

Math 241
Calculus III

Thomas
Honold

Administrative
Things

Introductory
Remarks

Real Vectors and
Vector Spaces

The Field of Real
Numbers

Analytic
Geometry in
 \mathbb{R}^2 and \mathbb{R}^3

Linear and
Affine
Subspaces

Representations
of Lines and
Planes in \mathbb{R}^3

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Administrative Things

2 Introductory Remarks

Real Vectors and Vector Spaces
The Field of Real Numbers

3 Analytic Geometry in \mathbb{R}^2 and \mathbb{R}^3

4 Linear and Affine Subspaces

5 Representations of Lines and Planes in \mathbb{R}^3

Math 241
Calculus III

Thomas
Honold

Administrative
Things

Introductory
Remarks

Real Vectors and
Vector Spaces

The Field of Real
Numbers

Analytic
Geometry in
 \mathbb{R}^2 and \mathbb{R}^3

Linear and
Affine
Subspaces

Representations
of Lines and
Planes in \mathbb{R}^3

Today's Lecture: Introduction

Lecturer

Prof. Dr. Thomas Honold

ZJU-UIUC Institute

International Campus, Haining

Office: Room 415, ZJUI Building (1C)

Office hours: Mon, 17:00 – 18:30

Email: honold@zju.edu.cn

Teaching Staff

Teaching Assistants

Li Pengyu pengyu.20@intl.zju.edu.cn

Hu Kejia kejia.20@intl.zju.edu.cn

Chen Zhenbo zhenbo.22@intl.zju.edu.cn

Yu Jiarui jiarui.20@intl.zju.edu.cn

Ren Hao haor.20@intl.zju.edu.cn

Dong Wei weid.21@intl.zju.edu.cn

Wei Peiran peiran.22@intl.zju.edu.cn

Qiu Xubin xubin.20@intl.zju.edu.cn

Guo Wangyihan wangyihan.22@intl.zju.edu.c

Weekly Timetable

Lecture A

Mon/Thu 10 – 11, LTW 102
Fri 9:30 – 10:30, LTW 102

Lecture B

Mon/Thu 11 – 12, LTW 102
Fri 10:30 – 11:30, LTW 102

Discussion Session

Wed, 18–20 (8 groups)

Homework

Homework is assigned on Wednesdays (sometimes Thursdays) and must be handed in on the next Wednesday before the discussion session. Late homework will not be accepted.

Textbook

[Ste21] J. Stewart, D. Clegg, S. Watson, *Calculus: Early Transcendentals*, 9th edition, **metric version**, Cengage Learning, 2021

Other editions of Stewart's book will not be supported.

Course Contents (tentative)

Administrative
Things

Introductory
Remarks

Real Vectors and
Vector Spaces

The Field of Real
Numbers

Analytic
Geometry in
 \mathbb{R}^2 and \mathbb{R}^3

Linear and
Affine
Subspaces

Representations
of Lines and
Planes in \mathbb{R}^3

Week	Topics	[Ste21] Sections
1,2	Analytic Geometry	12.1–12.6
3,4	Vector Functions	13.1–13.4
5,6,7,8	Partial Derivatives	14.1–14.8
9,10,11	Multiple Integrals	15.1–15.10
12,13,14	Vector Calculus	16.1–16.9

The lecture won't follow the textbook strictly (regarding notation, mathematical depth, and invoking tools from Linear Algebra).

Examination Regulations

Calculation of the final score

40 % final exam (closed book)

30 % 3 midterm exams (closed book),
score weights 10% + 20% + 10%

20 % homework

10 % discussion session work

The total midterm score (30) will be computed as

$$\max \{b + c, a + b/2 + c, a + b\},$$

where a (≤ 10), b (≤ 20), c (≤ 10) denote the individual midterm scores. (In other words, the bottom 25 % of the total midterm score is discarded.)

The Calculus III midterms are usually held in course weeks 5, 9, 13 (tentative). Exact dates will be fixed in due course.

Math 241
Calculus III

Thomas
Honold

Administrative
Things

Introductory
Remarks

Real Vectors and
Vector Spaces

The Field of Real
Numbers

Analytic
Geometry in
 \mathbb{R}^2 and \mathbb{R}^3

Linear and
Affine
Subspaces

Representations
of Lines and
Planes in \mathbb{R}^3

Course Website

Blackboard

Computations

For numerical and symbolic computations, graphing functions, etc., I will use the computer algebra system **SageMath** www.sagemath.org. SageMath forms an essential component of the Linear Algebra course Math257, and you are advised to install it on your laptop and experiment with it.

Some Advice Before We Start

- Attend each class! (cf. next slide)
- Solve (well, at least try hard to solve) each exercise!
- Don't hesitate to ask (stupid) questions!
- Don't take everything in the textbook too literally!

Attendance Control

Starting with this Fall semester, lecture attendance will be checked (electronically), and multiple unauthorized absence penalized through score deduction. Similarly for discussion sessions.

Students can be exempted from class attendance, but only for very important reasons and with prior authorization.

Vectors

The subject of Linear Algebra

First Definition

A *vector* (with entries in the field \mathbb{R} of real numbers) is an *n-tuple* (or sequence of length *n*)

$$\mathbf{v} = (v_1, v_2, \dots, v_n) \quad \text{with } v_i \in \mathbb{R}$$

for some integer $n \geq 1$; n is called the *dimension* of \mathbf{v} . The set of all *n*-dimensional vectors is denoted by \mathbb{R}^n (and called “real *n*-space” or *n-dimensional Euclidean space*).

Equality of vectors

Two vectors $\mathbf{u} = (u_1, u_2, \dots, u_n)$, $\mathbf{v} = (v_1, v_2, \dots, v_n)$ are equal if $u_1 = v_1 \wedge u_2 = v_2 \wedge \dots \wedge u_n = v_n$ (equality of tuples/sequences).

Row versus Column Vectors

A vector \mathbf{v} can either be represented as a row or a column, e.g.,

$$\mathbf{v} = \begin{pmatrix} 2 & -1 & 5 \end{pmatrix} \quad \text{or} \quad \mathbf{v} = \begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix}.$$

Such vectors are referred to as *row vectors*, respectively, *column vectors*. Switching between row and column vectors is named *transposition* (e.g., $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is the transpose of $(1, -1)$ and vice versa).

Question

Is a vector \mathbf{v} equal to its transpose?

Answer: Yes and No.

Certainly a 1-dimensional vector is equal to its transpose.

Transposition identifies a vector and its transpose vector, so they cannot be distinguished.

Later we will give a more precise definition of row/column vectors and transposition, as a result of which n -dimensional row/column vectors with $n > 1$ are not equal to their transposes.

Operations on Vectors

Two vectors of the same dimension (length) n can be added, and a vector of (any) dimension n can be multiplied by a real number (“scalar”). The result is again a vector of dimension n .

Vector addition

For $\mathbf{u} = (u_1, u_2, \dots, u_n)$, $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ we define $\mathbf{u} + \mathbf{v} \in \mathbb{R}^n$ by

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n).$$

Scalar multiplication

For $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ and $c \in \mathbb{R}$ we define $c\mathbf{v} \in \mathbb{R}^n$ by

$$c\mathbf{v} = (cv_1, cv_2, \dots, cv_n).$$

Example (using 2-dimensional column vectors)

Suppose $\mathbf{v} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$, $\mathbf{w} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$.

$$\mathbf{v} + \mathbf{w} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix} = \begin{pmatrix} 4 - 1 \\ 2 + 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \mathbf{w} + \mathbf{v},$$

$$2\mathbf{v} = \begin{pmatrix} 2v_1 \\ 2v_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix},$$

$$(-1)\mathbf{v} = \begin{pmatrix} -v_1 \\ -v_2 \end{pmatrix} = \begin{pmatrix} -4 \\ -2 \end{pmatrix} = -\mathbf{v},$$

$$\mathbf{v} - \mathbf{w} = \mathbf{v} + (-1)\mathbf{w} = \begin{pmatrix} 4 \\ 2 \end{pmatrix} + \begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 4 - (-1) \\ 2 - 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \end{pmatrix},$$

$$1\mathbf{v} = \begin{pmatrix} 1 \cdot v_1 \\ 1 \cdot v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \mathbf{v},$$

$$0\mathbf{v} = \begin{pmatrix} 0 \cdot v_1 \\ 0 \cdot v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \mathbf{0} \quad (\text{the all-zero vector})$$

Some of the computations generalize to arbitrary vectors; cf. next slide.

Vector Space Axioms/Laws

Let n be a fixed positive integer.

For all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and all real numbers $c, d \in \mathbb{R}$ the following are true:

- V1 $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (*commutative law*)
- V2 $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (*associative law*)
- V3 $\mathbf{u} + \mathbf{0} = \mathbf{u}$ (*additive identity*)
- V4 $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ (*additive inverse*)
- V5 $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ (*1st distributive law*)
- V6 $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$ (*2nd distributive law*)
- V7 $(cd)\mathbf{u} = c(d\mathbf{u})$ (*mixed associative law*)
- V8 $1\mathbf{u} = \mathbf{u}$ (*identity law*)

Here $\mathbf{0} = (0, 0, \dots, 0)$ denotes the all-zero vector of \mathbb{R}^n .

This says that \mathbb{R}^n , together with the operations of vector addition and scalar multiplication, forms a vector space over \mathbb{R} according to the following definition.

Definition (Abstract vector space)

Let V be a non-empty set (of “vectors”), $0 = 0_V \in V$ a distinguished element (“zero vector”), K a field (such as \mathbb{Q} , \mathbb{R} , \mathbb{C}) with multiplicative identity 1, and $V \times V \rightarrow V$, $(u, v) \mapsto u + v$, (“vector addition”), $K \times V \rightarrow V$, $(c, u) \mapsto cu$ (“scalar multiplication”) two binary operations. We say that V forms a *vector space over K* if for all $u, v, w \in V$ and $c, d \in K$ the axioms (V1)–(V8) of the previous slide (mutatis mutandis) hold, except that (V4) is replaced by “for all $u \in V$ there exists an element $u' \in V$ such that $u + u' = 0_V$.

Notes

- This provides the second definition of the term “vector”: A *vector* is just an element of an abstract vector space V as defined above.
- The element u' in Axiom (V4) is uniquely determined and denoted by $-u$ (as in the case of \mathbb{R}^n).
- $V \times V = \{(u, v); u, v \in V\}$, $K \times V = \{(c, v); c \in K, v \in V\}$ are so-called *cartesian products* (sets of all ordered pairs formed from two given sets); compare with $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$.

Notes cont'd

- There are many vector spaces over \mathbb{R} other than the *standard* spaces \mathbb{R}^n . For example, the set of all functions $f: [0, 1] \rightarrow \mathbb{R}$ forms a vector space over \mathbb{R} relative to the operations $(f + g)(x) = f(x) + g(x)$ and $(cf)(x) = cf(x)$ for $x \in [0, 1]$ (“point-wise addition and scalar multiplication”).
- In Digital Communication vector spaces over the binary field $\mathbb{F}_2 = \{0, 1\}$ (with modulo 2 addition $1 + 1 = 0$) are particularly important. For example, consider all bit strings $x_1 x_2 \dots x_8$, $x_i \in \mathbb{F}_2$, of length 8 (bytes) with an additional parity bit $x_9 = x_1 + \dots + x_8 \pmod{2}$ added, such as $10110000|1$ or $10111000|0$. The $2^8 = 256$ binary words of length 9 obtained in this way form a vector space over \mathbb{F}_2 , a so-called *binary linear code* (check the details, if you like).
- A given set V may form a vector space in different ways (i.e., with respect to different operations, over different fields, etc.).
- (Abstract) Vector spaces and their properties are the subject of *Linear Algebra* and discussed in Math257.

A Note about Fields

In Pure Mathematics (as opposed to Physics) a *field* is an algebraic structure consisting of a set F and two operations $(a, b) \mapsto a + b$ (addition) and $(a, b) \mapsto a \cdot b = ab$ (multiplication), which satisfy the usual laws of algebra known from \mathbb{R} . In particular there must exist two distinguished elements $0, 1 \in F$ satisfying $a + 0 = a$ and $a \cdot 1 = a$ for all $a \in F$, and all equations $a + x = b$ ($a, b \in F$) and $ax = b$ ($a, b \in F$, $a \neq 0$) must be uniquely solvable (thus providing definitions of differences $b - a$ and quotients b/a).

A Hierarchy of Number Systems

$$\mathbb{N} = \{1, 2, 3, \dots\} \quad (\text{natural numbers})$$

$$\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, 3, \dots\} \quad (\text{integers})$$

$$\mathbb{Q} = \left\{ c \in \mathbb{R}; c = \frac{p}{q} \text{ for some } p \in \mathbb{Z}, q \in \mathbb{N} \right\} \quad (\text{rational numbers})$$

$$\begin{aligned} \mathbb{A} = \left\{ c \in \mathbb{R}; a_0 + a_1 c + \dots + a_r c^r = 0 \right. \\ \left. \text{for some } a_i \in \mathbb{Z}, \text{ not all zero} \right\} \quad (\text{real algebraic numbers}) \end{aligned}$$

$$\mathbb{R} \quad (\text{real numbers})$$

$$\mathbb{C} = \{a + bi; a, b \in \mathbb{R}\} \quad (\text{complex numbers})$$

Out of these, \mathbb{Q} , \mathbb{A} , \mathbb{R} and \mathbb{C} form fields (with respect to the usual addition and multiplication), but \mathbb{N} and \mathbb{Z} do not.

In \mathbb{N} and \mathbb{Z} the equation $2x = 1$ has no solution. As disproof for \mathbb{N} (but not for \mathbb{Z}) we could also use an additive equation like $x + 1 = 0$. Thus in a sense \mathbb{Z} is closer to a field than \mathbb{N} . In fact the only thing which prevents \mathbb{Z} to be a field is the non-solvability of $ax = 1$ for some nonzero $a \in \mathbb{Z}$. In Abstract Algebra a structure like \mathbb{Z} is referred to as a *commutative ring*.

When proving that \mathbb{Q} , \mathbb{A} , \mathbb{R} and \mathbb{C} are fields, the tedious part is the construction of \mathbb{R} together with its addition and multiplication, and the verification of the field laws. For \mathbb{C} one can verify the laws directly from the definitions $(a + bi) + (c + di) = a + c + (b + d)i$, $(a + bi)(c + di) = (ac - bd) + (ad + bc)i$. For \mathbb{Q} and \mathbb{A} it suffices to check that both contain 0 and 1 and with any numbers a, b also $a \pm b$, ab , and (provided that $b \neq 0$) a/b .

