 0$ for all $(u, v) \in E$. Suppose the function Φ is one-one. Then

$$\hat{\boldsymbol{n}}\,(u, v) := \frac{(\Phi_u \times \Phi_v)(u, v)}{\|(\Phi_u \times \Phi_v)(u, v)\|}$$

is a unit normal vector at a point $P := \Phi(u, v)$ on the geometric surface $S := \Phi(E)$, and so is $-\hat{\boldsymbol{n}}\,(u, v)$.

Next, let D be a subset of $\mathbb{R}^3$, and consider a function $F : D \to \mathbb{R}$. Suppose the equation $F(x, y, z) = 0$ implicitly defines a surface S in $\mathbb{R}^3$. If F has partial derivatives at all points of D and $\nabla F \neq 0$ on D, then

$$\frac{\nabla F(P)}{\|\nabla F(P)\|}$$

is a unit normal vector at $P \in S$, and so is $-\nabla F(P)/\|\nabla F(P)\|$.

Let S be a (geometric) surface in $\mathbb{R}^3$. If there exists a continuous function $P \longmapsto \boldsymbol{n}(P)$ from S to $\mathbb{R}^3$ such that $\boldsymbol{n}(P)$ is a unit normal vector at P, then we say that the surface S is **orientable**, and such an assignment is called an **orientation** of S. A surface S with a given orientation is called **oriented**.

Suppose the surface S is oriented. If a parametrization Φ of S yields the same unit normal vectors as specified by the given orientation, then Φ is called an **orientation-preserving parametrization** of S, and Φ^{op} is called an **orientation-reversing parametrization** of S.

Suppose $\Phi : E \to \mathbb{R}^3$ is a smooth orientation-preserving parametrization of S which is one-one except possibly on a subset of E of two-dimensional content zero. Let $\boldsymbol{F} : S \to \mathbb{R}^3$ be a continuous vector field. Then we define

$$\iint_S \boldsymbol{F} \cdot d\boldsymbol{S} := \iint_\Phi \boldsymbol{F} \cdot d\boldsymbol{S}.$$

Examples

(i) Let $E \subset \mathbb{R}^2$ have an area, and $f: E \to \mathbb{R}$ be a smooth scalar field. Consider the graph $S := \{(x, y, f(x, y)) : (x, y) \in E\}$ of f. For $P := (x, y, z) \in S$, define
$\boldsymbol{n}(P) := (-f_x(P), -f_y(P), 1)/\|(-f_x(P), -f_y(P), 1)\|$.
This continuous assignment of upward unit normal vectors gives an orientation of S. Hence S is orientable. Clearly, the parametrization of S given by $\Phi(x, y) := (x, y, f(x, y))$ for $(x, y) \in E$, is orientation-preserving.

(ii) Let $S := \{(x, y, z) :\in \mathbb{R}^3 : x^2 + y^2 = a^2 \text{ and } 0 \le z \le h\}$.
For $P := (x, y, z) \in S$, define $\boldsymbol{n}(P) := (-x/a, -y/a, 0)$.
This continuous assignment of inward unit normal vectors gives an orientation of S. Hence the cylinder S is orientable.

Let $E := [0, h] \times [-\pi, \pi]$ and

$\Psi(z, \theta) := (a \cos\theta, a \sin\theta, z)$ for $(z, \theta) \in E$.

If $P := \Psi(z, \theta) = (x, y, z) \in S$, then

$$\frac{\Psi_z \times \Psi_\theta}{\|\Psi_z \times \Psi_\theta\|}(z, \theta) = \frac{(-a\cos\theta, -a\sin\theta, 0)}{a} = \left(-\frac{x}{a}, -\frac{y}{a}, 0\right) = \boldsymbol{n}(P).$$

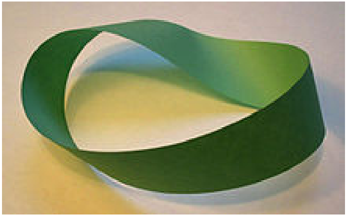
Hence Ψ is an orientation-preserving parametrization.

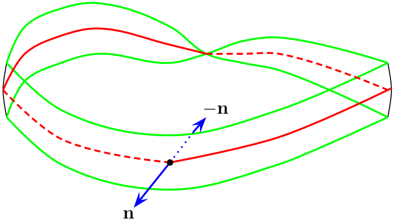
(iii) Let $S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = a^2\}$. For $P := (x, y, z) \in S$, define $\boldsymbol{n}(P) := (x/a, y/a, z/a)$. This continuous assignment of outward unit normal vectors gives an orientation of the sphere S. Hence S is orientable. Let $E := (0, \pi) \times [-\pi, \pi]$ and $\Phi(\varphi, \theta) := (a \sin\varphi \cos\theta, a \sin\varphi \sin\theta, a \cos\varphi)$ for $(\varphi, \theta) \in E$. If $P := \Phi(\varphi, \theta) = (x, y, z) \in S$, then

$$\frac{\Phi_\varphi \times \Phi_\theta}{\|\Phi_\varphi \times \Phi_\theta\|}(\varphi, \theta) = \frac{(a \sin\varphi)\Phi(\varphi, \theta)}{a^2 \sin\varphi} = \left(\frac{x}{a}, \frac{y}{a}, \frac{z}{a}\right) = \boldsymbol{n}(P).$$

Hence Φ is an orientation-preserving parametrization of $S \setminus \{(0, 0, \pm a)\}$.

(iv) Möbius strip





(Artwork by David Bebbennick)

A parametrization of a Möbius strip is given as follows. Let $E := [-\pi, \pi] \times [-1, 1]$, and for $(u, v) \in E$, let $\Phi(u, v) := (x(u, v), y(u, v), z(u, v))$, where

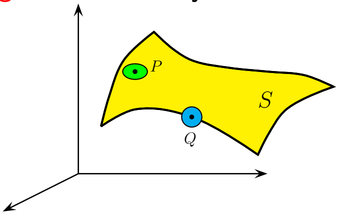
$$\begin{aligned} x(u, v) &:= \big(1 + (v/2)\cos(u/2)\big)\cos u, \\ y(u, v) &:= \big(1 + (v/2)\cos(u/2)\big)\sin u, \\ z(u, v) &:= (v/2)\sin(u/2). \end{aligned}$$

It can be shown that a Möbius strip is not orientable.

Intrinsic Boundary of a Surface

Let S be a (geometric) surface in $\mathbb{R}^3$, and suppose S is a closed subset of $\mathbb{R}^3$. Typically, no point of S is an interior point of S as a subset of $\mathbb{R}^3$, and so $\partial S = S$.

Let P be a point on S. It is called an **intrinsic interior point** of S if there is a two dimensional disk-like neighbourhood of P which is contained in S, that is, there is a curvilinear patch on S around P which can be flattened into a disk. Otherwise, the point P is called an **intrinsic boundary point** of S. We shall denote the set of all intrinsic boundary points of S by $\eth S$. It consists of the 'edges' of S, if any.



In the figure on the previous page, P is an intrinsic interior point of S, whereas Q is an intrinsic boundary point of S.

If $\eth S = \emptyset$, then S is called a **surface without edges**. It is also known as a **closed surface**. (Compare the definition of a closed geometric curve.)

If S is a planar surface, that is, $S := D$, a subset of $\mathbb{R}^2$, then $\eth S$ of S coincides with the usual boundary ∂D of D.

Examples

(i) If $S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = a^2\}$, then $\eth S = \emptyset$.

(ii) If S is the union of the three sets
$[a, b]\times[c, d]\times\{p, q\}, [a, b]\times\{c, d\}\times[p, q], \{a, b\}\times[c, d]\times[p, q]$,
then $\eth S = \emptyset$.

(iii) If $S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = a^2, \ 0 \le z \le h\}$,
then $\eth S$ is the union of the two sets
$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = a^2 \text{ and } z = 0\}$ and
$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = a^2 \text{ and } z = h\}$.