Examples of algebraic numbers are $\sqrt{2}$ (solving $x^2 - 2 = 0$), $\sqrt[3]{5}$ (solving $x^3 - 5 = 0$), but also more exotic numbers such as the unique real solution of $x^5 + x + 1 = 0$. The numbers $e = 2.71828\dots$ and $\pi = 3.14159\dots$ are transcendental (i.e., not algebraic), but the proofs of these facts are difficult. Further examples of transcendental numbers are provided by limits $\sum_{n=0}^{\infty} a_n$ of rapidly converging series with coefficients $a_n \in \mathbb{Q}$, for example $\sum_{n=0}^{\infty} \frac{1}{2^{n!}}$.

Exercise

Find a nontrivial polynomial equation for each of $\sqrt{2} + \sqrt[3]{5}$, $\sqrt{2} - \sqrt[3]{5}$, $\sqrt{2} \cdot \sqrt[3]{5}$, and $\sqrt{2}/\sqrt[3]{5}$.

Almost All Real Numbers are Transcendental

In the chain of inclusions

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{A} \subset \mathbb{R} \subset \mathbb{C}$$

the first three steps and the last step are small, but the step $\mathbb{A} \subset \mathbb{R}$ is big!

Definition

An infinite set S is said to be *countable* (or *denumerable*) if it can be put in 1-1 correspondence with the natural numbers, i.e.,

$S = \{s_1, s_2, s_3, \dots\}$ with $s_i \neq s_j$ for $i \neq j$.

\mathbb{N} is trivially countable, as is $\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \dots\}$.

\mathbb{Q} is countable as well, as the following enumeration of all positive rational numbers shows:

$$\frac{0}{0}, \quad \frac{1}{0}, \quad \frac{0}{1}, \quad \frac{2}{0}, \quad \frac{1}{1}, \quad \frac{0}{2}, \quad \frac{3}{0}, \quad \frac{2}{1}, \quad \frac{1}{2}, \quad \frac{0}{3}, \quad \frac{4}{0}, \quad \frac{3}{1}, \quad \frac{2}{2}, \quad \frac{1}{3}, \quad \frac{0}{4}, \quad \dots$$

Undefined numbers and numbers that appear a second time, i.e., $\frac{a}{b}$ with $\gcd(a, b) > 1$, are discarded. The resulting sequence contains every positive rational number exactly once.

Likewise, \mathbb{A} is countable. Since nonzero polynomials have only finitely many zeros, it suffices to enumerate polynomials with integer coefficients. This can be done in a similar way as for \mathbb{Q} .

But the interval $[0, 1]$, and hence \mathbb{R} , is uncountable! One possible proof, called CANTOR's *diagonal argument*, uses the fact that every real number a satisfying $0 < a \leq 1$ has a unique “non-terminating” decimal expansion $a = 0.a_1a_2a_3\dots$ with $a_k \neq 0$ for infinitely many k ; cf. subsequent lemma. Now suppose, by contradiction, that $(0, 1] = \{a, b, c, \dots, z, \dots\}$ and consider the doubly-infinite array

$$a = 0.a_1a_2a_3\dots$$

$$b = 0.b_1b_2b_3\dots$$

$$c = 0.c_1c_2c_3\dots$$

$$\vdots$$

The trick is to consider the “diagonal” number $\delta = 0.a_1b_2c_3\dots$. Since we have $10 > 2$ digits available, we can change each digit of δ into a different and nonzero digit a'_1, b'_2, c'_3, \dots . The resulting number $\delta' = 0.a'_1b'_2c'_3\dots$ satisfies $\delta' \in (0, 1]$ and (using uniqueness of the decimal expansion) is different from all numbers a, b, c, \dots, z, \dots . Contradiction!

Decimal Expansion

Lemma

Every real number $a \in (0, 1]$ has a unique non-terminating decimal expansion $a = 0.a_1 a_2 a_3 \dots$, i.e., $a = \sum_{k=1}^{\infty} a_k 10^{-k}$ with $a_k \in \{0, 1, \dots, 9\}$ for all k and $a_k \neq 0$ for infinitely many k .

For numbers a of the form $0.a_1 a_2 \dots a_k 000 \dots$ with $a_k \neq 0$ (i.e. $10^k a$ is an integer, but $10^{k-1} a$ is not), the non-terminating expansion is $0.a_1 a_2 \dots a_{k-1} (a_k - 1) 999 \dots$

Proof.

Assume w.l.o.g. $a \neq 1$ and define a sequence of digits $a_0, a_1, a_2, \dots \in \{0, 1, \dots, 9\}$ and a sequence of “remainders” $r_0, r_1, r_2, \dots \in [0, 1)$ recursively by $r_0 = a$ and $10r_{k-1} = a_k + r_k$ for $k \geq 1$. By induction we have $10r_{k-1} \in [0, 10)$, so that a_k and r_k are well-defined.

$$r_1 = 10a - a_1,$$

$$r_2 = 10(10a - a_1) - a_2 = 10^2 a - 10a_1 - a_2,$$

$$r_3 = 10(10^2 a - 10a_1 - a_2) - a_3 = 10^3 a - 10^2 a_1 - 10a_2 - a_3, \quad \text{and}$$

$$r_k = 10^k a - 10^{k-1} a_1 - 10^{k-2} a_2 - \dots - 10a_{k-1} - a_k \quad \text{in general.}$$

Proof cont'd.

Hence we have

$$\frac{r_k}{10^k} = a - \frac{a_1}{10} - \frac{a_2}{10^2} - \cdots - \frac{a_k}{10^k}.$$

Since $0 \leq r_k < 1$, this shows $a = \sum_{k=1}^{\infty} a_k 10^{-k}$.

The uniqueness part is left as an exercise. □

Notes

- For any integer $b \geq 2$ there is a similar “base- b expansion” $a = \sum_{k=1}^{\infty} a_k b^{-k}$, $a_k \in \{0, 1, \dots, b-1\}$, of $a \in (0, 1]$.
- The decimal expansion (and similarly any base- b expansion) can be extended to real numbers $a > 0$ by allowing an arbitrary integer as starting index, i.e., $a = \sum_{k=k_0}^{\infty} a_k 10^{-k}$ with $k_0 \in \mathbb{Z}$. If $k_0 \leq 0$, this is written as $a = a_{k_0} \dots a_{-1} a_0 . a_1 a_2 a_3 \dots$. Here $a_{k_0} \dots a_{-1} a_0$ is the decimal representation of the integer part $\lfloor a \rfloor$ of a (provided that a is not an integer).

\mathbb{R} can be put into 1-1 correspondence with any subinterval of positive length; cf. the accompanying exercises. The common cardinality of all intervals (and \mathbb{R}) is referred to as *cardinality of the continuum*.

Remark

It is possible to assign to very general subsets $S \subseteq [0, 1]$ a number $v(S) \in [0, 1]$ that can be interpreted as total length of S , or as the probability that a randomly chosen number $x \in [0, 1]$ is contained in S ; cf. the chapter on multidimensional integration. It turns out that all countable subsets S have $v(S) = 0$ and $v([0, 1] \setminus S) = 1$. Choosing $S = \mathbb{A}$ gives the announced property “allmost all real numbers in $[0, 1]$ are transcendental”.

Exercise

- ① Explain how to obtain an enumeration of \mathbb{Q} from that of the positive rational numbers.
- ② Find an enumeration of the set of polynomials with integer coefficients.
- ③ Find a 1-1 correspondence (bijection) between \mathbb{R} and \mathbb{C} .
- ④ Prove the uniqueness part of the lemma about decimal expansions.

But what is \mathbb{R} ?

The answer is provided by the following theorem, whose proof is beyond the scope of this course.

Theorem

- ① *There exists a complete ordered field.*
- ② *If F_1 and F_2 are complete ordered fields then there exists a bijection $f: F_1 \rightarrow F_2$ satisfying $f(a + b) = f(a) + f(b)$, $f(ab) = f(a)f(b)$ and $a \leq b \Leftrightarrow f(a) \leq f(b)$ for all $a, b \in F_1$ (a so-called order isomorphism).*

The unique (up to order isomorphism) complete ordered field is called *field of real numbers* and denoted by \mathbb{R} .

Explanations

- *Ordered field* refers to a field F together with a total order relation \leq on F which is compatible with the field operations in the sense that $a, b > 0$ imply $a + b > 0$ and $ab > 0$. (One can also specify the order relation in terms of its *domain* $P = \{x \in F; x > 0\}$ *of positivity*, which should satisfy $F = P \uplus \{0\} \uplus (-P)$ (disjoint decomposition), $P + P \subseteq P$, and $P \cdot P \subseteq P$, by defining $x > y \leftrightarrow x - y \in P$.)
- An ordered field F is said to be *complete* if every non-empty subset $S \subseteq F$ that is bounded from above (i.e., there exists $c \in F$ such that $x \leq c$ for all $x \in S$) has a least upper bound c^* (i.e., $c^* \leq c$ for all upper bounds c of S).

The least upper bound of S , if applicable, is uniquely determined and called *supremum* of S (notation $c^* = \sup S$).

One can derive all important properties of \mathbb{R} from the field axioms, the order axioms listed above, and the completeness property. (One such property is that the order relation on \mathbb{R} is *Archimedean*, i.e., for every $x \in \mathbb{R}$ there exists an $n \in \mathbb{N}$ such that $\underbrace{1 + 1 + \cdots + 1}_{n \text{ summands}} > x$.)

Assignments for Week 1

Reading Assignment

Read Sections 12.1–12.3 and 12.5 in [Ste21].

Tutorial/Homework Assignment

Worksheet 1 will be provided on Wed Sep 20 and is due on the same day. Homework 1 will be provided until Thu Sep 21 and is due on Wed Sep 27.

Coordinate systems

Basic observation

We can represent points of any plane in the real world (such as the plane determined by the whiteboard) by ordered pairs $(x_1, x_2) \in \mathbb{R}^2$ (i.e., two-dimensional real vectors), provided we have chosen

- ① a fixed reference point O (called *origin*);
- ② two orthogonal directions through O , called *coordinate axes*.
- ③ A unit of measurement, such as 1 m.

A point P is represented by (x_1, x_2) , where $|x_1|, |x_2|$ denote the distances from P to the respective coordinate axis and the signs are determined by the corresponding projection points P_1, P_2 .

The correspondence $P \triangleq (x_1, x_2)$ appears to yield a “bijection” from the physical object “plane determined by the whiteboard” to the mathematical object \mathbb{R}^2 .

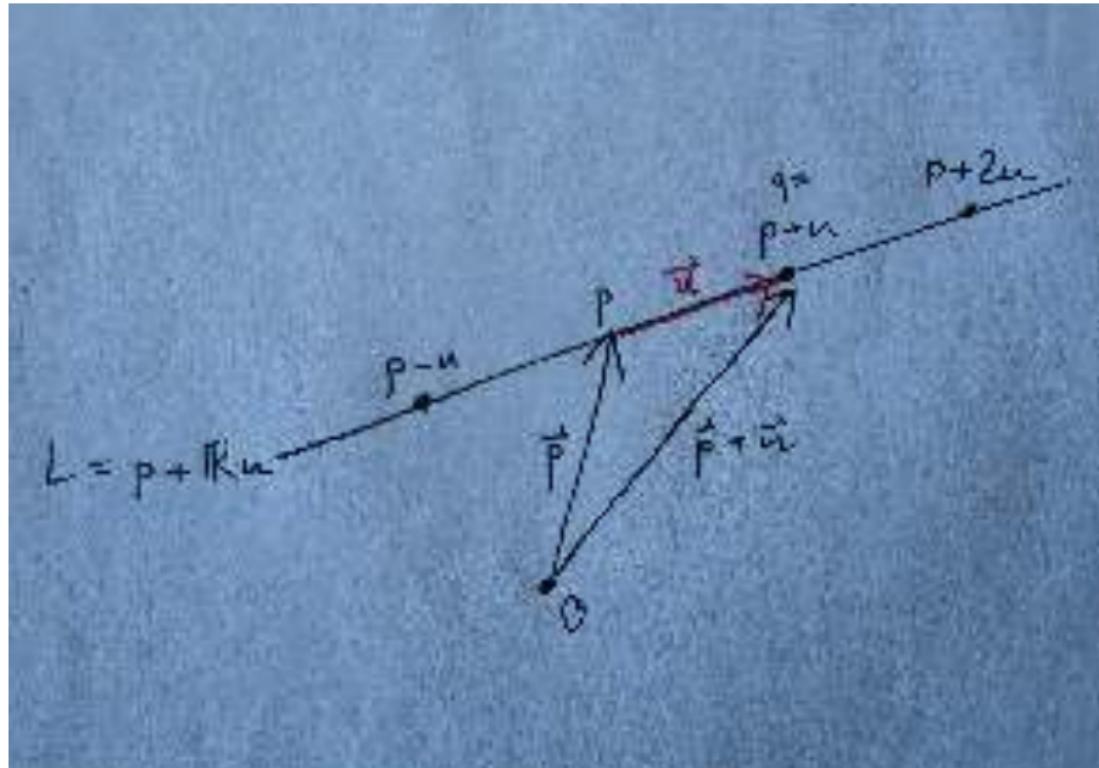


Figure: Several points on the line $L = \mathbf{p} + \mathbb{R}\mathbf{u}$ through \mathbf{p} and $\mathbf{q} = \mathbf{p} + \mathbf{u}$; points are identified with vectors; arrows are shown only for the purpose of illustrating the geometric meaning of vector addition

Moreover,

- The physical *line* through any two distinct points $P \triangleq (p_1, p_2)$ and $Q \triangleq (q_1, q_2)$ is represented by the mathematical line

$$\begin{aligned} l &= \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + \mathbb{R} \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \end{pmatrix} = \left\{ \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + \lambda \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 \end{pmatrix}; \lambda \in \mathbb{R} \right\} \\ &= \left\{ (1 - \lambda) \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + \lambda \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}; \lambda \in \mathbb{R} \right\} \end{aligned}$$

Using vectorial notation $\mathbf{p} = (p_1, p_2)$, $\mathbf{q} = (q_1, q_2)$ this simplifies to $l = \mathbf{p} + \mathbb{R}(\mathbf{q} - \mathbf{p}) = \{\mathbf{p} + \lambda(\mathbf{q} - \mathbf{p}); \lambda \in \mathbb{R}\}$ and can also be written as $l = \{\lambda_1 \mathbf{p} + \lambda_2 \mathbf{q}; \lambda_1, \lambda_2 \in \mathbb{R}, \lambda_1 + \lambda_2 = 1\}$.

The vector $\mathbf{u} = \mathbf{q} - \mathbf{p}$ is sometimes called *direction vector* of the line l . (Note, however, that any multiple $\lambda \mathbf{u}$, $\lambda \neq 0$, and in particular $-\mathbf{u}$ is a direction vector of l as well.)