MA 105 : Calculus
Division 1, Lecture 25

Prof. Sudhir R. Ghorpade
IIT Bombay

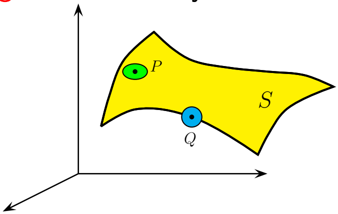
Recap of the previous lecture

- Surface integral of a scalar field. Examples
- Reparametrization of a surface
- Invariance of surface integrals of scalar fields under reparametrixations
- Notion of a geometric surface
- Surface integral of a vector field.
- Differential notaion (involving wedge products) and physical interpretation. Examples
- Invariance of surface integrals of vector fields up to a sign under reparametrizations
- Opposite of a parametrized surface
- Oriented surfaces
- Orientation-preserving/reversing parametrizations
- Examples, including Mobius strip and its paramerization
- Intrinsic boundary of a surface

Intrinsic Boundary of a Surface

Let S be a (geometric) surface in $\mathbb{R}^3$, and suppose S is a closed subset of $\mathbb{R}^3$. Typically, no point of S is an interior point of S as a subset of $\mathbb{R}^3$, and so $\partial S = S$.

Let P be a point on S. It is called an **intrinsic interior point** of S if there is a two dimensional disk-like neighbourhood of P which is contained in S, that is, there is a curvilinear patch on S around P which can be flattened into a disk. Otherwise, the point P is called an **intrinsic boundary point** of S. We shall denote the set of all intrinsic boundary points of S by $\eth S$. It consists of the 'edges' of S, if any.



In the figure on the previous page, P is an intrinsic interior point of S, whereas Q is an intrinsic boundary point of S.

If $\eth S = \emptyset$, then S is called a **surface without edges**. It is also known as a **closed surface**. (Compare the definition of a closed geometric curve.)

If S is a planar surface, that is, $S := D$, a subset of $\mathbb{R}^2$, then $\eth S$ of S coincides with the usual boundary ∂D of D.

Examples

(i) If $S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = a^2\}$, then $\eth S = \emptyset$.

(ii) If S is the union of the three sets
$[a, b]\times[c, d]\times\{p, q\}, [a, b]\times\{c, d\}\times[p, q], \{a, b\}\times[c, d]\times[p, q]$,
then $\eth S = \emptyset$.

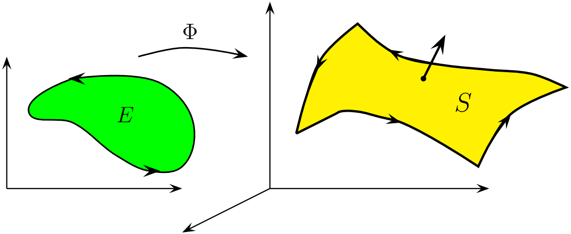
(iii) If $S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = a^2,\ 0 \le z \le h\}$,
then $\eth S$ is the union of the two sets
$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = a^2 \text{ and } z = 0\}$ and
$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = a^2 \text{ and } z = h\}$.

Orientation of the Intrinsic Boundary

An orientation of a surface induces an orientation on its intrinsic boundary as follows.

Suppose S is an oriented surface, and let $\boldsymbol{n}(P)$ be the specified unit normal vector at a point P on S. As we travel along any part of the intrinsic boundary $\eth S$ of the surface S with $\boldsymbol{n}(P)$ as our upright direction, suppose the surface S lies on our left. This gives the **induced orientation** on $\eth S$.

Note: If E is a subset of $\mathbb{R}^2$, and $\Phi : E \to \mathbb{R}^3$ is a smooth orientation-preserving parametrization of the surface S, then $\Phi(\partial E) = \eth S$, and the induced orientation on the intrinsic boundary $\eth S$ of S corresponds to the positive orientation of the boundary ∂E of the parameter set $E \subset \mathbb{R}^2$ defined earlier.



Example

Suppose the unit upper hemisphere $S := \{(x, y, z) : x^2 + y^2 + z^2 = 1 \text{ and } z \geq 0\}$ is oriented by the unit normal vector $\boldsymbol{n}(P) := (x, y, z)$ for $P := (x, y, z) \in S$.

Let $E := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ and $\Phi(x, y) := \left(x, y, \sqrt{1 - x^2 - y^2}\right)$ for $(x, y) \in E$. Then Φ is an orientation-preserving parametrization of S, and the induced orientation on $\partial S = \{(x, y, z) : x^2 + y^2 + z^2 = 1 \text{ and } z = 0\}$ corresponds to the counterclockwise orientation of $\partial E := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

The Stokes Theorem

An important result of Stokes relates line integrals along the edges of a surface to flux integrals across the surface. It is a 'curved version' of the Green theorem, and can be viewed as an analogue of Part II of the Fundamental Theorem of Calculus. The proof of the Stokes theorem heavily depends on the proof of the Green theorem.

Theorem

Let S be an oriented (geometric) surface in $\mathbb{R}^3$ which is a closed and bounded subset of $\mathbb{R}^3$, which is piecewise C^2 and whose piecewise smooth intrinsic boundary ∂S consists of a finite number of nonintersecting simple closed curves along with the induced orientation. Let $\boldsymbol{F}$ be a smooth vector field on an open subset containing S. Then

$$\int_{\partial S} \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_S (\operatorname{curl} \boldsymbol{F}) \cdot d\boldsymbol{S}.$$

Sketch of Proof:

Let the (geometric) surface S correspond to $\Phi : E \to \mathbb{R}^3$, where E is a closed and bounded subset of $\mathbb{R}^2$. Suppose $\Phi(u, v) := (x(u, v), y(u, v), z(u, v))$ for $(u, v) \in E$, and $\boldsymbol{F} := (P, Q, R)$ on an open subset of $\mathbb{R}^3$ containing S. Then

$$\int_{\eth S} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\eth S} P\,dx + Q\,dy + R\,dz.$$

Let ∂E be given by a path $\boldsymbol{\gamma} := (u, v)$. Then $\eth S$ is given by the path $\Phi \circ \boldsymbol{\gamma}$. Since, by the chain rule, $(x \circ \boldsymbol{\gamma})' = x_u u' + x_v v'$,

$$\int_{\eth S} P\,dx = \int_{\partial E} (P \circ \Phi)\,(x_u du + x_v dv).$$

Let us write $g := P \circ \Phi$ on E. Then

$$\int_{\eth S} P\,dx = \int_{\partial E} (g\,x_u) du + (g\,x_v) dv.$$

The Green theorem shows that

$$\begin{aligned}\int_{\partial E} (g\, x_u) du + (g\, x_v) dv &= \iint_E \left((g\, x_v)_u - (g\, x_u)_v\right) d(u,v) \\ &= \iint_E (g_u\, x_v - g_v\, x_u) d(u,v)\end{aligned}$$

since $g\, x_{vu} - g\, x_{uv} = 0$ by the Mixed Partials Theorem.
(Note: $\Phi(u,v) = (x(u,v), y(u,v), z(u,v))$ is piecewise C^2.)

But since $g(u,v) = P(x(u,v), y(u,v), z(u,v))$, we obtain

$$g_u = P_x\, x_u + P_y\, y_u + P_z\, z_u, \quad g_v = P_x\, x_v + P_y\, y_v + P_z\, z_v,$$

by the chain rule. Hence

$$\begin{aligned}g_u\, x_v - g_v\, x_u &= -P_y(x_u\, y_v - x_v\, y_u) + P_z(z_u\, x_v - z_v\, x_u) \\ &= -P_y \frac{\partial(x,y)}{\partial(u,v)} + P_z \frac{\partial(z,x)}{\partial(u,v)}.\end{aligned}$$

It follows that

$$\int_{\partial S} P\,dx = \iint_E \left(-P_y \frac{\partial(x,y)}{\partial(u,v)} + P_z \frac{\partial(z,x)}{\partial(u,v)} \right) d(u,v).$$

In a similar manner, we obtain

$$\int_{\partial S} Q\,dy = \iint_E \left(-Q_z \frac{\partial(y,z)}{\partial(u,v)} + Q_x \frac{\partial(x,y)}{\partial(u,v)} \right) d(u,v).$$

and

$$\int_{\partial S} R\,dz = \iint_E \left(-R_x \frac{\partial(z,x)}{\partial(u,v)} + R_y \frac{\partial(y,z)}{\partial(u,v)} \right) d(u,v).$$

Hence $\displaystyle\int_{\partial S} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\partial S} P\,dx + Q\,dy + R\,dz$ is equal to

$$\iint_E \left((R_y - Q_z)\frac{\partial(y,z)}{\partial(u,v)} + (P_z - R_x)\frac{\partial(z,x)}{\partial(u,v)} + (Q_x - P_y)\frac{\partial(x,y)}{\partial(u,v)} \right) d(u,v).$$

On the other hand,

$$\iint_S (\operatorname{curl} \boldsymbol{F}) \cdot d\boldsymbol{S} = \iint_E \operatorname{curl} \boldsymbol{F}(\Phi(u,v)) \cdot (\Phi_u \times \Phi_v)\, d(u,v),$$

where

$$\operatorname{curl} \boldsymbol{F} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ P & Q & R \end{vmatrix} = (R_y - Q_z,\ P_z - R_x,\ Q_x - P_y),$$

and

$$\Phi_u \times \Phi_v = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} = \left(\frac{\partial(y,z)}{\partial(u,v)}, \frac{\partial(z,x)}{\partial(u,v)}, \frac{\partial(x,y)}{\partial(u,v)} \right).$$

Thus

$$\int_{\partial S} \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_S (\operatorname{curl} \boldsymbol{F}) \cdot d\boldsymbol{S}, \quad \text{as desired.} \quad \square$$

The conclusion of the Stokes theorem is also written as follows:

$$\int_{\partial S} P\,dx + Q\,dy + R\,dz$$
$$= \iint_S (R_y - Q_z)\,dy \wedge dz + (P_z - R_x)\,dz \wedge dx + (Q_x - P_y)\,dx \wedge dy.$$

Here is a version of the Stokes theorem when S does not have intrinsic boundary.

Corollary

Let S be a closed oriented smooth surface in $\mathbb{R}^3$ (and so $\partial S = \emptyset$). Suppose $\boldsymbol{F}$ is a smooth vector field on an open subset containing S. Then $\iint_S (\text{curl}\,\boldsymbol{F}) \cdot d\boldsymbol{S} = 0$.

Proof: Introduce a hole in S by cutting out a small piece along a smooth simple closed curve C on S. Let S_1 denote the part of S cut out, and let S_2 denote the remaining part of S.

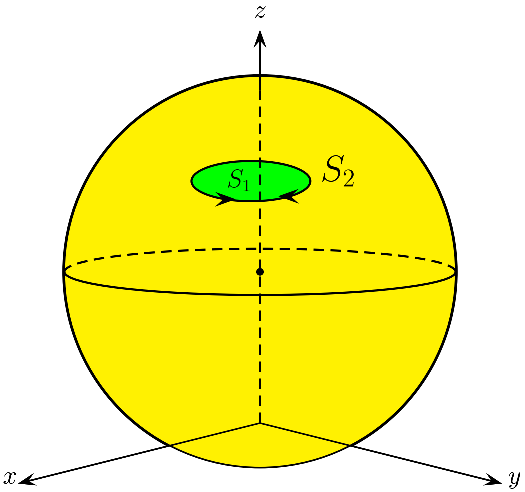


Figure: The Stokes theorem for S with $\partial S = \emptyset$.

Then $$\iint_S (\text{curl}\, \boldsymbol{F}) \cdot d\boldsymbol{S} = \iint_{S_1} (\text{curl}\, \boldsymbol{F}) \cdot d\boldsymbol{S} + \iint_{S_2} (\text{curl}\, \boldsymbol{F}) \cdot d\boldsymbol{S}.$$

by the domain additivity. Now the Stokes theorem shows that

$$\iint_{S_1} (\text{curl}\, \boldsymbol{F}) \cdot d\boldsymbol{S} = \int_{\eth S_1} \boldsymbol{F} \cdot d\boldsymbol{s} \text{ and } \iint_{S_2} (\text{curl}\, \boldsymbol{F}) \cdot d\boldsymbol{S} = \int_{\eth S_2} \boldsymbol{F} \cdot d\boldsymbol{s}.$$

We observe that the intrinsic boundary $\eth S_1$ of S_1 is the closed curve C with the orientation induced by the orientation on S_1. Since $\eth S = \emptyset$, the intrinsic boundary $\eth S_2$ of S_2 is also C. But the orientations induced on C by the orientations on S_1 and on S_2 are opposite. Hence

$$\iint_S (\text{curl}\, \boldsymbol{F}) \cdot d\boldsymbol{S} = \int_{\eth S_1} \boldsymbol{F} \cdot d\boldsymbol{s} + \int_{\eth S_2} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_C \boldsymbol{F} \cdot d\boldsymbol{s} - \int_C \boldsymbol{F} \cdot d\boldsymbol{s} = 0.$$

$\square$

Green's Theorem: Let D be a closed and bounded subset of $\mathbb{R}^2$ such that the boundary ∂D of D consists of a finite number of nonintersecting simple closed piecewise smooth curves, and suppose it is positively oriented. Let $\boldsymbol{F} := (P, Q)$ be a smooth vector field on an open subset Ω of $\mathbb{R}^2$ containing D. Then

$$\int_{\partial D} P\, dx + Q\, dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) d(x, y).$$

Stokes' Theorem: Let S be an oriented (geometric) surface in $\mathbb{R}^3$ which is a closed and bounded subset of $\mathbb{R}^3$, which is piecewise C^2 and whose piecewise smooth intrinsic boundary $\eth S$ consists of a finite number of nonintersecting simple closed curves along with the induced orientation. Let $\boldsymbol{F}$ be a smooth vector field on an open subset containing S. Then

$$\int_{\eth S} \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_S (\operatorname{curl} \boldsymbol{F}) \cdot d\boldsymbol{S}.$$

Surface-Independence

Let $E \subset \mathbb{R}^2$, and consider a parametrized surface $\Phi : E \to \mathbb{R}^3$. Let $D \subset \mathbb{R}^3$. We say that Φ lies in D if $\Phi(E) \subset D$.

Let $\boldsymbol{F} : D \to \mathbb{R}^3$ be a continuous vector field. We say that flux integrals of $\boldsymbol{F}$ are **surface-independent** in D if

$$\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{\widetilde{\Phi}} \boldsymbol{F} \cdot d\boldsymbol{S}$$

whenever the piecewise C^2 parametrized surfaces $\Phi : E \to \mathbb{R}^3$ and $\widetilde{\Phi} : \widetilde{E} \to \mathbb{R}^3$ lie in D, and if $S := \Phi(E)$, $\widetilde{S} := \widetilde{\Phi}(\widetilde{E})$, then S and $\widetilde{S}$ are oriented (geometric) surfaces and have the same intrinsic boundary which consists of a finite number of nonintersecting simple closed curves and on which the induced orientation is the same. Note that the geometric surfaces S and $\widetilde{S}$ need not be the same.

(Compare the definition of path-independence of a vector field given in Lecture 31.)

Examples: Let $E := [0, \pi/2] \times [-\pi, \pi]$, and let $\Phi(\varphi, \theta) := (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$ for $(\varphi, \theta) \in E$. Also, let $\widetilde{E} := \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 \le 1\}$, and let $\widetilde{\Phi}(u, v) := (u, v, 0)$ for $(u, v) \in \widetilde{E}$. Then the oriented intrinsic boundaries of $S := \Phi(E)$ and $\widetilde{S} := \widetilde{\Phi}(E)$ are the same, namely $\{(x, y, 0) \in \mathbb{R}^3 : x^2 + y^2 = 1\}$, traced counterclockwise as seen from $(0, 0, 2)$.