- The *distance* between P and Q , relative to the chosen unit of measurement, is given by

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} = |\mathbf{p} - \mathbf{q}| = d(\mathbf{p} - \mathbf{q}, \mathbf{0}).$$

Here $|\mathbf{x}| = \sqrt{x_1^2 + x_2^2}$ is the *length* of the vector $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$.

- Lines $l_1 \triangleq \mathbf{p} + \mathbb{R}\mathbf{a}$ and $l_2 \triangleq \mathbf{p} + \mathbb{R}\mathbf{b}$ through the same point $P \triangleq \mathbf{p} = (p_1, p_2)$ (or their direction vectors \mathbf{a} and \mathbf{b}) are *orthogonal* (notation $l_1 \perp l_2$, respectively, $\mathbf{a} \perp \mathbf{b}$) iff

$$\mathbf{a} \cdot \mathbf{b} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} := a_1 b_1 + a_2 b_2 = 0.$$

It is a speciality of the 2-dimensional case that the line orthogonal to $\mathbf{p} + \mathbb{R} \left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \right)$ at the point \mathbf{p} is uniquely determined and given by $\mathbf{p} + \mathbb{R} \left(\begin{pmatrix} -a_2 \\ a_1 \end{pmatrix} \right)$.

- Generally, the *angle* $\phi \in [0, \pi]$ between l_1, l_2 at a point $\mathbf{p} \in l_1 \cap l_2$ is determined by

$$\cos \phi = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|}$$

The angle is *acute* if $\mathbf{a} \cdot \mathbf{b} > 0$ and *obtuse* if $\mathbf{a} \cdot \mathbf{b} < 0$.

For lines l_1, l_2 as above there are two angles of intersection at a point $\mathbf{p} \in l_1 \cap l_2$, which add up to 180° .

Reason: placing \mathbf{a} by $-\mathbf{a}$, which is also a direction vector of l_1 , changes the sign of $\mathbf{a} \cdot \mathbf{b}$ and turns the above equation into

$$\cos(\pi - \phi) = -\cos \phi = \frac{(-\mathbf{a}) \cdot \mathbf{b}}{|-\mathbf{a}| |\mathbf{b}|}.$$

- Every vector $\mathbf{a} \in \mathbb{R}^2$ is a positive multiple of a (uniquely determined) unit length vector:

$$\mathbf{a} = |\mathbf{a}| \left(\frac{1}{|\mathbf{a}|} \mathbf{a} \right) = |\mathbf{a}| \begin{pmatrix} \frac{a_1}{\sqrt{a_1^2 + a_2^2}} \\ \frac{a_2}{\sqrt{a_1^2 + a_2^2}} \end{pmatrix}.$$

The vector $\left(\frac{a_1}{\sqrt{a_1^2 + a_2^2}}, \frac{a_2}{\sqrt{a_1^2 + a_2^2}} \right)$ has length 1 and determines/is equal to the intersection point of the line $\mathbb{R}\mathbf{a}$ with the unit circle $x_1^2 + x_2^2 = 1$.

- The *area A* of the parallelogram spanned by vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ is given by the *2×2 -determinant*

$$A = \pm \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} := \pm (a_1 b_2 - a_2 b_1) = \pm \begin{pmatrix} -a_2 \\ a_1 \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

$$\text{Proof: } A^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 \sin^2 \phi = |\mathbf{a}|^2 |\mathbf{b}|^2 - |\mathbf{a}|^2 |\mathbf{b}|^2 \cos^2 \phi = (a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1 b_1 + a_2 b_2)^2 = (a_1 b_2 - a_2 b_1)^2.$$

We have $a_1 b_2 - a_2 b_1 = 0$ iff \mathbf{a}, \mathbf{b} are linearly dependent (i.e., scalar multiples of each other), and one can show that $a_1 b_2 - a_2 b_1 > 0$ iff the move from $\frac{1}{|\mathbf{a}|} \mathbf{a}$ to $\frac{1}{|\mathbf{b}|} \mathbf{b}$ on the shorter arc of the unit circle is counterclockwise.

- The *orthogonal projection* $\text{proj}_{\mathbf{a}}(\mathbf{b})$ of a vector $\mathbf{b} \in \mathbb{R}^2$ onto the line $\mathbb{R}\mathbf{a}$ (which passes through the origin) spanned by a vector $\mathbf{a} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}$ is the unique vector $\lambda\mathbf{a} \in \mathbb{R}\mathbf{a}$ satisfying $\mathbf{b} - \lambda\mathbf{a} \perp \mathbf{a}$; cf. picture on the next slide.

Solving for λ gives

$$(\mathbf{b} - \lambda\mathbf{a}) \cdot \mathbf{a} = 0 \iff \mathbf{b} \cdot \mathbf{a} - \lambda \mathbf{a} \cdot \mathbf{a} = 0 \iff \lambda = \lambda^* := \frac{\mathbf{b} \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}}$$

It follows that $\text{proj}_{\mathbf{a}}(\mathbf{b}) = \lambda^*\mathbf{a} = \frac{\mathbf{b} \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a}$.

The *distance* from \mathbf{b} to the line $\mathbb{R}\mathbf{a}$ is then computed as

$$\begin{aligned} d(\mathbf{b}, \mathbb{R}\mathbf{a}) &= \min \{ |\mathbf{b} - \lambda\mathbf{a}| ; \lambda \in \mathbb{R} \} \\ &= |\mathbf{b} - \lambda^*\mathbf{a}| = \sqrt{|\mathbf{b}|^2 - |\lambda^*\mathbf{a}|^2} \\ &= \sqrt{\mathbf{b} \cdot \mathbf{b} - \lambda^{*2} \mathbf{a} \cdot \mathbf{a}} = \sqrt{\mathbf{b} \cdot \mathbf{b} - \frac{(\mathbf{b} \cdot \mathbf{a})^2}{\mathbf{a} \cdot \mathbf{a}}}. \end{aligned}$$

Reasons: (i) $\mathbf{0}, \lambda^*\mathbf{a}, \mathbf{b}$ form a right triangle, implying

$|\mathbf{b}|^2 = |\lambda^*\mathbf{a}|^2 + |\mathbf{b} - \lambda^*\mathbf{a}|^2$ by Pythagoras' Law.

(ii) For any other point $\lambda\mathbf{a}$ on the line $\mathbb{R}\mathbf{a}$, the points $\lambda\mathbf{a}, \lambda^*\mathbf{a}, \mathbf{b}$ form another right triangle, and hence (Pythagoras' Law strikes again!) $|\mathbf{b} - \lambda\mathbf{a}|^2 = |\mathbf{b} - \lambda^*\mathbf{a}|^2 + |\lambda^*\mathbf{a} - \lambda\mathbf{a}|^2 > |\mathbf{b} - \lambda^*\mathbf{a}|^2$.

Also note that $|\lambda^* \mathbf{a}| = \frac{|\mathbf{b} \cdot \mathbf{a}|}{|\mathbf{a}|}$, and hence the angle ϕ between \mathbf{a} and \mathbf{b} is given by $\cos \phi = \frac{|\lambda^* \mathbf{a}|}{|\mathbf{b}|} = \frac{|\mathbf{b} \cdot \mathbf{a}|}{|\mathbf{a}| |\mathbf{b}|} = \frac{|\mathbf{a} \cdot \mathbf{b}|}{|\mathbf{a}| |\mathbf{b}|}$. Thus the stated formula for ϕ follows from elementary plane geometry and the formula for the orthogonal projection.

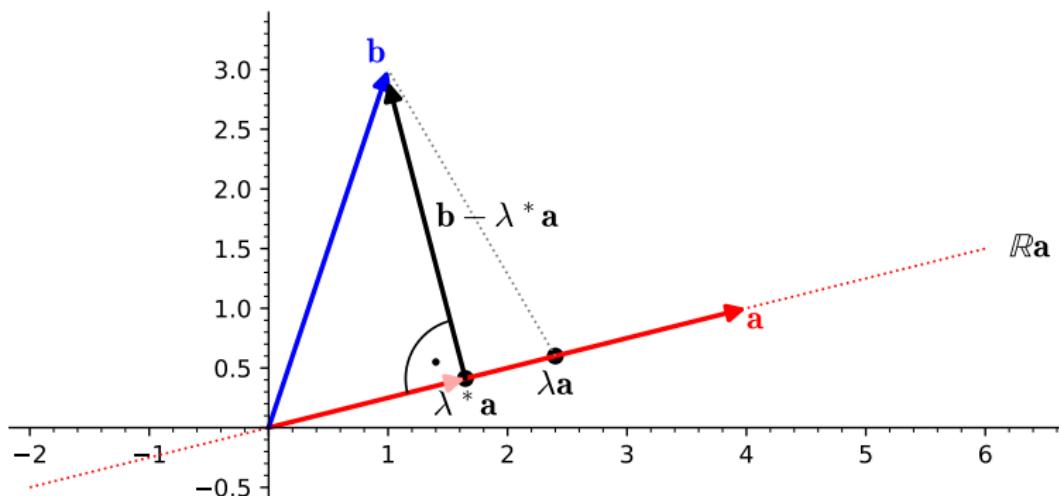


Figure: Illustration of $\text{proj}_{\mathbf{a}}(\mathbf{b}) = \lambda^* \mathbf{a}$, $\lambda^* = \frac{\mathbf{b} \cdot \mathbf{a}}{\mathbf{a} \cdot \mathbf{a}}$

In this example, $\mathbf{a} = (4, 1)$, $\mathbf{b} = (1, 3)$, and

$$\text{proj}_{\mathbf{a}}(\mathbf{b}) = \frac{7}{17}(4, 1) = \left(\frac{28}{17}, \frac{7}{17} \right) \approx (1.65, 0.41), \quad d(\mathbf{b}, \mathbb{R}\mathbf{a}) = \frac{11}{\sqrt{17}} \approx 2.68.$$

Switching between different coordinate systems

Setting $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (0, 1)$ (the so-called *standard basis vectors*), we have

$$P \triangleq \mathbf{p} = (p_1, p_2) = p_1 \mathbf{e}_1 + p_2 \mathbf{e}_2,$$

i.e., the coordinates of P are the coefficients in the unique representation of \mathbf{p} as a linear combination of the basis vectors $\mathbf{e}_1, \mathbf{e}_2$.

Now suppose we are given a second coordinate system of our physical plane.

Observation

$O' \triangleq \mathbf{p} = (p_1, p_2)$ for some $\mathbf{p} \in \mathbb{R}^2$ and the new unit coordinate directions are represented by $\mathbf{p} + \mathbb{R}\mathbf{a}$, $\mathbf{p} + \mathbb{R}\mathbf{b}$ with $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ orthogonal and of the same length.

\implies A point Q with new coordinates (x'_1, x'_2) has old coordinates

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{p} + x'_1 \mathbf{a} + x'_2 \mathbf{b} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + x'_1 \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + x'_2 \begin{pmatrix} b_1 \\ b_2 \end{pmatrix};$$

i.e., the two coordinate vectors are related linearly.

Example

Suppose that the new coordinate system has origin $\mathbf{p} = (-1, 1)$, and the unit directions of the new coordinate axes are $\mathbf{a} = (2, 1)$ and $\mathbf{b} = (-1, 2)$. Verify that the new system is indeed a valid coordinate system, and find the coordinates of the old origin $(0, 0)$ and the point $(1, 3)$ in the new coordinate system.

Since $\mathbf{a} \cdot \mathbf{b} = 2 \cdot 1 + (-1) \cdot 2 = 0$, $|\mathbf{a}| = \sqrt{2^2 + 1^2} = \sqrt{5} = |\mathbf{b}|$, the new system is a valid coordinate system.

The new coordinates (x'_1, x'_2) of $(0, 0)$ are determined by

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} + x'_1 \begin{pmatrix} 2 \\ 1 \end{pmatrix} + x'_2 \begin{pmatrix} -1 \\ 2 \end{pmatrix},$$

which is equivalent to the linear system

$$\begin{array}{rclcrcl} 2x'_1 & - & x'_2 & = & 1 \\ x'_1 & + & 2x'_2 & = & -1 \end{array}$$

Solving the system gives $x'_1 = \frac{1}{5}$, $x'_2 = -\frac{3}{5}$, so that $(0, 0)$ has new coordinates $(\frac{1}{5}, -\frac{3}{5})$.

Doing the same computation with $\mathbf{q} = (1, 3)$ gives the system $2x'_1 - x'_2 = 2 \wedge x'_1 + 2x'_2 = 2$, which is solved by $(x'_1, x'_2) = (\frac{6}{5}, \frac{2}{5})$.

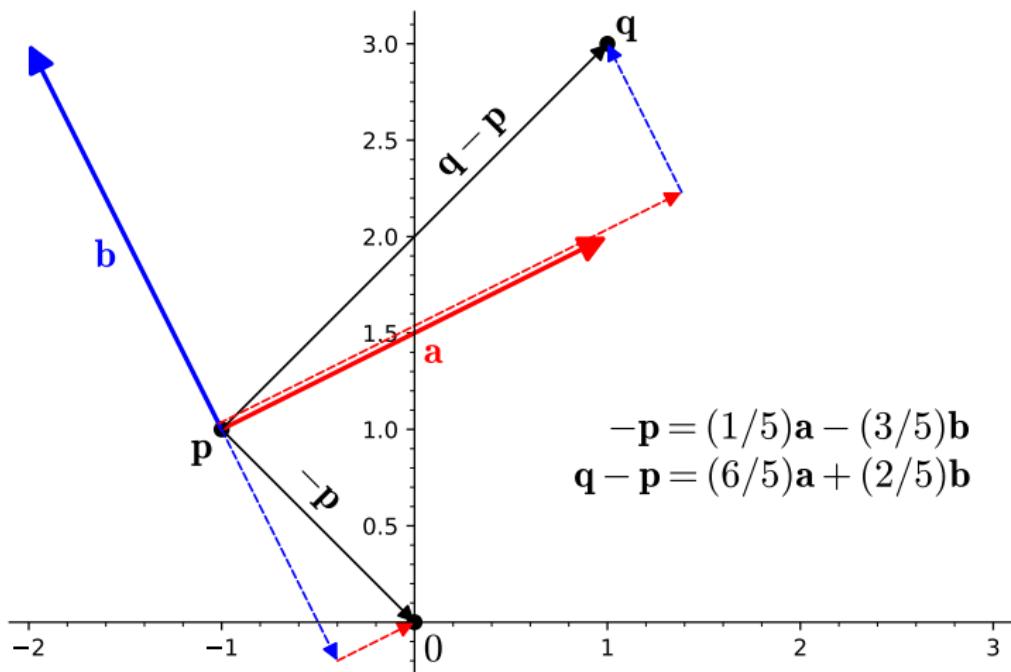


Figure: Illustration of coordinate change

The new coordinates of $\mathbf{x} = (x_1, x_2)$ are the coefficients x'_1, x'_2 in the representation $\mathbf{x} - \mathbf{p} = x'_1\mathbf{a} + x'_2\mathbf{b}$.