(i) Let $\boldsymbol{F}(x, y, z) := (y, z, x)$ for $(x, y, z) \in \mathbb{R}^3$. Then $\iint_\Phi \boldsymbol{F} \cdot d\boldsymbol{S}$ is equal to

$$\iint_E (\sin\varphi\sin\theta, \cos\varphi, \sin\varphi\cos\theta) \cdot (\sin\varphi)\Phi(\varphi, \theta) d(\varphi, \theta)$$

which is equal to 0 since

$$\int_{-\pi}^{\pi} \sin\theta\cos\theta\, d\theta = \int_{-\pi}^{\pi} \sin\theta\, d\theta = \int_{-\pi}^{\pi} \cos\theta\, d\theta = 0.$$

Also, since $\widetilde{\Phi}_u \times \widetilde{\Phi}_v = (1, 0, 0) \times (0, 1, 0) = (0, 0, 1)$, we obtain

$$\iint_{\widetilde{\Phi}} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{\widetilde{E}} (v, 0, u) \cdot (0, 0, 1) d(u, v) = \iint_{\widetilde{E}} u \, d(u, v)$$
$$= \int_{-\pi}^{\pi} \int_0^1 (r \cos\theta) r \, dr \, d\theta = 0.$$

In fact, we shall presently see that flux integrals of $\boldsymbol{F}$ are surface-independent in $\mathbb{R}^3$.

(ii) Let $\boldsymbol{F}(x, y, z) := (x, y, z)$ for $(x, y, z) \in \mathbb{R}^3$. Then

$$\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{E} \Phi(\varphi, \theta) \cdot (\sin\varphi) \Phi(\varphi, \theta) d(\varphi, \theta) = 2\pi.$$

On the other hand,

$$\iint_{\widetilde{\Phi}} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{\widetilde{E}} (u, v, 0) \cdot (0, 0, 1) d(u, v) = 0.$$

Thus flux integrals of $\boldsymbol{F}$ are not surface-independent in $\mathbb{R}^3$.

Theorem

Let $D \subset \mathbb{R}^3$ be open, and let $\boldsymbol{F}$ be a continuous curl field on D. Then flux integrals of $\boldsymbol{F}$ are surface-independent in D.

Proof: Let $\boldsymbol{G}$ be a smooth vector field on D such that $\boldsymbol{F} = \operatorname{curl} \boldsymbol{G}$ on D. Let $\Phi : E \to \mathbb{R}^3$ be a piecewise C^2 parametrized surface in D, and let the intrinsic boundary $\eth S$ of the oriented surface $S := \Phi(E)$ consist of a finite number of nonintersecting simple closed curves.

If $\eth S$ is given the induced orientation, then by the Stokes theorem,

$$\iint_{\Phi} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{S} \operatorname{curl} \boldsymbol{G} \cdot d\boldsymbol{S} = \int_{\eth S} \boldsymbol{G} \cdot d\boldsymbol{s},$$

which clearly depends only on $\boldsymbol{G}$, and on $\eth S$ and its orientation, but not on Φ. Hence the result follows. $\square$

If $\boldsymbol{F}$ is a curl field, then the evaluation of a surface integral $\iint_\Phi \boldsymbol{F} \cdot d\boldsymbol{S}$ can be simplified by replacing Φ by a parametrized surface having the same oriented intrinsic boundary. If we actually know a vector field $\boldsymbol{G}$ such that $\operatorname{curl} \boldsymbol{G} = \boldsymbol{F}$, then we can reduce our problem to the calculation of a line integral.

Example: Let $\boldsymbol{F}(x, y, z) := (y, z, x)$ for $(x, y, z) \in \mathbb{R}^3$. Consider $\boldsymbol{G}(x, y, z) := (z^2/2, x^2/2, y^2/2)$ for $(x, y, z) \in \mathbb{R}^3$. Then $\operatorname{curl} \boldsymbol{G} = \boldsymbol{F}$ on $\mathbb{R}^3$, and so flux integrals of $\boldsymbol{F}$ are surface-independent in $\mathbb{R}^3$.
Let $E := [0, \pi/2] \times [-\pi, \pi]$, and for $(\varphi, \theta) \in E$, let $\Phi(\varphi, \theta) := (\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi)$. Let $S := \Phi(E)$. Then

$$\begin{aligned}\iint_\Phi \boldsymbol{F} \cdot d\boldsymbol{S} &= \iint_S \operatorname{curl} \boldsymbol{G} \cdot d\boldsymbol{S} = \int_{\partial S} \boldsymbol{G} \cdot d\boldsymbol{s} \\ &= \frac{1}{2}\int_{-\pi}^{\pi} (0, \cos^2 t, \sin^2 t) \cdot (-\sin t, \cos t, 0)\, dt = 0\end{aligned}$$

since $\cos^3 t = (\cos 3t + 3\cos t)/4$ for $t \in [-\pi, \pi]$.

It is important to know whether a continuous vector field $\boldsymbol{F}$ is in fact a curl field, that is, whether there is a smooth vector field $\boldsymbol{G}$ such that $\boldsymbol{F} = \operatorname{curl} \boldsymbol{G}$.
We already know a necessary condition when $\boldsymbol{F}$ is smooth:
If $\boldsymbol{F}$ is smooth and $\boldsymbol{F} = \operatorname{curl} \boldsymbol{G}$ on an open subset D of $\mathbb{R}^3$, then $\operatorname{div}(\boldsymbol{F}) = 0$ on D.

Examples:
(i) Let $D := \mathbb{R}^3$, and $\boldsymbol{F} := (P, Q, R)$, where $P(x, y, z) := x$, $Q(x, y, z) := y$ and $R(x, y, z) := z$ for $(x, y, z) \in D$. Since $\operatorname{div} \boldsymbol{F} = 3 \neq 0$, the vector field $\boldsymbol{F}$ is not a curl field on D.

(ii) Even if $\operatorname{div} \boldsymbol{F} = 0$ on an open subset D of $\mathbb{R}^3$, the vector field $\boldsymbol{F}$ may not be a curl field on D. For instance, let $D := \{(x, y, z) \in \mathbb{R}^3 : (x, y, z) \neq (0, 0, 0)\}$, and let $\boldsymbol{F} := (P, Q, R)$, where for $(x, y, z) \in D$,

$$P := \frac{x}{(x^2+y^2+z^2)^{3/2}}, \; Q := \frac{y}{(x^2+y^2+z^2)^{3/2}}, \; R := \frac{z}{(x^2+y^2+z^2)^{3/2}}.$$

Then $(x^2+y^2+z^2)^3 \operatorname{div} \boldsymbol{F} = (x^2+y^2+z^2)^3(P_x+Q_y+R_z) = 3(x^2+y^2+z^2)^{3/2} - 3(x^2+y^2+z^2)^{1/2}(x^2+y^2+z^2) = 0.$

Thus $\operatorname{div} \boldsymbol{F} = 0$ on D. But $\boldsymbol{F}$ is not a curl field on D since the standard unit sphere $S := \Phi([0,\pi]\times[-\pi,\pi])$ lies in D, and

$$\begin{aligned}\iint_{\Phi} \boldsymbol{F}\cdot d\boldsymbol{S} &= \iint_{E} \Phi(\varphi,\theta)\cdot(\sin\varphi)\Phi(\varphi,\theta)\,d(\varphi,\theta) \\ &= \int_{-\pi}^{\pi}\left(\int_0^{\pi} \sin\varphi\, d\varphi\right) d\theta = 4\pi \neq 0.\end{aligned}$$

But if $\boldsymbol{F} = \operatorname{curl} \boldsymbol{G}$, then $\iint_{\Phi} \boldsymbol{F}\cdot d\boldsymbol{S} = \iint_{S} \operatorname{curl} \boldsymbol{G}\cdot d\boldsymbol{S} = 0$ since $\partial S = \emptyset$. (Recall the corollary of the Stokes theorem.)

If $D := I\times J\times K$, where I, J, K are open intervals in $\mathbb{R}$, and $\boldsymbol{F}$ is a smooth vector field on D such that $\operatorname{div} \boldsymbol{F} = 0$ on D, then $\boldsymbol{F}$ is indeed a curl field. In fact, this also holds if D is an open subset of $\mathbb{R}^3$ such that every closed surface in D is the boundary of a solid lying entirely in D.

Let $D := I \times J \times K$, where I, J, K are open intervals in $\mathbb{R}$. Suppose $\boldsymbol{F} := (P, Q, R)$ is a smooth vector field on D such that $\operatorname{div} \boldsymbol{F} = 0$ on D. We find a vector field $\boldsymbol{G} := (L, M, N)$ on D such that $\boldsymbol{F} = \operatorname{curl} \boldsymbol{G}$, that is,

$$(P, Q, R) = (N_y - M_z, L_z - N_x, M_x - L_y).$$

Let $x_0 \in I$, $z_0 \in K$.

Define $L := 0$ and $N(x, y, z) := -\int_{x_0}^{x} Q(u, y, z)du$.

Then $L_z - N_x = 0 + Q = Q$. Next, define $M(x, y, z) := \int_{x_0}^{x} R(u, y, z)du - \int_{z_0}^{z} P(x_0, y, v)dv$. Then $M_x - L_y = R - 0 = R$.

Finally, by differentiating under the integral sign,

$$\begin{aligned} N_y(x, y, z) &= -\int_{x_0}^{x} Q_y(u, y, z)du = \int_{x_0}^{x} (P_x + R_z)(u, y, z)du \\ &= P(x, y, z) - P(x_0, y, z) + \int_{x_0}^{x} R_z(u, y, z)du, \end{aligned}$$

and

$$M_z(x,y,z) = \int_{x_0}^{x} R_z(u,y,z)du - P(x_0,y,z).$$

Hence $N_y - M_z = P$. Thus $\boldsymbol{F} = \operatorname{curl} \boldsymbol{G}$.

Further, if there is a vector field $\boldsymbol{H}$ on D such that $\boldsymbol{F} = \operatorname{curl} \boldsymbol{H}$ as well, then $\operatorname{curl}(\boldsymbol{H} - \boldsymbol{G}) = 0$ on D, and so $\boldsymbol{H} = \boldsymbol{G} + \nabla f$, where f is a smooth scalar field on D, as we have mentioned while considering 'path-independence' of vector fields in Lecture 32. This will follow as a consequence of the Stokes theorem since D is simply connected.

Example:

Let $\boldsymbol{F} := (P,Q,R)$, where $P := x$, $Q = -2\,y$ and $R := z$ for $(x,y,z) \in \mathbb{R}^3$. Clearly, $\operatorname{div} \boldsymbol{F} = 0$ on $\mathbb{R}^3$.
Let $x_0 := 0 =: z_0$. Define $L := 0$, $N := 2\,x\,y$ and $M := z\,x$ as above. If we let $\boldsymbol{G} := (0, z\,x, 2\,x\,y)$, then $\boldsymbol{F} = \operatorname{curl} \boldsymbol{G}$. Also, if $\operatorname{curl} \boldsymbol{H} = \boldsymbol{F}$, then $\boldsymbol{H} = \boldsymbol{G} + (f_x, f_y, f_z)$, where f is a smooth scalar field on $\mathbb{R}^3$.

MA 105 : Calculus
Division 1, Lecture 26

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Orientation of the intrinsic boundary
- The Stokes Theorem
- Sketch of proof of Stokes Theorem
- Vanishing of surface integrals of curl fields as a Corollary
- Surface-independence. Examples
- Surface-independence of flux integrals of continuous curl fields
- Necessary condition for a smooth vector field to be a curl field
- Example to show that this necessary condition is not sufficient.
- Sufficiency in case of domains of the type I x J x K, where I, J, K are intervals

Let us recall the statement of Stokes' Theorem.

Let S be an oriented (geometric) surface in $\mathbb{R}^3$ which is a closed and bounded subset of $\mathbb{R}^3$, which is piecewise C^2 and whose piecewise smooth intrinsic boundary $\eth S$ consists of a finite number of nonintersecting simple closed curves along with the induced orientation. Let $\boldsymbol{F}$ be a smooth vector field on an open subset containing S. Then

$$\int_{\eth S} \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_S (\operatorname{curl} \boldsymbol{F}) \cdot d\boldsymbol{S}.$$

(i) Calculation of the flux integral of a curl field

Example:

Let $\boldsymbol{F}(x, y, z) := (y, -x, e^{xz})$ for $(x, y, z) \in \mathbb{R}^3$, and let
$S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + (z - \sqrt{3})^2 = 4 \text{ and } z \geq 0\}$,
be oriented by the outward unit normal vectors.
In order to find $\iint_S (\operatorname{curl} \boldsymbol{F}) \cdot d\boldsymbol{S}$, we note that the induced orientation on $\eth S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1 \text{ and } z = 0\}$ is anticlockwise as seen from the point $(0, 0, 4)$.

By the Stokes theorem,

$$\begin{aligned}
\iint_S (\operatorname{curl} \boldsymbol{F}) \cdot d\boldsymbol{S} &= \int_{\eth S} \boldsymbol{F} \cdot d\boldsymbol{s} \\
&= \int_{-\pi}^{\pi} (\sin t, -\cos t, e^{\cos t \cdot 0}) \cdot (-\sin t, \cos t, 0) dt \\
&= \int_{-\pi}^{\pi} -(\sin^2 t + \cos^2 t) dt = -2\pi.
\end{aligned}$$

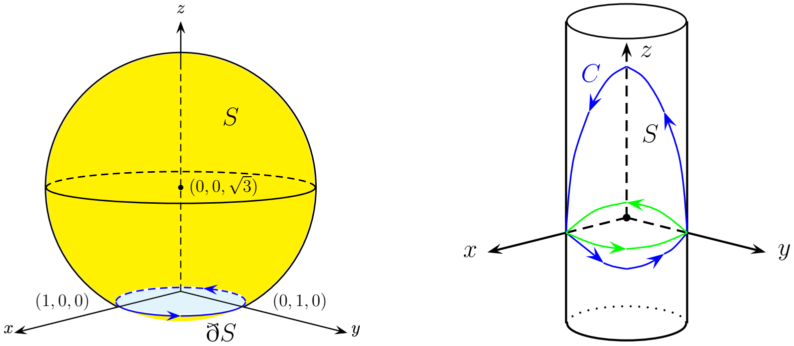


Figure: Examples (i) and (ii) of the use of the Stokes theorem

(ii) Calculation of the line integral along an oriented boundary

Example:
Let $\boldsymbol{F}(x, y, z) := (-y^3, x^3, -z^3)$ for $(x, y, z) \in \mathbb{R}^3$, and let C denote the intersection of the cylinder $x^2 + y^2 = 1$ and the plane $x+y+z=1$, oriented by the anticlockwise motion on the projection $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ of C on the xy-plane.
In order to find $\int_C \boldsymbol{F} \cdot d\boldsymbol{s}$, consider the surface

$$S := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq 1 \text{ and } x + y + z = 1\}.$$

Let us think of S as the graph of the function $f : D_0 \to \mathbb{R}$, where $D_0 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ and for $(x, y) \in D_0$, $f(x, y) := 1 - x - y$. Let S be oriented by the upward unit normal vectors. Then $\eth S = C$, and the induced orientation is the same as the given orientation of C. Hence

$$d\boldsymbol{S} = (-f_x, -f_y, 1)d(x, y) = (1, 1, 1)d(x, y).$$

Also,

$$\operatorname{curl}\boldsymbol{F} = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ -y^3 & x^3 & -z^3 \end{vmatrix} = (0, 0, 3x^2 + 3y^2),$$

and so $(\operatorname{curl}\boldsymbol{F})\cdot d\boldsymbol{S} = 3(x^2 + y^2)d(x, y)$.

By the Stokes theorem,

$$\begin{aligned} \int_C \boldsymbol{F}\cdot d\boldsymbol{s} &= \iint_S (\operatorname{curl}\boldsymbol{F})\cdot d\boldsymbol{S} = 3\iint_{D_0} (x^2 + y^2)d(x, y) \\ &= 3\int_0^1 \int_{-\pi}^{\pi} r^2 r\, dr\, d\theta = \frac{3\pi}{2}. \end{aligned}$$

Consequences of the Stokes Theorem

Proposition

Let $\boldsymbol{F}$ be a smooth vector field on an open subset D of $\mathbb{R}^3$ such that $\operatorname{curl} \boldsymbol{F} = \boldsymbol{0}$ on D.
(i) Suppose S is a bounded oriented piecewise C^2 surface in D, and let $\eth S$ denote its intrinsic boundary with the induced orientation, as in the Stokes theorem. Then $\int_{\eth S} \boldsymbol{F} \cdot d\boldsymbol{s} = 0$.