Notes

- In [Ste21] points in the plane are denoted, e.g., by $P(2, 1)$ (perhaps indicating that P is considered as a physical object), and the vector moving a point P to a point Q through vector addition is denoted by \overrightarrow{PQ} . In the lecture **points and vectors are the same thing**, and we write $P = (2, 1)$, or rather $\mathbf{p} = (2, 1)$, instead. The notation $\overrightarrow{\mathbf{pq}} = \overrightarrow{PQ}$ becomes unnecessary, since $\overrightarrow{\mathbf{pq}}$ is just $\mathbf{q} - \mathbf{p}$.
- The map $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, $(x'_1, x'_2) \mapsto (x_1, x_2)$ affording a coordinate change is a particular example of an *affine map*. It is a bijection (1-1 correspondence) and preserves lines and angles, but not necessarily distances. In fact distances are scaled by the constant factor $\sqrt{a_1^2 + a_2^2}$, as one can easily show. In the preceding example this factor is $\sqrt{5}$, so that $|\mathbf{x} - \mathbf{y}| = \sqrt{5} |\mathbf{x}' - \mathbf{y}'|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$.

For Three-Dimensional Space

- Any fixed physical point O , three pairwise orthogonal directions through O , and a unit of measurement determine a coordinate system.
- Points of the physical world (universe) are represented by triples $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$.
- The distance between $P \triangleq \mathbf{p} = (p_1, p_2, p_3)$ and $Q \triangleq \mathbf{q} = (q_1, q_2, q_3)$ is given by

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}.$$

Similarly for orthogonality, angles between vectors, etc.

However, \mathbb{R}^3 has a richer geometric structure than \mathbb{R}^2 , leading to additional geometric problems and corresponding algebraic concepts. One such concept is the cross product $\mathbf{a} \times \mathbf{b} \in \mathbb{R}^3$ of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$, which is discussed in Section 12.4 of our textbook and will be later in the lecture. Another one are distance computations between points, lines and planes (whereas in \mathbb{R}^2 the only such case is the distance from a point to a line).

Question

Your dormitory is $a = 4$ meters long, $b = 2$ meters wide and $c = 3$ meters high. What is the length of its (room) diagonals?

Answer

We don't presuppose the 3-dimensional distance formula, but derive it from Pythagoras' Theorem in the plane.

Denote the common length of the four diagonals by D and the length of a floor diagonal by d (measured in m). Since the floor is a rectangle with side lengths a and b , Pythagoras' Theorem gives $d^2 = a^2 + b^2$. Now look at the triangle formed by the end points P, Q of a floor diagonal and the vertex of the ceiling vertically above Q , say. This triangle is a right triangle with side lengths d , c , D , and hence $d^2 + c^2 = D^2$ by another application of Pythagoras' Theorem. Putting everything together gives $D^2 = a^2 + b^2 + c^2$, i.e., $D = \sqrt{a^2 + b^2 + c^2}$ as stated in the distance formula.

Thus the diagonals of your dormitory have length $\sqrt{4^2 + 2^2 + 3^2} = \sqrt{29} \approx 5.4$ [m].

Distance Computations in \mathbb{R}^3

We consider only the distance $d(l_1, l_2)$ of two lines $l_1 = \mathbf{a}_1 + \mathbb{R}\mathbf{u}_1$, $l_2 = \mathbf{a}_2 + \mathbb{R}\mathbf{u}_2$, which is defined as

$$\begin{aligned} d(l_1, l_2) &= \min \{ |\mathbf{x} - \mathbf{y}| ; \mathbf{x} \in l_1, \mathbf{y} \in l_2 \} \\ &= \min \{ |\mathbf{a}_1 + \lambda_1 \mathbf{u}_1 - \mathbf{a}_2 - \lambda_2 \mathbf{u}_2| ; \lambda_1, \lambda_2 \in \mathbb{R} \} \\ &= \min \{ |\mathbf{a}_1 - \mathbf{a}_2 - (\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2)| ; \lambda_1, \lambda_2 \in \mathbb{R} \}. \end{aligned}$$

(Note that the signs of λ_1, λ_2 do not matter, since $-\lambda$ runs through \mathbb{R} if λ does.)

The last expression shows $d(l_1, l_2) = d(\mathbf{b}, U)$ with $\mathbf{b} = \mathbf{a}_1 - \mathbf{a}_2$ and $U = \mathbb{R}\mathbf{u}_1 + \mathbb{R}\mathbf{u}_2$.

⇒ Computing distances between lines is equivalent to computing distances from a point to a plane through the origin (if $\mathbb{R}\mathbf{u}_1 \neq \mathbb{R}\mathbf{u}_2$, i.e., l_1 and l_2 are not parallel and U is a plane) or from a point to a line through the origin (if $\mathbb{R}\mathbf{u}_1 = \mathbb{R}\mathbf{u}_2$, i.e., l_1 and l_2 are parallel and $U = \mathbb{R}\mathbf{u}_1$ is a line).

Solution

As in the 2-dimensional case the key idea is to find a point $\mathbf{x} = \lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 \in U$ such that $\mathbf{b} - \mathbf{x}$ is orthogonal to \mathbf{u}_1 and \mathbf{u}_2 .

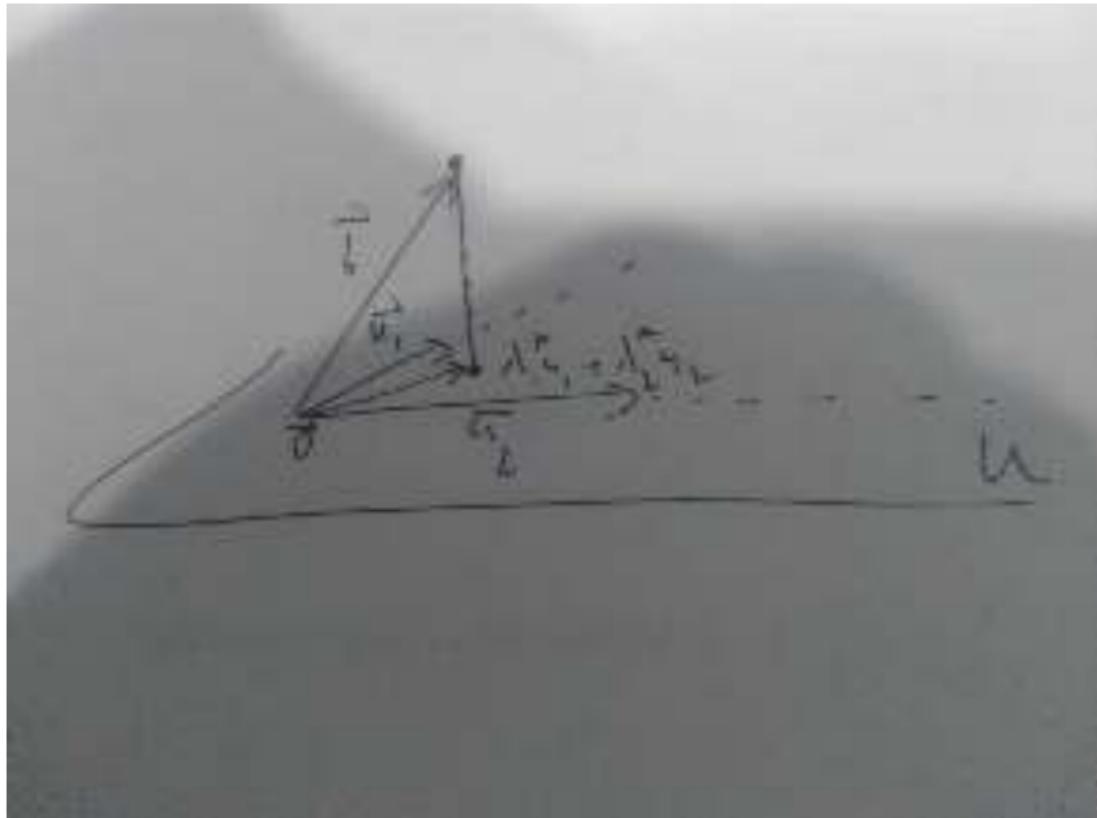


Figure: Orthogonal projection of \mathbf{b} to $U = \langle \mathbf{u}_1, \mathbf{u}_2 \rangle$

Solution cont'd

⇒ We need to solve the system of linear equations

$$\begin{cases} (\mathbf{b} - \lambda_1 \mathbf{u}_1 - \lambda_2 \mathbf{u}_2) \cdot \mathbf{u}_1 = 0 \\ (\mathbf{b} - \lambda_1 \mathbf{u}_1 - \lambda_2 \mathbf{u}_2) \cdot \mathbf{u}_2 = 0 \end{cases} \iff \begin{cases} (\mathbf{u}_1 \cdot \mathbf{u}_1)\lambda_1 + (\mathbf{u}_1 \cdot \mathbf{u}_2)\lambda_2 = \mathbf{u}_1 \cdot \mathbf{b} \\ (\mathbf{u}_1 \cdot \mathbf{u}_2)\lambda_1 + (\mathbf{u}_2 \cdot \mathbf{u}_2)\lambda_2 = \mathbf{u}_2 \cdot \mathbf{b} \end{cases}$$

This is a system of 2 linear equations for the 2 unknowns λ_1, λ_2 .

One can show that this system has a unique solution $(\lambda_1^*, \lambda_2^*)$, provided \mathbf{u}_1 and \mathbf{u}_2 are linearly independent (\rightarrow Cauchy-Schwarz Inequality). The corresponding point $\mathbf{x}^* = \lambda_1^* \mathbf{u}_1 + \lambda_2^* \mathbf{u}_2$ then satisfies $d(l_1, l_2) = d(\mathbf{b}, U) = |\mathbf{b} - \mathbf{x}^*|$.

If \mathbf{u}_1 and \mathbf{u}_2 are linearly dependent, the formula derived in the 2-dimensional case applies: $d(l_1, l_2) = d(\mathbf{b}, \mathbb{R}\mathbf{u}_1) = |\mathbf{b} - \lambda^* \mathbf{u}_1|$ with $\lambda^* = \frac{\mathbf{u}_1 \cdot \mathbf{b}}{\mathbf{u}_1 \cdot \mathbf{u}_1}$.

With a little imagination you can now guess how distance computations between general affine subspaces in \mathbb{R}^n are done and which formulas thereby arise.

The preceding argument shows that the function $\mathbf{x} \mapsto d(\mathbf{b}, \mathbf{x})$ attains a minimum on U . This fact is quite nontrivial (and false for many “nonlinear” subsets of \mathbb{R}^3). A valid definition of the distance between $U, V \subseteq \mathbb{R}^3$ (or \mathbb{R}^n) is $d(U, V) = \inf\{|\mathbf{x} - \mathbf{y}| ; \mathbf{x} \in U, \mathbf{y} \in V\}$.

Example

Determine the distance d between the lines

$$l_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \mathbb{R} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad l_2 = \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix} + \mathbb{R} \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix},$$

which turn out to be skew.

We have seen that d is equal to the distance between the point $\mathbf{b} = (0, 1, 0) - (3, 3, 0) = (-3, -2, 0)$ and the plane $U = \mathbb{R}(1, 1, 1) + \mathbb{R}(0, -2, 1)$. (It is clear that U is a plane in this case.)

The dot products we need are

$$\mathbf{u}_1 \cdot \mathbf{b} = (1, 1, 1) \cdot (-3, -2, 0) = -5,$$

$$\mathbf{u}_2 \cdot \mathbf{b} = (0, -2, 1) \cdot (-3, -2, 0) = 4,$$

$$\mathbf{u}_1 \cdot \mathbf{u}_1 = |(1, 1, 1)|^2 = 3,$$

$$\mathbf{u}_2 \cdot \mathbf{u}_2 = |(0, -2, 1)|^2 = 5,$$

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = (1, 1, 1) \cdot (0, -2, 1) = -1.$$

Example (cont'd)

This gives the linear system

$$\begin{array}{rcl} 3\lambda_1 & - & \lambda_2 = -5, \\ -\lambda_1 & + & 5\lambda_2 = 4. \end{array}$$

The solution is $\lambda_1^* = -\frac{3}{2}$, $\lambda_2^* = \frac{1}{2}$, from which obtain

$$\mathbf{x}^* = -\frac{3}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} -3/2 \\ -5/2 \\ -1 \end{pmatrix},$$

$$d = \left| \begin{pmatrix} -3 \\ -2 \\ 0 \end{pmatrix} - \begin{pmatrix} -3/2 \\ -5/2 \\ -1 \end{pmatrix} \right| = \left| \begin{pmatrix} -3/2 \\ 1/2 \\ 1 \end{pmatrix} \right| = \left| \frac{1}{2} \begin{pmatrix} -3 \\ 1 \\ 2 \end{pmatrix} \right| = \frac{1}{2} \sqrt{14}$$

Since $d > 0$ (equivalently, $\mathbf{b} - \mathbf{x}^* \neq \mathbf{0}$), the lines l_1 and l_2 are *skew* (i.e., disjoint as point sets).

The unique pair $\mathbf{p}_1 \in l_1$, $\mathbf{p}_2 \in l_2$ of points realizing the distance can be found by inserting $-\lambda_1^*$, λ_2^* into the parametric forms

$$l_1 = \mathbf{a}_1 + \mathbb{R}\mathbf{u}_1, \quad l_2 = \mathbf{a}_2 + \mathbb{R}\mathbf{u}_2.$$

Result: $\mathbf{p}_1 = \left(\frac{3}{2}, \frac{5}{2}, \frac{3}{2}\right)$, $\mathbf{p}_2 = \left(3, 2, \frac{1}{2}\right)$

Afternote

In [Ste21], p. 870 f you can find a different formula for the distance from a point to a plane in \mathbb{R}^3 , which uses a normal vector of the plane. For distance computations in \mathbb{R}^3 this method is admittedly shorter, but it has no analog in higher dimensions, where our method works just as well. Also it doesn't yield the points $\mathbf{p}_1, \mathbf{p}_2$ on l_1, l_2 realizing the distance.