In particular, if $\eth S = C_1 \cup (-C_2)$, so that C_1 and $-C_2$ have the induced orientation, then $\int_{C_1} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{C_2} \boldsymbol{F} \cdot d\boldsymbol{s}$.

(ii) If D is simply connected, then $\boldsymbol{F}$ is a gradient field on D.

Proof:
(i) By the Stokes theorem,

$$\int_{\eth S} \boldsymbol{F} \cdot d\boldsymbol{s} = \iint_S (\operatorname{curl} \boldsymbol{F}) \cdot d\boldsymbol{S} = \iint_S \boldsymbol{0} \cdot d\boldsymbol{S} = 0.$$

Next, suppose $\eth S = C_1 \cup (-C_2)$, so that C_1 and $-C_2$ have the induced orientation. Then

$$\int_{C_1} \boldsymbol{F}\cdot d\boldsymbol{s} - \int_{C_2} \boldsymbol{F}\cdot d\boldsymbol{s} = \int_{C_1 - C_2} \boldsymbol{F}\cdot d\boldsymbol{s} = \int_{\eth S} \boldsymbol{F}\cdot d\boldsymbol{s} = 0.$$

(ii) Suppose D is simply connected. Let C denote a simple closed smooth curve in D. Then there is a smooth surface S in D such that $\eth S = C$, and so $\int_C \boldsymbol{F}\cdot d\boldsymbol{s} = \int_{\eth S} \boldsymbol{F}\cdot d\boldsymbol{s} = 0$.

This also holds if a smooth closed curve C in D intersects itself, since we can break up C into several simple closed smooth curves. As a result, line integrals of $\boldsymbol{F}$ are path independent in D, and so $\boldsymbol{F}$ is a gradient field on D. $\square$

Remark:

The last statement in part (i) above is known as the invariance of line integrals, or as the deformation principle for line integrals along paths in $\mathbb{R}^3$. (See the figure on the next slide.)

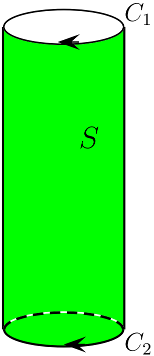


Figure: If curl $\boldsymbol{F} = \mathbf{0}$ on S, then $\displaystyle\int_{C_1} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{C_2} \boldsymbol{F} \cdot d\boldsymbol{s}$.

Examples: (i) Let $\boldsymbol{F} := (P, Q, R)$, where $P := -y/(x^2+y^2)$, $Q := x/(x^2+y^2)$ and $R := 1$ for $(x, y, z) \in \mathbb{R}^3$ with $(x, y) \neq 0$. We have seen in Lecture 33 that $Q_x = P_y$, and so $\operatorname{curl}(\boldsymbol{F}) = (0, 0, Q_x - P_y) = \boldsymbol{0}$. Consider the upper hemisphere $H := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1 \text{ and } z \geq 0\}$. Let C_1 denote the equator given by $x^2 + y^2 = 1$ and $z = 0$. Also, let a plane in $\mathbb{R}^3$ intersect H in a circle C_2 such that $C_1 \cap C_2 = \emptyset$ and $(0, 0, 1) \notin C_2$.

Let S denote the part of the hemisphere H bounded by C_1 and C_2, and let S be oriented by the outward unit normal vectors. Suppose C_1 and $-C_2$ have the induced orientation, so that both C_1 and C_2 are traced anticlockwise as seen from the point $(0, 0, 2)$. Then

$$\int_{C_2} -\frac{y}{x^2+y^2}dx + \frac{x}{x^2+y^2}\,dy + dz = \int_{C_1} -y\,dx + x\,dy = 2\pi,$$

as we had calculated earlier.

(ii) Let $D := \mathbb{R}^3$, and let $\boldsymbol{F}(x, y, z) := (y\, e^z,\ x\, e^z,\ x\, y\, e^z)$ for $(x, y, z) \in D$. Is $\boldsymbol{F}$ a gradient field? Now

$$(\operatorname{curl} \boldsymbol{F})(x, y, z) = (xe^z - xe^z, ye^z - ye^z, e^z - e^z) = (0, 0, 0)$$

for all $(x, y, z) \in D$. Also, D is simply connected. Hence $\boldsymbol{F}$ must be a gradient field. In fact, if we let $f(x, y, z) := x\, y\, e^z$ for $(x, y, z) \in D$, then $\nabla f = \boldsymbol{F}$ on D.

Recall the Green theorem for a multiply connected region D in $\mathbb{R}^2$: If ∂D denotes the positively oriented boundary of D, then

$$\int_{\partial D} P\,dx + Q\,dy = \iint_D (Q_x - P_y)d(x,y).$$

(i) (Circulation-Curl form) Let $\boldsymbol{F} := (P, Q, 0)$. Then

$$\int_{\partial D} \boldsymbol{F}\cdot d\boldsymbol{s} = \iint_D (\operatorname{curl} \boldsymbol{F})\cdot \boldsymbol{k}\, d(x,y).$$

(ii) (Flux-Divergence form) Let $\boldsymbol{F} := (Q, -P, 0)$. Then

$$\int_{\partial D} \boldsymbol{F}\cdot \hat{\boldsymbol{n}}\, ds = \iint_D (\operatorname{div} \boldsymbol{F})\, d(x,y).$$

We have seen that the Stokes theorem generalizes the form (i) above from a region D in $\mathbb{R}^2$ to a surface in $\mathbb{R}^3$. We shall now show that the Gauss divergence theorem generalizes the form (ii) above from a region D in $\mathbb{R}^2$ to a solid in $\mathbb{R}^3$.

Let D be a closed and bounded subset of $\mathbb{R}^3$. Suppose that the (usual) boundary ∂D of D consists of a finite number of orientable, piecewise smooth, nonintersecting closed geometric surfaces. (Thus all these surfaces are without any 'edges', that is, their intrinsic boundary is empty.) If each of these surfaces is so oriented that the unit normal vectors point out of the solid D, then we say that the boundary ∂D of D is **positively oriented**.

(Compare the definition of the positively oriented boundary of a multiply connected domain in $\mathbb{R}^2$.)

Example: Let $D := \{(x, y, z) \in \mathbb{R}^3 : 1 \leq x^2 + y^2 + z^2 \leq 4\}$. Then $\partial D = S_1 \cup S_2$ is positively oriented if $S_1 := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ is oriented by the inward unit normal vectors, and $S_2 := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 4\}$ is oriented by the outward unit normal vectors. Note that $\eth S_1 = \emptyset = \eth S_2$.

Theorem (Gauss Divergence Theorem)

Let D be a closed and bounded subset of $\mathbb{R}^3$ whose boundary ∂D consists of a finite number of orientable, piecewise smooth, nonintersecting closed geometric surfaces. Suppose ∂D is positively oriented. Let $\boldsymbol{F}$ be a smooth vector field on an open subset of $\mathbb{R}^3$ containing D. Then

$$\iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S} = \iiint_{D} (\operatorname{div} \boldsymbol{F}) d(x, y, z).$$

Proof: Let $\boldsymbol{F} := (P, Q, R)$. Then the surface integral $\iint_{\partial D} (P, Q, R) \cdot d\boldsymbol{S}$ on the left side is well-defined since P, Q, R are continuous on ∂D. The triple integral $\iiint_D (P_x + Q_y + R_z)\, d(x, y, z)$ on the right side is well-defined since P_x, Q_y and R_z are continuous on D and since ∂D is of three-dimensional content zero. (See Lecture 27.)

We show that

$$\iint_{\partial D} (0, 0, R) \cdot d\boldsymbol{S} = \iiint_D R_z \, d(x, y, z).$$

The other two equalities would follow similarly.

Let us suppose that D is an elementary region given as follows.