The method in [Ste21] is based on the fact that in \mathbb{R}^3 there is a unique line through $\mathbf{0}$ orthogonal to U and that $\mathbf{b} - \mathbf{x}^*$ can also be computed as the orthogonal projection of \mathbf{b} to this line. A particular vector orthogonal to $\mathbf{u}_1, \mathbf{u}_2$ ("normal vector of U ") is $\mathbf{n} = (-3, 1, 2)$, and hence we have

$$d = \left| \frac{\mathbf{b} \cdot \mathbf{n}}{\|\mathbf{n}\|} \mathbf{n} \right| = \frac{|\mathbf{b} \cdot \mathbf{n}|}{\|\mathbf{n}\|} = \frac{(-3, -2, 0) \cdot (-3, 1, 2)}{\sqrt{14}} = \frac{7}{\sqrt{14}} = \frac{1}{2}\sqrt{14}.$$

The Link with Geometry

Vectors are just points

We know that points in the Euclidean plane are represented as (x_1, x_2) with $x_1, x_2 \in \mathbb{R}$ (or, in slightly different but equivalent notation, as (x, y) with $x, y \in \mathbb{R}$).

$\Rightarrow \mathbb{R}^2$ is a model for the Euclidean plane.

Similarly, $\mathbb{R}^3 = \{\mathbf{x} = (x_1, x_2, x_3); x_1, x_2, x_3 \in \mathbb{R}\}$ is a model for 3-dimensional Euclidean space.

Vectors are like arrows

Adding a vector $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ to the “origin” $\mathbf{0} = (0, 0)$ gives the point (v_1, v_2) . But \mathbf{v} can be added to other points (i.e., vectors), affecting the *translation*

$$T: \mathbb{R}^2 \rightarrow \mathbb{R}^2, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \mathbf{x} + \mathbf{v} = \begin{pmatrix} x_1 + v_1 \\ x_2 + v_2 \end{pmatrix}$$

The “displacement vector” \mathbf{v} of T can be recovered from any pair $\mathbf{x}, T(\mathbf{x})$ as $\mathbf{v} = T(\mathbf{x}) - \mathbf{x}$.

\Rightarrow We have a 1-1 correspondence between vectors in \mathbb{R}^2 and translations of the Euclidean plane; similarly for \mathbb{R}^3 , etc.

Problem

Consider the vectors $\mathbf{v} = (4, 2)$, $\mathbf{w} = (-1, 2)$ in \mathbb{R}^2 .

- Visualize $\mathbf{v} \pm \mathbf{w}$ in the Euclidean plane.
- Repeat a) with $\mathbf{v} = (3, 0)$.

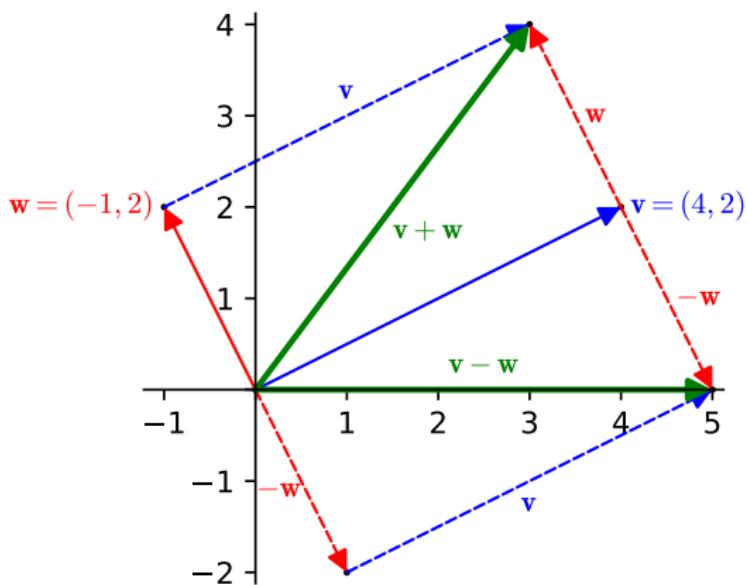


Figure: Solution to a)

Linear Combinations

Vector addition and scalar multiplication can be iterated.
For example, given $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ we can form the new vectors

$$2\mathbf{v} + \mathbf{w} = (2\mathbf{v}) + \mathbf{w} = \begin{pmatrix} 2v_1 \\ 2v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} 2v_1 + w_1 \\ 2v_2 + w_2 \end{pmatrix},$$

$$2\mathbf{v} - 3\mathbf{w} = (2\mathbf{v}) - 3\mathbf{w} = \begin{pmatrix} 2v_1 - 3w_1 \\ 2v_2 - 3w_2 \end{pmatrix}, \quad \text{and generally}$$

$$c\mathbf{v} + d\mathbf{w} = \begin{pmatrix} cv_1 + dw_1 \\ cv_2 + dw_2 \end{pmatrix} \quad \text{for } c, d \in \mathbb{R}.$$

Definition (linear combination)

Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r \in \mathbb{R}^n$ be vectors and $c_1, c_2, \dots, c_r \in \mathbb{R}$ be numbers. The vector

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_r\mathbf{v}_r \in \mathbb{R}^n$$

is called *linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ with coefficients c_1, c_2, \dots, c_r* .

Problem

Consider again $\mathbf{v} = (4, 2)$, $\mathbf{w} = (-1, 2)$ in \mathbb{R}^2 . Visualize the vectors $c\mathbf{v} + d\mathbf{w}$ for $c, d \in \{-2, -1, 0, 1, 2\}$ in the Euclidean plane.

Definition (subspace of \mathbb{R}^n)

A subset $U \subseteq \mathbb{R}^n$ is called a (*linear*) subspace of \mathbb{R}^n if $\mathbf{0} \in U$ and $\mathbf{u}, \mathbf{v} \in U$ implies $\mathbf{u} + \mathbf{v} \in U$ and $\lambda\mathbf{u} \in U$ for all $\lambda \in \mathbb{R}$.

Equivalently, $U \neq \emptyset$ and U is closed with respect to taking linear combinations of vectors in U .

Examples

- ① $\{\mathbf{0}\}$ and \mathbb{R}^n itself are subspaces of \mathbb{R}^n (the so-called *trivial* subspaces).
- ② Lines in \mathbb{R}^2 , as well as lines and planes in \mathbb{R}^3 , are subspaces of \mathbb{R}^2 , resp., \mathbb{R}^3 , iff they pass through the origin.
- ③ The set of all linear combinations $\lambda_1\mathbf{v}_1 + \cdots + \lambda_r\mathbf{v}_r$, $\lambda_i \in \mathbb{R}$, of a given set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ of vectors in \mathbb{R}^n forms a subspace of \mathbb{R}^n , the so-called (*linear*) *span* of $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$.

Definition (affine combination)

An *affine combination* of $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{R}^n$ is a linear combination $\lambda_1\mathbf{v}_1 + \dots + \lambda_r\mathbf{v}_r$ satisfying $\lambda_1 + \dots + \lambda_r = 1$.

Theorem

For a non-empty subset $A \subseteq \mathbb{R}^n$ the following are equivalent:

- ① $\mathbf{a}, \mathbf{b} \in A$ implies $\mathbf{a} + \mathbb{R}(\mathbf{b} - \mathbf{a}) \subseteq A$.
- ② A is closed with respect to taking affine combinations of vectors in A .
- ③ $A = \mathbf{a} + U$ for some linear subspace $U \subseteq \mathbb{R}^n$ and some (any) $\mathbf{a} \in A$.

Definition

A (non-empty) subset $A \subseteq \mathbb{R}^n$ is called an *affine subspace* of \mathbb{R}^n if it has the equivalent properties listed in the theorem.

Proof of the theorem.

(1) \implies (2): Since $\mathbf{a} + \mathbb{R}(\mathbf{b} - \mathbf{a}) = \{(1 - \lambda)\mathbf{a} + \lambda\mathbf{b}; \lambda \in \mathbb{R}\}$ is the set of affine combinations of \mathbf{a} and \mathbf{b} , A is closed w.r.t. taking affine combinations of 2 vectors. But one can easily show that an affine combination of $r > 2$ vectors can also be obtained by repeatedly taking affine combinations of 2 vectors.

Proof cont'd.

Here is the computation for $r = 3$: If $\lambda_1 + \lambda_2 + \lambda_3 = 1$ and $\lambda_1 + \lambda_2 \neq 0$, we can write

$$\lambda_1 \mathbf{a} + \lambda_2 \mathbf{b} + \lambda_3 \mathbf{c} = (\lambda_1 + \lambda_2) \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \mathbf{a} + \frac{\lambda_2}{\lambda_1 + \lambda_2} \mathbf{b} \right) + \lambda_3 \mathbf{c}.$$

This represents any affine combination of three vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ with $\lambda_3 \neq 1$ as an iteration of affine combinations involving only two vectors. Since not all of $\lambda_1, \lambda_2, \lambda_3$ can be equal to one, such a representation can always be found, permuting $\mathbf{a}, \mathbf{b}, \mathbf{c}$ first if necessary.

(2) \Rightarrow (1): Clear

(2) \Rightarrow (3): We set $U = \{\mathbf{a} - \mathbf{b}; \mathbf{a}, \mathbf{b} \in A\}$ and check the defining properties of subspaces.

$\mathbf{0} \in U$, since $A \neq \emptyset$ and hence $\mathbf{0} = \mathbf{a} - \mathbf{a}$ for some $\mathbf{a} \in A$.

Suppose $\mathbf{u} = \mathbf{a} - \mathbf{b} \in U$ (where $\mathbf{a}, \mathbf{b} \in A$) and $\lambda \in \mathbb{R}$.

$$\Rightarrow \lambda \mathbf{u} = \lambda(\mathbf{a} - \mathbf{b}) = \underbrace{\lambda \mathbf{a} + (1 - \lambda) \mathbf{b}}_{\in A} - \underbrace{\mathbf{b}}_{\in A} \in U$$

Proof cont'd.

Finally suppose $\mathbf{u} = \mathbf{a} - \mathbf{b}$, $\mathbf{v} = \mathbf{a}' - \mathbf{b}' \in U$.

$$\implies \frac{1}{2}(\mathbf{u} + \mathbf{v}) = \underbrace{\frac{1}{2}(\mathbf{a} + \mathbf{a}')}_{\in A} - \underbrace{\frac{1}{2}(\mathbf{b} + \mathbf{b}')}_{\in A} \in U.$$

Setting $\lambda = 2$ in the previous step, we obtain $\mathbf{u} + \mathbf{v} \in U$ as well. Hence U is a subspace.

The equation $A = \mathbf{a} + U = \{\mathbf{a} + \mathbf{u}; \mathbf{u} \in U\}$ is proved by writing $\mathbf{x} \in A$ as $\mathbf{x} = \mathbf{a} + \underbrace{\mathbf{x} - \mathbf{a}}_{\in U}$ and, for the inclusion $\mathbf{a} + U \subseteq A$, by

observing that $\mathbf{a} + \mathbf{b} - \mathbf{c}$ ($\mathbf{b}, \mathbf{c} \in A$) is an affine combination of $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

(3) \implies (2): Suppose $\mathbf{x} = \mathbf{a} + \mathbf{u}$, $\mathbf{y} = \mathbf{a} + \mathbf{v} \in A$ and $\lambda_1, \lambda_2 \in \mathbb{R}$ with $\lambda_1 + \lambda_2 = 1$.

$$\implies \lambda_1 \mathbf{x} + \lambda_2 \mathbf{y} = (\lambda_1 + \lambda_2) \mathbf{a} + \lambda_1 \mathbf{u} + \lambda_2 \mathbf{v} = \mathbf{a} + \lambda_1 \mathbf{u} + \lambda_2 \mathbf{v} \in A$$

i.e., $A = \mathbf{a} + U$ satisfies (2). □

Examples

- ① Lines in \mathbb{R}^2 , as well as lines and planes in \mathbb{R}^3 are affine subspaces.
- ② The set of all affine combinations of a given set $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ of vectors in \mathbb{R}^n forms an affine subspace of \mathbb{R}^n .

Example 1 contains only special cases of Example 2.

In order to understand Example 2, take e.g. $r = 3$. We can rewrite an affine combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ as

$$\begin{aligned}\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \lambda_3 \mathbf{v}_3 &= (1 - \lambda_2 - \lambda_3) \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \lambda_3 \mathbf{v}_3 \\ &= \mathbf{v}_1 + \lambda_2 (\mathbf{v}_2 - \mathbf{v}_1) + \lambda_3 (\mathbf{v}_3 - \mathbf{v}_1).\end{aligned}$$

This shows that the set consisting of all affine combinations of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ has the form $\mathbf{v}_1 + \mathbb{R}(\mathbf{v}_2 - \mathbf{v}_1) + \mathbb{R}(\mathbf{v}_3 - \mathbf{v}_1) = \mathbf{v}_1 + U$ with U a (linear) subspace of \mathbb{R}^n (the span of $\mathbf{v}_2 - \mathbf{v}_1$ and $\mathbf{v}_3 - \mathbf{v}_1$).

$\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ span a plane iff the “direction vectors” $\mathbf{v}_2 - \mathbf{v}_1, \mathbf{v}_3 - \mathbf{v}_1$ are linearly independent (i.e., not scalar multiples of each other).

Note that $\mathbf{v}_2 - \mathbf{v}_1$ and $\mathbf{v}_3 - \mathbf{v}_1$ normally don't belong to this plane (except for the case that the plane contains the origin). Rather, points on the plane are obtained by taking affine combinations of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, for example $\frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_2)$, $\frac{1}{3}(\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3)$, or $\frac{1}{2}\mathbf{v}_1 + \frac{1}{4}\mathbf{v}_2 + \frac{1}{4}\mathbf{v}_3$.

A Further Note

Last year I was asked by students how the concept of “line segment” fits into the framework of linear and affine subspaces. Here is the answer.

Recall that the *line segment* between (distinct) points $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ is defined as

$$\begin{aligned} [\mathbf{a}, \mathbf{b}] &= \{\mathbf{a} + \lambda(\mathbf{b} - \mathbf{a}); 0 \leq \lambda \leq 1\} \\ &= \{\lambda_1 \mathbf{a} + \lambda_2 \mathbf{b}; \lambda_1 + \lambda_2 = 1, \lambda_1, \lambda_2 \geq 0\}. \end{aligned}$$

Convex Sets and Convex Combinations

- Affine combinations with positive coefficients are called *convex combinations*. The line segment $[\mathbf{a}, \mathbf{b}]$ consists of all convex combinations of \mathbf{a} and \mathbf{b} .
- A set $K \subseteq \mathbb{R}^n$ is said to be *convex*, if $\mathbf{a}, \mathbf{b} \in K$ implies $[\mathbf{a}, \mathbf{b}] \subseteq K$. Thus K is convex if it is closed with respect to taking convex combinations of points in K . (Again this property transfers from combinations of 2 points to combinations of $r \geq 2$ points.)

Convex Sets cont'd

- The set of all convex combinations of fixed vectors $\mathbf{v}_1, \dots, \mathbf{v}_r \in \mathbb{R}^n$ is an example of a convex set. It is called *convex hull* of $\mathbf{v}_1, \dots, \mathbf{v}_r$. Examples are line segments (convex hull of 2 distinct points, viz. the endpoints), interiors of triangles (convex hull of 3 non-collinear points, viz. the vertices of the triangle), and the unit square $[0, 1]^2 \subset \mathbb{R}^2$ (convex hull of $(0, 0), (0, 1), (1, 0), (1, 1)$).
- Further examples of convex sets are provided by *balls* (such as $x^2 + y^2 < 1$ and $x^2 + y^2 \leq 1$ in \mathbb{R}^2), but there are many more.

Convex sets play an important role in Optimization Theory (e.g., Linear and Convex Programming). Occasionally we'll meet convex sets again as the course proceeds, but an in-depth discussion of them is beyond the scope of an undergraduate Calculus course.

Problem

- a) Given vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r \in \mathbb{R}^3$, determine the set of all linear combinations of these vectors (a set of vectors, and hence points, in \mathbb{R}^3).
- b) Given a further vector $\mathbf{a} \in \mathbb{R}^3$, determine $\{\mathbf{a} + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_r\mathbf{v}_r; c_i \in \mathbb{R}\}$.

Solution

Given the solution to Part a), Part b) is easy to solve:

$$\{\mathbf{a} + c_1\mathbf{v}_1 + \dots + c_r\mathbf{v}_r; c_i \in \mathbb{R}\} = T(\{c_1\mathbf{v}_1 + \dots + c_r\mathbf{v}_r; c_i \in \mathbb{R}\})$$

is the image of the point set in a) under the translation
 $T(\mathbf{x}) = \mathbf{x} + \mathbf{a}$.

The solution to a) is not easy. Only special cases are straightforward to solve:

Solution (cont'd)

$r = 1$ $\{c\mathbf{v}; c \in \mathbb{R}\} = \mathbb{R}\mathbf{v}$ is the line in \mathbb{R}^3 through the origin with "direction" vector \mathbf{v} if $\mathbf{v} \neq \mathbf{0}$, and reduces to the singleton set $\{\mathbf{0}\}$ (the origin) if $\mathbf{v} = \mathbf{0}$.

$r = 2$ W.l.o.g. we assume $\mathbf{v}_1, \mathbf{v}_2 \neq \mathbf{0}$;
 $S = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2; c_1, c_2 \in \mathbb{R}\}$ is a line (through the origin) if there exists $d \in \mathbb{R}$ such that $\mathbf{v}_2 = d\mathbf{v}_1$, i.e., if \mathbf{v}_1 and \mathbf{v}_2 are linearly dependent. If \mathbf{v}_1 and \mathbf{v}_2 are linearly independent then S is a plane through the origin (more precisely, the plane determined by the three points $\mathbf{0}, \mathbf{v}_1, \mathbf{v}_2$).

It turns out that in general (for arbitrary r) the point set $\{c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r; c_i \in \mathbb{R}\}$ is either a (i) point, or (ii) a line, or (iii) a plane, or (iv) the hole space \mathbb{R}^3 . Hence the same is true for the point set in b), the only difference being that this set does not necessarily contain the origin.

Problem

- a) Show that the vectors $\mathbf{v} = (1, 1, 0)$, $\mathbf{w} = (0, 1, 1)$ fill a plane and describe that plane.
- b) Find a vector that is not on this plane.

Solution

Since \mathbf{v} , \mathbf{w} are linearly independent (i.e., not scalar multiples of each other), they span a plane, viz.,

$$E = \{\lambda\mathbf{v} + \mu\mathbf{w}; \lambda, \mu \in \mathbb{R}\} = \{(\lambda, \lambda + \mu, \mu); \lambda, \mu \in \mathbb{R}\}.$$

If you are not yet convinced that E is a plane, observe that every vector in E is orthogonal to $(1, -1, 1)$, and conversely every vector $\mathbf{x} = (x_1, x_2, x_3)$ orthogonal to $(1, -1, 1)$, i.e.,

$x_1 - x_2 + x_3 = 0$, is in E :

$$\mathbf{x} = (x_1, x_1 + x_3, x_3) = x_1(1, 1, 0) + x_3(0, 1, 1).$$

Thus E consists of all points (vectors) satisfying the equation $x_1 - x_2 + x_3 = 0$. It is then trivial to find vectors not in E , e.g., $(1, 0, 0)$.

Equational Representation

The representation of lines/planes discussed so far may be called “parametric”. There exists alternative representations as solution sets of linear equations.

Planes

Every plane in \mathbb{R}^3 is the solution set of a single linear equation

$$a_1x_1 + a_2x_2 + a_3x_3 = b$$

with $a_1, a_2, a_3, b \in \mathbb{R}$ and not all a_i equal to zero (i.e., $(a_1, a_2, a_3) \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$).

Conversely, every such point set is a plane in \mathbb{R}^3 .

Linear equations $a_1x_1 + a_2x_2 + a_3x_3 = b$ and $a'_1x_1 + a'_2x_2 + a'_3x_3 = b'$ represent the same plane iff the vectors (a_1, a_2, a_3, b) and (a'_1, a'_2, a'_3, b') are scalar multiples of each other.

Lines

Every line in \mathbb{R}^3 is the intersection of 2 planes (in many ways) and hence the solution set of a system of 2 linear equations.

Conversely, the intersection of any 2 distinct planes in \mathbb{R}^3 is a line.

Points

Every point $\mathbf{v} = (v_1, v_2, v_3)$ in \mathbb{R}^3 is the intersection of 3 planes (in many ways). For example, we can take the planes with equations $x_1 = v_1$, $x_2 = v_2$, $x_3 = v_3$.

The proofs of all these facts become conceptual and easy, once we have developed the full machinery of Linear Algebra (but right now some of them are not easy).

Problem

- a) Find an equational representation of the plane with parametric representation

$$\mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix} + c_1 \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix} + c_2 \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}, \quad c_1, c_2 \in \mathbb{R}$$

- b) Find a parametric representation of the plane

$$x_1 + x_2 + x_3 = 1$$

- c) Explain how to make the equational representation of planes in \mathbb{R}^3 canonical (i.e., every plane should have a unique associated linear equation of the given form).

Problem

- a) Compute a parametric representation for the intersection of the two planes in \mathbb{R}^3 with equations $x_1 + x_2 - 2x_3 = 4$ and $-2x_1 - x_2 + 5x_3 = 0$, thereby showing that this intersection is a line.
- b) Represent the line in \mathbb{R}^3 through the two points $(1, 2, 1)$ and $(3, 0, -1)$ as solution set of a system of 2 linear equations.

Math 241
Calculus III

Thomas
Honold

Lengths and
Dot Products

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Lengths and Dot Products

Today's Lecture:

The Dot Product

Definition

The *dot product* (or *standard inner product*) of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ is defined as $\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$.

Properties

For all $\mathbf{a}_1, \mathbf{a}_2, \mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$ the following are true:

D1 $(\mathbf{a}_1 + \mathbf{a}_2) \cdot \mathbf{b} = \mathbf{a}_1 \cdot \mathbf{b} + \mathbf{a}_2 \cdot \mathbf{b};$

D2 $(c\mathbf{a}_1) \cdot \mathbf{b} = c(\mathbf{a}_1 \cdot \mathbf{b});$

D3 $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}.$

D4 $\mathbf{a} \cdot \mathbf{a} \geq 0$ with equality iff $\mathbf{a} = \mathbf{0}.$

(D1),(D2) are equivalent to

$$(c_1\mathbf{a}_1 + c_2\mathbf{a}_2) \cdot \mathbf{b} = c_1(\mathbf{a}_1 \cdot \mathbf{b}) + c_2(\mathbf{a}_2 \cdot \mathbf{b})$$

for all $\mathbf{a}_1, \mathbf{a}_2, \mathbf{b} \in \mathbb{R}^n$ and $c_1, c_2 \in \mathbb{R}$ (*linearity* of the dot product in the 1st argument). The symmetry property (D3) implies that the dot product is also linear in the 2nd argument, i.e., a so-called *bi-linear form*. Property (D4) is referred to as *positive definite*.

Orthogonality of Vectors

Definition

Two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ are said to be *orthogonal* or *perpendicular* if $\mathbf{a} \cdot \mathbf{b} = 0$.

This definition reduces to the usual concept of orthogonality in 2 and 3 dimensions.

Normal vectors in \mathbb{R}^3

The set of vectors $\mathbf{a} \in \mathbb{R}^3$ orthogonal to a given vector $\mathbf{a} \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$ is the plane with equation $\mathbf{a} \cdot \mathbf{x} = a_1x_1 + a_2x_2 + a_3x_3 = 0$.

For a general plane H with equation $a_1x_1 + a_2x_2 + a_3x_3 = b$ and an arbitrarily chosen point $\mathbf{p} \in H$ we have $a_1p_1 + a_2p_2 + a_3p_3 = b$ and hence

$$\begin{aligned}\mathbf{x} \in H &\iff a_1x_1 + a_2x_2 + a_3x_3 = a_1p_1 + a_2p_2 + a_3p_3 \\ &\iff a_1(x_1 - p_1) + a_2(x_2 - p_2) + a_3(x_3 - p_3) = 0 \\ &\iff \mathbf{a} \cdot (\mathbf{x} - \mathbf{p}) = 0.\end{aligned}$$

The vector \mathbf{a} is called *normal vector* of the plane H . It is unique up to scalar multiples and characterized by the fact that it is orthogonal to any difference ("direction vector") $\mathbf{x} - \mathbf{y}$ with $\mathbf{x}, \mathbf{y} \in H$.

Problem

Adapt the concept of a normal vector to the plane \mathbb{R}^2 . Which geometric objects have normal vectors in this case?

Problem

A plane H in \mathbb{R}^3 with equation $a_1x_1 + a_2x_2 + a_3x_3 = 0$ partitions the whole space into 3 sets:

$$H^+ = \{\mathbf{x} \in \mathbb{R}^3; a_1x_1 + a_2x_2 + a_3x_3 > 0\},$$

$$H = \{\mathbf{x} \in \mathbb{R}^3; a_1x_1 + a_2x_2 + a_3x_3 = 0\},$$

$$H^- = \{\mathbf{x} \in \mathbb{R}^3; a_1x_1 + a_2x_2 + a_3x_3 < 0\},$$

and similarly for lines in \mathbb{R}^2 . Can you distinguish the “halfspaces” H^+ and H^- geometrically by a property satisfied by $\mathbf{a} = (a_1, a_2, a_3)$ and the points in H^+ , H^- ?

Length and Distance

Definition

- ① The (*Euclidean*) *length* of a vector $\mathbf{a} \in \mathbb{R}^n$ is defined as

$$|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}.$$

- ② The (*Euclidean*) *distance* of two points (vectors) $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ is defined as the length of their difference, i.e.

$$d(\mathbf{a}, \mathbf{b}) = |\mathbf{a} - \mathbf{b}|.$$

Thus $d(\mathbf{a}, \mathbf{0}) = |\mathbf{a}|$ (length = distance from the origin) and $d(\mathbf{a} + \mathbf{c}, \mathbf{b} + \mathbf{c}) = |\mathbf{a} + \mathbf{c} - \mathbf{b} - \mathbf{c}| = d(\mathbf{a}, \mathbf{b})$ (invariance of the Euclidean distance under translations $\mathbf{x} \mapsto \mathbf{x} + \mathbf{c}$).

In 2 and 3 dimensions these definitions reduce to the usual concepts of length and distance (known perhaps already from high school).

Pythagoras' Law

Pythagoras' Law generalizes to \mathbb{R}^n : Using Properties (D1) and (D3) of the dot product, we have

$$\begin{aligned} |\mathbf{a} + \mathbf{b}|^2 &= (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) \\ &= \mathbf{a} \cdot \mathbf{a} + \mathbf{a} \cdot \mathbf{b} + \mathbf{b} \cdot \mathbf{a} + \mathbf{b} \cdot \mathbf{b} \\ &= |\mathbf{a}|^2 + 2(\mathbf{a} \cdot \mathbf{b}) + |\mathbf{b}|^2. \end{aligned}$$

If \mathbf{a} and \mathbf{b} are orthogonal, this reduces to

$$\begin{aligned} |\mathbf{a} + \mathbf{b}|^2 &= |\mathbf{a}|^2 + |\mathbf{b}|^2, \quad \text{and similarly} \\ |\mathbf{a} - \mathbf{b}|^2 &= |\mathbf{a}|^2 + |\mathbf{b}|^2. \end{aligned}$$

This says that the triangle with vertices $\mathbf{0}, \mathbf{a}, \mathbf{b}$, which is a right triangle with 3rd side $\pm(\mathbf{a} - \mathbf{b})$, satisfies Pythagoras' Law.

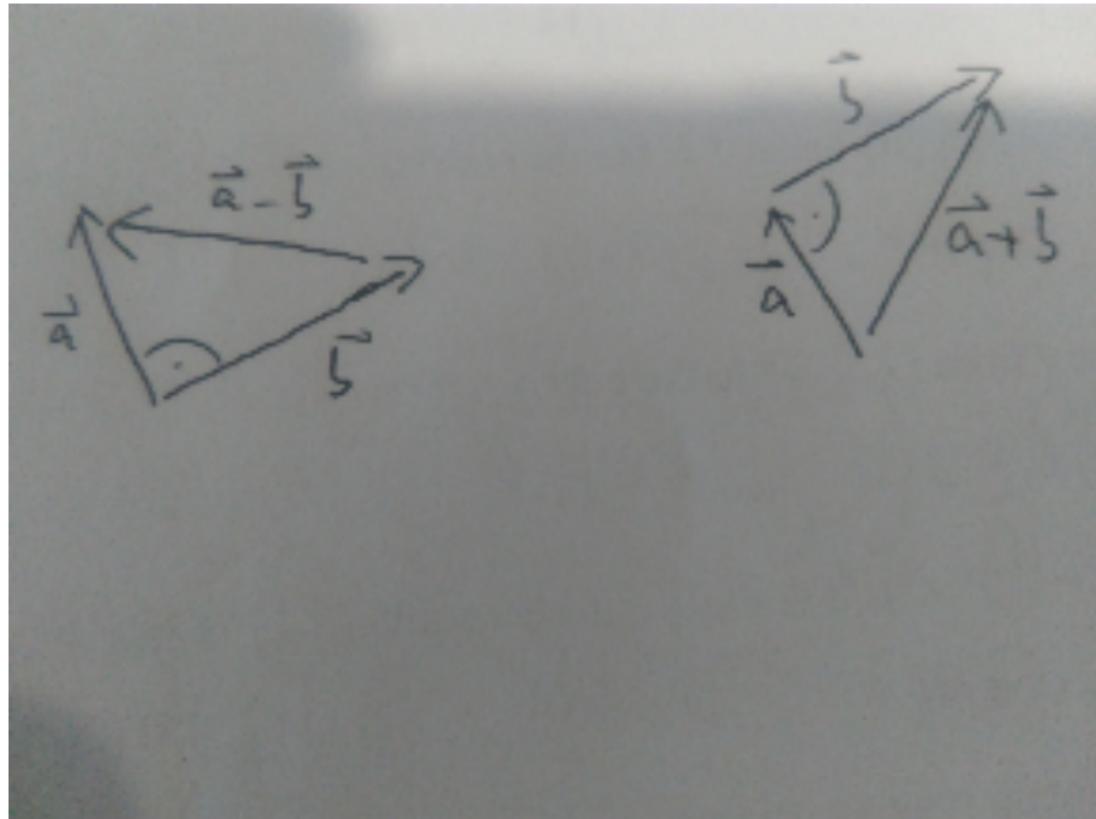


Figure: The two cases of Pythagoras' Law

The CAUCHY-SCHWARZ Inequality

Perhaps the most important inequality of Mathematics

Theorem

For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ we have $|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}| |\mathbf{b}|$.