Let $D_0 \subset \mathbb{R}^2$, and let $\psi_1, \psi_2 : D_0 \to \mathbb{R}$ be continuous functions. Let

$$D := \{(x, y, z) \in \mathbb{R}^3 : \psi_1(x, y) \leq z \leq \psi_2(x, y)\},$$

and suppose ∂D consists of a single piecewise smooth, positively oriented, closed geometric surface S, where $S := S_1 \cup S_2$, with

$S_1 := \{(x, y, z) \in \mathbb{R}^3 : z = \psi_1(x, y)\}$ is its lower part, and

$S_2 := \{(x, y, z) \in \mathbb{R}^3 : z = \psi_2(x, y)\}$ is its upper part.

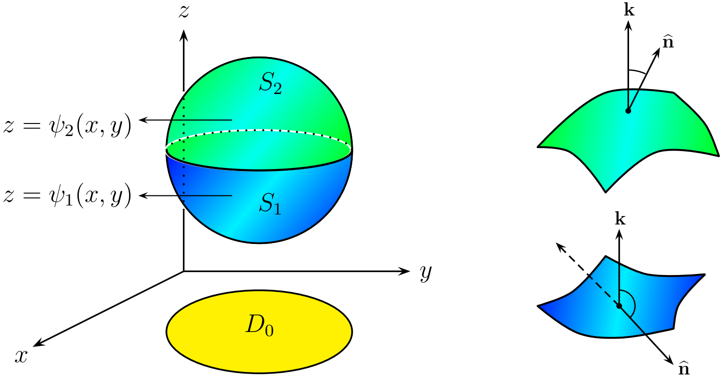


Figure: An outward normal $\hat{\boldsymbol{n}}$ to the upper part S_2 is upward, and an outward normal $\hat{\boldsymbol{n}}$ to the lower part S_1 is downward.

The upper part S_2 is given by $\Phi(x,y) := (x,y,\psi_2(x,y))$ for $(x,y) \in D_0$, and so the normal vector

$$\Phi_x \times \Phi_y = \big(1, 0, (\psi_2)_x\big) \times \big(0, 1, (\psi_2)y\big) = \big(-(\psi_2)_x, -(\psi_2)_y, 1\big),$$

points 'upward', that is, it goes out of the solid D. Hence Φ is an orientation-preserving parametrization of S_2, and so

$$\begin{aligned} \iint_{S_2} (0,0,R) \cdot d\mathbf{S} &= \iint_{D_0} \big(0, 0, R(\Phi(x,y))\big) \cdot (\Phi_x \times \Phi_y) d(x,y) \\ &= \iint_{D_0} R(x,y,\psi_2(x,y)) d(x,y). \end{aligned}$$

The lower part S_1 is given by $\Psi(x,y) := (x,y,\psi_1(x,y))$ for $(x,y) \in D_0$. Now the 'upward' normal vector $\Psi_x \times \Psi_y = \big(-(\psi_1)_x, -(\psi_1)_y, 1\big)$ points inward to the solid D.

Hence Ψ is an orientation-reversing parametrization of S_1, and so

$$\iint_{S_1} (0,0,R) \cdot d\boldsymbol{S} = -\iint_{D_0} R(x,y,\psi_1(x,y))\,d(x,y).$$

Combining the above two equalities, and using domain additivity (since $S_1 \cup S_2 = S = \partial D$), we obtain

$$\begin{aligned} \iint_{\partial D} (0,0,R) \cdot d\boldsymbol{S} &= \iint_{D_0} \left(\int_{\psi_1(x,y)}^{\psi_2(x,y)} R_z \, dz \right) d(x,y) \\ &= \iiint_D R_z \, d(x,y,z) \end{aligned}$$

by Part II of the FTC, and by the Fubini theorem for the triple integrals, which is also known as the Cavalieri Principle.

The above proof can be modified to treat the general case. $\square$

MA 105 : Calculus
Division 1, Lecture 27

Prof. Sudhir R. Ghorpade
IIT Bombay

Recap of the previous lecture

- Computational Uses of the Stokes Theorem
  - Calculation of flux integrals of continuous curl fields
  - Calculation of line integrals along an oriented boundary
- Consequences of the Stokes Theorem
- Invariance of line integrals or the deformation principle
- Positive orientation for the boundary of a solid in 3-space
- Gauss Divergence Theorem and its proof

Let us recall:

Theorem (Gauss Divergence Theorem)

Let D be a closed and bounded subset of $\mathbb{R}^3$ whose boundary ∂D consists of a finite number of orientable, piecewise smooth, nonintersecting closed geometric surfaces. Suppose ∂D is positively oriented. Let $\boldsymbol{F}$ be a smooth vector field on an open subset of $\mathbb{R}^3$ containing D. Then

$$\iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S} = \iiint_{D} (\operatorname{div} \boldsymbol{F})\, d(x, y, z).$$

The Gauss Divergence theorem can also be stated as follows:

$$\iint_{\partial D} P\, dy \wedge dz + Q\, dz \wedge dx + R\, dx \wedge dy$$
$$= \iiint_D (P_x + Q_y + R_z) d(x, y, z),$$

where P, Q, R are smooth real-valued functions defined on an open subset of $\mathbb{R}^3$ containing a closed bounded subset D of $\mathbb{R}^3$ whose boundary ∂D consists of a finite number of orientable, piecewise smooth, nonintersecting closed geometric surfaces, and is oriented by normal vectors pointing out of D.

Physical Interpretation of the Gauss Divergence Theorem:

Suppose a solid body D in $\mathbb{R}^3$ is enclosed by a closed geometric surface S, oriented in the direction of the outward normals. Let $\boldsymbol{F}$ be a vector field on D. The Gauss divergence theorem says that the flux of $\boldsymbol{F}$ across S is equal to the triple integral of the divergence of the vector field $\boldsymbol{F}$ over D.

Computational Uses of Gauss Divergence Theorem

(i) Calculation of a triple integral

Example: Finding the volume of a subset of $\mathbb{R}^3$

Let D be a closed and bounded subset of $\mathbb{R}^3$ whose positively oriented boundary ∂D consists of a finite number of orientable, piecewise smooth, closed geometric surfaces. Note that

$$\mathrm{Vol}\,(D) := \iiint_D 1_D \, d(x, y, z).$$

Let $\boldsymbol{F}$ be a smooth vector field such that $\operatorname{div} \boldsymbol{F} = 1$ on D. By the Gauss divergence theorem,

$$\mathrm{Vol}\,(D) = \iiint_D (\operatorname{div} \boldsymbol{F}) d(x, y, z) = \iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S}.$$

Thus if $\boldsymbol{F} := (P, Q, R)$, and $P_x + Q_y + R_z = 1$ on D, then

$$\mathrm{Vol}\,(D) = \iint_{\partial D} P\, dy \wedge dz + Q\, dz \wedge dx + R\, dx \wedge dy.$$

For instance, for $(x, y, z) \in D$, we may let

$$P(x, y, z) := \frac{x}{3}, \quad Q(x, y, z) := \frac{y}{3}, \quad R(x, y, z) := \frac{z}{3}.$$

Then clearly $P_x + Q_y + R_z = 1$ on D, and so

$$\mathrm{Vol}\,(D) = \frac{1}{3} \iint_{\partial D} x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy.$$

Suppose the positively oriented boundary ∂D of D is parametrized by $\Phi(u, v) := (x(u, v), y(u, v), z(u, v))$ for $(u, v) \in E$, where E is a subset of $\mathbb{R}^2$. Then

$$x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy = \det \begin{bmatrix} x & y & z \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{bmatrix} d(u, v).$$

The determinant of the matrix on the right is called the **Wronskian** of x, y, z, and it is denoted by $W(x, y, z)$. Thus

$$\text{Vol}\,(D) = \frac{1}{3}\iint_E W(x, y, z)(u, v)d(u, v).$$

We note that $\text{Vol}\,(D)$ is also equal to

$$\iint_{\partial D} x\,dy\wedge dz \quad \text{or} \quad \iint_{\partial D} y\,dz\wedge dx \quad \text{or} \quad \iint_{\partial D} z\,dx\wedge dy$$

if we let $P := x$, $Q := 0 =: R$, or $Q := y$, $P := 0 =: R$, or $R := z$, $P := 0 =: Q$, respectively.