Equality holds iff \mathbf{a} and \mathbf{b} are linearly dependent.

An equivalent form of the Cauchy-Schwarz Inequality is obtained by squaring:

$$(a_1 b_1 + \cdots + a_n b_n)^2 \leq (a_1^2 + \cdots + a_n^2)(b_1^2 + \cdots + b_n^2)$$

The case $n = 2$

Here it is $(a_1 b_1 + a_2 b_2)^2 \leq (a_1^2 + a_2^2)(b_1^2 + b_2^2)$ and expands into

$$a_1^2 b_1^2 + 2a_1 b_1 a_2 b_2 + a_2^2 b_2^2 \leq a_1^2 b_1^2 + (a_1^2 b_2^2 + a_2^2 b_1^2) + a_2^2 b_2^2$$

$$\iff a_1^2 b_2^2 + a_2^2 b_1^2 - 2a_1 a_2 b_1 b_2 \geq 0$$

$$\iff (a_1 b_2 - a_2 b_1)^2 = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}^2 \geq 0$$

Together with the well-known fact that a 2×2 determinant vanishes iff its columns (or rows) are linearly dependent, the theorem follows in this case.

Proof in the general case.

W.l.o.g assume $\mathbf{a} \neq \mathbf{0}$ and consider

$$Q(x) = |\mathbf{x}\mathbf{a} + \mathbf{b}|^2 = (\mathbf{x}\mathbf{a} + \mathbf{b}) \cdot (\mathbf{x}\mathbf{a} + \mathbf{b}).$$

Expanding the dot product, we see that this function is a quadratic $Ax^2 + Bx + C$:

$$Q(x) = (\mathbf{a} \cdot \mathbf{a})x^2 + 2(\mathbf{a} \cdot \mathbf{b})x + \mathbf{b} \cdot \mathbf{b},$$

and $A = \mathbf{a} \cdot \mathbf{a} > 0$ by (D4). The discriminant of $Q(x)$ is

$$\Delta = B^2 - 4AC = 4(\mathbf{a} \cdot \mathbf{b})^2 - 4(\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) = 4((\mathbf{a} \cdot \mathbf{b})^2 - |\mathbf{a}|^2 |\mathbf{b}|^2).$$

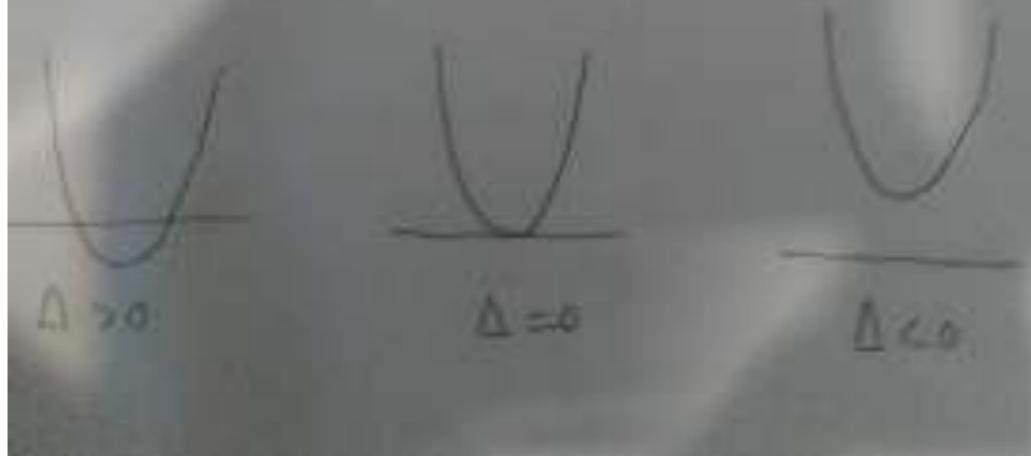
Property (D4) gives $Q(x) \geq 0$ for all $x \in \mathbb{R}$, and from high school we know that this implies $\Delta \leq 0$: A quadratic with no real zeros (exactly one real zero) has discriminant $\Delta < 0$ (resp., $\Delta = 0$).

This proves the first part of the theorem.

Now suppose equality holds, i.e., $\Delta = 0$. Then $Q(\lambda) = 0$ for some $\lambda \in \mathbb{R}$ (in fact $\lambda = -\frac{B}{2A} = -\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}}$), and hence $\lambda\mathbf{a} + \mathbf{b} = \mathbf{0}$ by (D4).

This shows that \mathbf{a} and \mathbf{b} are linearly dependent. Conversely, if $\mathbf{b} = \lambda\mathbf{a}$, say, then $|\mathbf{a} \cdot \mathbf{b}| = |\lambda(\mathbf{a} \cdot \mathbf{a})| = |\lambda| |\mathbf{a}^2| = |\mathbf{a}| |\lambda\mathbf{a}| = |\mathbf{a}| |\mathbf{b}|$. □

Quadratic Functions with Area
max $\Delta > 0$ - wide



Problem (Requires complex numbers)

The dot product of two complex vectors $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{C}^n$ and $\mathbf{w} = (w_1, \dots, w_n) \in \mathbb{C}^n$ is defined as

$\mathbf{z} \cdot \mathbf{w} = z_1\overline{w}_1 + z_2\overline{w}_2 + \dots + z_n\overline{w}_n$ and the length of $\mathbf{z} \in \mathbb{C}^n$ as
 $|\mathbf{z}| = \sqrt{\mathbf{z} \cdot \mathbf{z}}$.

- ① Derive properties analogous to (D1)–(D4) of the complex dot product.
- ② Show that the Cauchy-Schwarz Inequality generalizes to \mathbb{C}^n (where of course linear dependence of \mathbf{z}, \mathbf{w} over \mathbb{C} is involved in the 2nd part).

Problem

Prove the following quantitative version of the Cauchy-Schwarz Inequality:

$$\left(\sum_{i=1}^n a_i^2 \right) \left(\sum_{i=1}^n b_i^2 \right) - \left(\sum_{i=1}^n a_i b_i \right)^2 = \sum_{1 \leq i < j \leq n} (a_i b_j - a_j b_i)^2$$

for any choice of real (or complex) numbers $a_1, \dots, a_n, b_1, \dots, b_n$.

Consequences

The most important consequence of the Cauchy-Schwarz Inequality is the so-called triangle inequality for the Euclidean distance on \mathbb{R}^n .

Triangle Inequality

For $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$ we have

$$d(\mathbf{a}, \mathbf{b}) \leq d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b}).$$

Drawing a picture for $n = 2$ justifies the name “triangle inequality”.

Proof.

Setting $\mathbf{x} = \mathbf{a} - \mathbf{c}$, $\mathbf{y} = \mathbf{c} - \mathbf{b}$ changes the inequality into the equivalent inequality $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$, which is more convenient to prove. Squaring both sides of the latter inequality gives

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2(\mathbf{x} \cdot \mathbf{y}), \\ (|\mathbf{x}| + |\mathbf{y}|)^2 &= |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2|\mathbf{x}||\mathbf{y}|. \end{aligned}$$

⇒ The Cauchy-Schwarz Inequality implies the triangle inequality. □

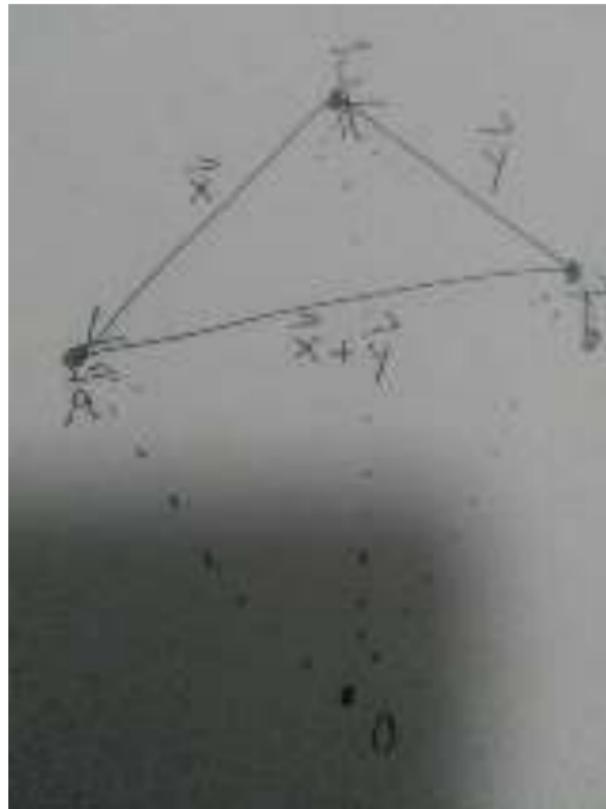


Figure: Illustration of the triangle inequality
 $d(\mathbf{a}, \mathbf{b}) \leq d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b})$

Angle between two Vectors

The Cauchy-Schwarz Inequality can be written in the form

$$-1 \leq \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|} \leq 1.$$

Definition

Suppose $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ are nonzero vectors. The angle $\theta \in [0, \pi]$ between \mathbf{a} and \mathbf{b} is defined by

$$\cos \theta = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|}.$$

Since $x \mapsto \cos x$ maps $[0, \pi]$ bijectively onto $[-1, 1]$, this is a valid definition. (It can also be stated as $\theta = \arccos\left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|}\right)$.)

The definition of θ implies

$|\mathbf{a} - \mathbf{b}|^2 = \mathbf{a}^2 + \mathbf{b}^2 - 2(\mathbf{a} \cdot \mathbf{b}) = |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta$ (*law of cosines*), and hence reduces in the plane case $n = 2$ to the usual definition of angles in plane geometry.

A Fact Used Earlier

When computing the distance between two lines with different directions $\mathbb{R}\mathbf{u}_1, \mathbb{R}\mathbf{u}_2$ in \mathbb{R}^3 , i.e., $\mathbf{u}_1, \mathbf{u}_2$ are linearly independent, we claimed that the system

$$(\mathbf{u}_1 \cdot \mathbf{u}_1)\lambda_1 + (\mathbf{u}_1 \cdot \mathbf{u}_2)\lambda_2 = \mathbf{u}_1 \cdot \mathbf{b}$$

$$(\mathbf{u}_1 \cdot \mathbf{u}_2)\lambda_1 + (\mathbf{u}_2 \cdot \mathbf{u}_2)\lambda_2 = \mathbf{u}_2 \cdot \mathbf{b}$$

has a unique solution. Now we can prove this:

Multiplying the first equation by $\mathbf{u}_2 \cdot \mathbf{u}_2$, the second by $\mathbf{u}_1 \cdot \mathbf{u}_2$, and subtracting, we get

$$[(\mathbf{u}_1 \cdot \mathbf{u}_1)(\mathbf{u}_2 \cdot \mathbf{u}_2) - (\mathbf{u}_1 \cdot \mathbf{u}_2)^2]\lambda_1 = (\mathbf{u}_2 \cdot \mathbf{u}_2)(\mathbf{u}_1 \cdot \mathbf{b}) - (\mathbf{u}_1 \cdot \mathbf{u}_2)(\mathbf{u}_2 \cdot \mathbf{b}),$$

and similarly

$$[(\mathbf{u}_1 \cdot \mathbf{u}_1)(\mathbf{u}_2 \cdot \mathbf{u}_2) - (\mathbf{u}_1 \cdot \mathbf{u}_2)^2]\lambda_2 = -(\mathbf{u}_1 \cdot \mathbf{u}_2)(\mathbf{u}_1 \cdot \mathbf{b}) + (\mathbf{u}_1 \cdot \mathbf{u}_1)(\mathbf{u}_2 \cdot \mathbf{b}).$$

The Cauchy-Schwarz Inequality gives

$(\mathbf{u}_1 \cdot \mathbf{u}_1)(\mathbf{u}_2 \cdot \mathbf{u}_2) - (\mathbf{u}_1 \cdot \mathbf{u}_2)^2 > 0$, and hence the system has a unique solution.

Math 241
Calculus III

Thomas
Honold

Determinants
and Cross
Products

Determinants
Cross Products

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Determinants and Cross Products

- Determinants
- Cross Products

Determinants
and Cross
Products

Determinants
Cross Products

Today's Lecture:

Determinants

Although determinants are defined for arbitrary square matrices, we will restrict ourselves in the examples to the cases of 2×2 and 3×3 matrices. (In the exercises you have to compute some 4×4 determinants.)

Definition

- For $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbb{R}^{2 \times 2}$ we define its *determinant* as

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}.$$

- For $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \in \mathbb{R}^{3 \times 3}$ we define

$$\begin{aligned}\det(\mathbf{A}) = & + a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}.\end{aligned}$$

Instead of $\det(\mathbf{A})$ one also writes $|\mathbf{A}|$.

Question

What is the general rule behind these definitions?

Answer

- For each permutation π of $\{1, 2, \dots, n\}$, the determinant contains exactly one summand of the form

$$\pm a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)}.$$

- The sign of the term is $+1$ or -1 according to whether the permutation is *even* or *odd*.

Definition (sign of a permutation)

A permutation of $\{1, 2, \dots, n\}$ is *even* if it is composed of an even number of 2-element swaps $i \leftrightarrow j$, and *odd* otherwise.

2-element swaps $i \leftrightarrow j$ (which leave all letters in $\{1, \dots, n\} \setminus \{i, j\}$ untouched) are called *transpositions* and are denoted by (i, j) .

It can be shown that every permutation π of $\{1, 2, \dots, n\}$ can be represented as a composition $\pi = \tau_1 \circ \tau_2 \circ \cdots \circ \tau_r$ of transpositions and that the parity of r in all such representations must be the same.