Example Let $a, b, c, d > 0$ be with $d < \min\{a, b, c\}$, and let

$$D := \left\{ (x, y, z) \in \mathbb{R}^3 : \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1 \text{ and } x^2 + y^2 + z^2 \geq d^2 \right\}.$$

Now $S_1 := \{(x, y, z) \in \mathbb{R}^3 : x^2/a^2 + y^2/b^2 + z^2/c^2 = 1\}$, oriented by the outward normal vectors and $S_2 := \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = d^2\}$, oriented by the inward normal vectors, constitute the positively oriented boundary ∂D of D. An orientation-preserving parametrization of S_1 is given by $\Phi(u, v) := (x(u, v), y(u, v), z(u, v))$, where $x(u, v) := a \sin u \cos v$, $y(u, v) := b \sin u \sin v$, $z(u, v) := c \cos u$ for $(u, v) \in [0, \pi] \times [-\pi, \pi]$. We can check that the Wronskian $W(x, y, z)$ at $(u, v) \in [0, \pi] \times [-\pi, \pi]$ is equal to $abc \sin u$.

Also, let $a := d, b := d, c := d$ to obtain an orientation-reversing parametrization of S_2. Hence

$$\text{Vol}\,(D) = \frac{1}{3}(abc - d^3) \int_{-\pi}^{\pi} \left(\int_0^{\pi} \sin u \, du \right) dv = \frac{4\pi}{3}(abc - d^3).$$

(ii) Calculation of surface integral across an oriented boundary

Example:

Let $D := \{(x, y, z) \in \mathbb{R}^3 : 0 \le x \le 1,\ 0 \le y \le 2,\ 0 \le z \le 3\}$, and let $\boldsymbol{F}(x, y, z) := (x^2z^3,\ 2x\,y\,z^3,\ x\,z^4)$ for $(x, y, z) \in \mathbb{R}^3$.

Direct calculation of $\iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S}$ is long since the positively oriented boundary ∂D of D has 6 faces.

On the other hand, we note that $\operatorname{div} \boldsymbol{F}(x, y, z) = 2x\,z^3 + 2x\,z^3 + 4x\,z^3 = 8x\,z^3$, and by the Gauss divergence theorem, we easily obtain

$$\iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S} = \iiint_D (\operatorname{div} \boldsymbol{F})\, d(x, y, z) = \iiint_D 8x\,z^3 d(x, y, z) = 162$$

using the Fubini theorem for triple integrals.

Consequences of the Gauss Divergence Theorem

Proposition

Let $\boldsymbol{F}$ be a smooth vector field on an open subset containing a closed and bounded subset D of $\mathbb{R}^3$ such that $\operatorname{div} \boldsymbol{F} = 0$ on D. If ∂D consists of a finite number of orientable, piecewise smooth, nonintersecting closed geometric surfaces, and is oriented by normal vectors pointing out of the solid D, then

$$\iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S} = 0.$$

Proof:

By the Gauss divergence theorem,

$$\iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S} = \iiint_D (\operatorname{div} \boldsymbol{F}) d(x, y, z) = \iiint_D 0 \, d(x, y, z) = 0. \quad \square$$

Invariance of Some Surface Integrals

Suppose S_1 and S_2 are piecewise smooth closed geometric surfaces such that S_2 lies in the interior of S_1, that is, S_1 encloses S_2. Let D denote the subset of $\mathbb{R}^3$ consisting of S_1, S_2 and the region between them. Let P, Q, R be smooth scalar fields satisfying $P_x + Q_y + R_z = 0$ on an open set containing D. If $\boldsymbol{F} := (P, Q, R)$, then $\operatorname{div} \boldsymbol{F} = 0$ on D, and so

$$\iint_{\partial D} P\, dy \wedge dz + Q\, dz \wedge dx + R\, dx \wedge dy = \iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S} = 0,$$

provided $\partial D = S_1 \cup S_2$ is oriented by unit normal vectors pointing out of D. Then S_1 and S_2 are oriented by unit normal vectors going in opposite directions. Hence if S_1 and S_2 are oriented by unit normal vectors going in the same direction, then the surface integrals of $P\, dy \wedge dz + Q\, dz \wedge dx + R\, dx \wedge dy$ across S_1 and S_2 are equal.

This is called the deformation principle for surface integrals.

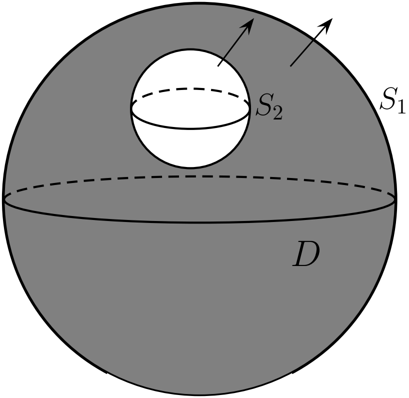


Figure: If div $\boldsymbol{F} = 0$ on D, then $\displaystyle\iint_{S_1} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{S_2} \boldsymbol{F} \cdot d\boldsymbol{S}$.

Example: Gauss Law in $\mathbb{R}^3$: Let S be a piecewise smooth closed geometric surface in $\mathbb{R}^3$, and let D denote the union of S and the region enclosed by S, so that S is the boundary ∂D of D. Suppose $\mathbf{0} \notin S$, and S is oriented by unit normal vectors that point outside D. Then

$$\iint_S \frac{\boldsymbol{r}}{r^3} \cdot d\boldsymbol{S} = \begin{cases} 0 & \text{if } \mathbf{0} \notin D, \\ 4\pi & \text{if } \mathbf{0} \in D, \end{cases}$$

where $\boldsymbol{r}(x, y, z) := (x, y, z)$ and $r(x, y, z) := \|\boldsymbol{r}(x, y, z)\|$ for $(x, y, z) \in \mathbb{R}^3$.

Proof: Let $\boldsymbol{F} := \boldsymbol{r}/r^3$, that is, $\boldsymbol{F} := (P, Q, R)$, where $P := x/r^3$, $Q := y/r^3$ and $R := z/r^3$ for $\boldsymbol{r} \neq \mathbf{0}$. As we saw earlier, $P_x + Q_y + R_z = \operatorname{div} \boldsymbol{F} = 0$ on $\mathbb{R}^3 \setminus \{\mathbf{0}\}$.

First, let $\mathbf{0} \notin D$. Then the vector field $\boldsymbol{F}$ is smooth on D. Hence $\iint_S \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{\partial D} \boldsymbol{F} \cdot d\boldsymbol{S} = \iiint_D \operatorname{div}(\boldsymbol{F}) d(x, y, z) = 0$ by the Gauss divergence theorem.

Next, let $\mathbf{0} \in D$. We have seen earlier that $\iint_{S_1} \boldsymbol{F} \cdot d\boldsymbol{S} = 4\pi$, where S_1 denotes the standard unit sphere in $\mathbb{R}^3$ (which is oriented by the outward unit normal vectors). Since $\mathbf{0}$ is an interior point of D, there is $\epsilon > 0$ such that the closed ball of radius ϵ and center $\mathbf{0}$ lies inside D. Let S_ϵ denote the standard sphere of radius ϵ (which is oriented by the outward unit normal vectors). Then

$$\iint_S \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{S_\epsilon} \boldsymbol{F} \cdot d\boldsymbol{S} = \iint_{S_1} \boldsymbol{F} \cdot d\boldsymbol{S} = 4\pi$$

by the deformation principle for surface integrals, applied first to the surfaces S and S_ϵ, and then to the surfaces S_ϵ and S_1. $\square$

We have now reached the end of this course. It is therefore a good time to thank you all for your cooperation during the semester and wish you all the best for the end-sem and your future endeavours!

P.S. We will now discuss Tutorial 14. There will also be a Help Session for D3 on Saturday, 9th November at 8.30 PM in LA 001. Interested students of D1 are welcome!