Example ($n = 3$)

There are $3! = 6$ permutations of $\{1, 2, 3\}$. The following table contains their array representations $\pi \triangleq \begin{pmatrix} 1 & 2 & 3 \\ \pi(1) & \pi(2) & \pi(3) \end{pmatrix}$ and their signs.

π	$(1 \ 2 \ 3)$	$(1 \ 2 \ 3)$	$(1 \ 2 \ 3)$	$(1 \ 2 \ 3)$	$(1 \ 2 \ 3)$	$(1 \ 2 \ 3)$
sign	+1	+1	+1	-1	-1	-1

The signs of the first permutation (identity map) and the 4th, 5th, 6th permutation (which are transpositions) are clear.

The 2nd and 3rd permutation are *cyclic*, viz.

$1 \mapsto 2 \mapsto 3 \mapsto 1$ and $1 \mapsto 3 \mapsto 2 \mapsto 1$. Representing them as $(1, 2, 3)$ and $(1, 3, 2)$, respectively, we have

$$(1, 2, 3) = (1, 2) \circ (2, 3),$$

$$(1, 3, 2) = (1, 3) \circ (3, 2),$$

where $\sigma \circ \tau$ means “first apply τ , then σ ”. Hence their signs are +1.

Properties of the Determinant

Algebraic properties (selection)

- 1 $\det(\mathbf{A}) = \det(\mathbf{A}^T)$; this is obvious for 2×2 matrices and can be checked with a little effort for 3×3 matrices.
- 2 The determinant is well-behaved with respect to elementary row (and column) operations: If \mathbf{B} arises from \mathbf{A} by an elementary row operation then according to the type
 - (i) $\det(\mathbf{B}) = -\det(\mathbf{A})$ $(R_i \leftrightarrow R_j)$
 - (ii) $\det(\mathbf{B}) = c \det(\mathbf{A})$ $(R_i = c R_i)$
 - (iii) $\det(\mathbf{B}) = \det(\mathbf{A})$ $(R_i = R_j + c R_i)$
- 3 The determinant of an upper (or lower) triangular matrix is the product of the entries on the main diagonal. (This follows, e.g., from Laplace expansion; see below.)
- 4 \mathbf{A} is invertible iff $\det(\mathbf{A}) \neq 0$.
- 5 $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ for all $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$.

Algebraic properties cont'd

- 6 *Laplace expansion along the first row:*

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix};$$

similarly for other rows or columns and for general n ; cf. next slide.

Notes

- Proofs of the properties can be found in any Linear Algebra book.
- Properties (2) and (3) provide an efficient method to compute the determinant of $\mathbf{A} \in \mathbb{R}^{n \times n}$: Use elementary row operations to transform \mathbf{A} into \mathbf{A}' in row-echelon form and record for each operation the change of the determinant according to (2). By (3), $\det(\mathbf{A}')$ is the product of the diagonal entries of \mathbf{A}' , which is equal to the product of the pivots if $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A}') = n$ and zero otherwise. From $\det(\mathbf{A}')$ we can then obtain $\det(\mathbf{A})$.
- For large n it is very inefficient to compute $\det(\mathbf{A})$ directly from the defining formula, which has $n!$ summands.

Notes cont'd

- The general Laplace expansion rule is

$$\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij}) \quad (\text{for Row } i)$$

$$= \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij}), \quad (\text{for Column } j)$$

where A_{ij} denotes the $(n - 1) \times (n - 1)$ submatrix of \mathbf{A} obtained by deleting Row i and Column j .

The resulting sign pattern is analogous to a chessboard with a “+” in the top-left corner, e.g.,

$$\begin{pmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{pmatrix}$$

for 4×4 matrices.

Notes cont'd

- The set S_n of all permutations of $\{1, 2, \dots, n\}$ forms a group under map composition, the so-called *symmetric group of degree n*. The defining formula for determinants, $\det(\mathbf{A}) = \sum_{\pi \in S_n} (-1)^\pi a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)}$ for $\mathbf{A} \in \mathbb{R}^{n \times n}$, is attributed to LEIBNIZ. It can be derived in a more conceptual way from the requirement that $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}$, $\mathbf{A} \mapsto \det(\mathbf{A})$ is a so-called *alternating multilinear form* on the space \mathbb{R}^n of column vectors, i.e., it satisfies

- (D1) For each $j \in \{1, \dots, n\}$ and any choice of vectors $\mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n \in \mathbb{R}^n$, the map $\mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{x} \rightarrow \det(\mathbf{a}_1 | \dots | \mathbf{a}_{j-1} | \mathbf{x} | \mathbf{a}_{j+1} | \dots | \mathbf{a}_n)$ is linear;
- (D2) $\det(\mathbf{A}) = 0$ whenever two columns of \mathbf{A} are equal.

It is fairly easy to see that (D1), (D2) imply

$$\det(\mathbf{A}) = \alpha \sum_{\pi \in S_n} (-1)^\pi a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)}, \quad \alpha = \det(\mathbf{I}_n),$$

and hence $\mathbf{A} \mapsto \det(\mathbf{A})$ is uniquely determined by (D1), (D2) and

- (D3) $\det(\mathbf{I}_n) = 1$.

Example

We compute the determinant of $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & 2 & -1 \end{pmatrix}$ in four different ways:

(1) Expansion along the first row:

$$\begin{aligned} \left| \begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & 2 & -1 \end{array} \right| &= 1 \cdot \left| \begin{array}{cc} 5 & 6 \\ 2 & -1 \end{array} \right| - 2 \left| \begin{array}{cc} 4 & 6 \\ -1 & -1 \end{array} \right| + 3 \left| \begin{array}{cc} 4 & 5 \\ -1 & 2 \end{array} \right| \\ &= 1 \cdot (-5 - 12) - 2(-4 + 6) + 3(8 + 5) \\ &= -17 - 4 + 39 = 18 \end{aligned}$$

(2) Expansion along the second column:

$$\begin{aligned} \left| \begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & 2 & -1 \end{array} \right| &= -2 \left| \begin{array}{cc} 4 & 6 \\ -1 & -1 \end{array} \right| + 5 \left| \begin{array}{cc} 1 & 3 \\ -1 & -1 \end{array} \right| - 2 \left| \begin{array}{cc} 1 & 3 \\ 4 & 6 \end{array} \right| \\ &= -2(-4 + 6) + 5(-1 + 3) - 2(6 - 12) \\ &= 1 - 4 + 10 + 12 = 18 \end{aligned}$$

Example (cont'd)

(3) Using Gaussian elimination:

$$\left| \begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & 2 & -1 \end{array} \right| = \left| \begin{array}{ccc} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 4 & 2 \end{array} \right| \quad (R2 = R2 - 4R1, R3 = R3 + R1)$$
$$= 3 \left| \begin{array}{ccc} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 4 & 2 \end{array} \right| \quad (R2 = \frac{1}{3}R2)$$
$$= 3 \left| \begin{array}{ccc} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & 0 & -6 \end{array} \right| \quad (R3 = R3 + 4R2)$$
$$= 3 \cdot 1(-1)(-6) = 18$$

Alice's Way to Compute a 3×3 Determinant

Example (cont'd)

(4) Using the following rather weird formula, valid only if the central matrix element a_{22} is nonzero:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \frac{1}{a_{22}} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}.$$

In the case under consideration we obtain

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & 2 & -1 \end{vmatrix} = \frac{1}{5} \begin{vmatrix} 1 \cdot 5 - 4 \cdot 2 & 2 \cdot 6 - 5 \cdot 3 \\ 4 \cdot 2 - (-1)5 & 5(-1) - 2 \cdot 6 \end{vmatrix}$$

$$= \frac{1}{5} \begin{vmatrix} -3 & -3 \\ 13 & -17 \end{vmatrix} = \frac{(-3)(-17) - 13(-3)}{5} = \frac{90}{5} = 18.$$

The formula is due to CHARLES LUTWIDGE DODGSON alias LEWIS CARROLL, author of *Alice's Adventures in Wonderland*.

Notes

- Gaussian elimination is the way to go for general $n \times n$ matrices with large n (starting with $n = 4$).
- When doing hand computations, you can also use a mix of elementary operations (both row and column operations) and Laplace expansion, e.g.,

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ -1 & 2 & -1 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 4 & 2 \end{vmatrix}$$
$$= 1 \cdot \begin{vmatrix} -3 & -6 \\ 4 & 2 \end{vmatrix} \quad (\text{Expand along 1st column})$$
$$= \dots$$

Geometric property

$|\det \mathbf{A}|$ is equal to the volume of the parallelepiped spanned by the rows (or columns) of \mathbf{A} .

Here, the *parallelepiped* $P = P(\mathbf{v}_1, \dots, \mathbf{v}_n)$ spanned by $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$ is defined as

$$P = \{c_1 \mathbf{v}_1 + \cdots + c_n \mathbf{v}_n; 0 \leq c_i \leq 1\}.$$

We shall prove this now for $n = 2$ and later (after the introduction of cross products) for $n = 3$.

Proof for $n = 2$.

Write $\mathbf{v}_1 = \mathbf{a}$, $\mathbf{v}_2 = \mathbf{b}$. From plane geometry we know $\text{vol}(P) = |\mathbf{a}| |\mathbf{b}| \sin \phi$, where $\phi \in [0, \pi]$ is the angle between \mathbf{a} and \mathbf{b} .

$$\begin{aligned}\implies \text{vol}(P)^2 &= |\mathbf{a}|^2 |\mathbf{b}|^2 \sin^2 \phi = |\mathbf{a}|^2 |\mathbf{b}|^2 - |\mathbf{a}|^2 |\mathbf{b}|^2 \cos^2 \phi \\ &= |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2 \\ &= (a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1 b_1 + a_2 b_2)^2 \\ &= (a_1 b_2 - a_2 b_1)^2,\end{aligned}\quad \text{as seen earlier.}$$

Cross Products

Definition

Suppose $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$. The *cross product* (or *vector product*) of \mathbf{a} and \mathbf{b} is the vector

$$\begin{aligned}\mathbf{a} \times \mathbf{b} &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix} \\ &= \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} \mathbf{e}_1 - \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} \mathbf{e}_2 + \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \mathbf{e}_3 \in \mathbb{R}^3.\end{aligned}$$

Mnemonic

$\mathbf{a} \times \mathbf{b}$ is obtained by formally expanding the “determinant”

$$\begin{vmatrix} a_1 & b_1 & \mathbf{e}_1 \\ a_2 & b_2 & \mathbf{e}_2 \\ a_3 & b_3 & \mathbf{e}_3 \end{vmatrix}$$

along the last column.

Geometric characterization

If \mathbf{a} and \mathbf{b} are linearly dependent then $\mathbf{a} \times \mathbf{b} = \mathbf{0}$.

Otherwise $\mathbf{a} \times \mathbf{b}$ is a nonzero vector orthogonal to \mathbf{a} and \mathbf{b} (hence contained in the line through the origin perpendicular to the plane generated by \mathbf{a}, \mathbf{b}), has length $|\mathbf{a}| |\mathbf{b}| \sin \phi$ (area of the parallelogram spanned by \mathbf{a}, \mathbf{b}), and the ordered basis $\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}$ has the same orientation as the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ (determined by the “right-hand rule”).

Proof.

Orthogonality to \mathbf{a}, \mathbf{b} is equivalent to

$$\begin{vmatrix} a_1 & b_1 & a_1 \\ a_2 & b_2 & a_2 \\ a_3 & b_3 & a_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & b_1 \\ a_2 & b_2 & b_2 \\ a_3 & b_3 & b_3 \end{vmatrix} = 0.$$

These identities can be verified directly, but also follow from the invariance of the determinant under elementary column operations (e.g., in the first matrix subtract Column 1 from Column 3 to obtain a matrix with a zero column, which clearly has determinant zero).

Proof cont'd.

The assertion about the length of $\mathbf{a} \times \mathbf{b}$ uses the formula

$$\begin{aligned} |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2 &= |\mathbf{a} \times \mathbf{b}|^2 \\ &= (a_2 b_3 - a_3 b_2)^2 + (a_1 b_3 - a_3 b_1)^2 + (a_1 b_2 - a_2 b_1)^2, \end{aligned}$$

which represents the case $n = 3$ of an exact formula for the error in the Cauchy-Schwarz Inequality stated earlier.

As in the case $n = 2$ we can rewrite this as

$$\begin{aligned} |\mathbf{a} \times \mathbf{b}|^2 &= |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 - |\mathbf{a}|^2 |\mathbf{b}|^2 \cos^2 \phi \\ &= |\mathbf{a}|^2 |\mathbf{b}|^2 \sin^2 \phi, \quad \text{proving the assertion.} \end{aligned}$$

For the last assertion it suffices to show $\det(\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}) > 0$.

$$\begin{aligned} \det(\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}) &= \begin{vmatrix} a_1 & b_1 & a_2 b_3 - a_3 b_2 \\ a_2 & b_2 & a_3 b_1 - a_1 b_3 \\ a_3 & b_3 & a_1 b_2 - a_2 b_1 \end{vmatrix} \\ &= (a_2 b_3 - a_3 b_2)^2 + (a_1 b_3 - a_3 b_1)^2 + (a_1 b_2 - a_2 b_1)^2 = |\mathbf{a} \times \mathbf{b}|^2 > 0, \end{aligned}$$

using Laplace expansion along the last column. □

The formula just used is a special case of the following.

Triple products

For $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ we have

$$\det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}.$$

This quantity (a real number) is called *triple product* in [Ste21].

The proof in the general case is the same:

$$\begin{aligned} \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} &= c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} - c_2 \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} + c_3 \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \\ &= \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}. \end{aligned}$$

Proof of the volume formula for $n = 3$.

Let $P = P(\mathbf{a}, \mathbf{b}, \mathbf{c})$ be the parallelepiped spanned by $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ (where w.l.o.g. $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent).

The height of P relative to the plane spanned by \mathbf{a}, \mathbf{b} is $|\mathbf{c}| |\cos \psi|$, where ψ is the angle between \mathbf{c} and $\mathbf{a} \times \mathbf{b}$.

$$\begin{aligned}\implies \text{vol}(P) &= \text{area of the base} \times \text{height} \\ &= \text{vol}(P(\mathbf{a}, \mathbf{b})) \times |\mathbf{c}| |\cos \psi| \\ &= |\mathbf{a} \times \mathbf{b}| |\mathbf{c}| |\cos \psi| \\ &= |(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}| \\ &= |\det(\mathbf{a}, \mathbf{b}, \mathbf{c})|.\end{aligned}$$



Using $|\cos \psi|$ is necessary, since $\psi > \pi/2$ is possible (i.e., the angle between \mathbf{c} and $\mathbf{a} \times \mathbf{b}$ can be obtuse). This occurs iff $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are negatively oriented.

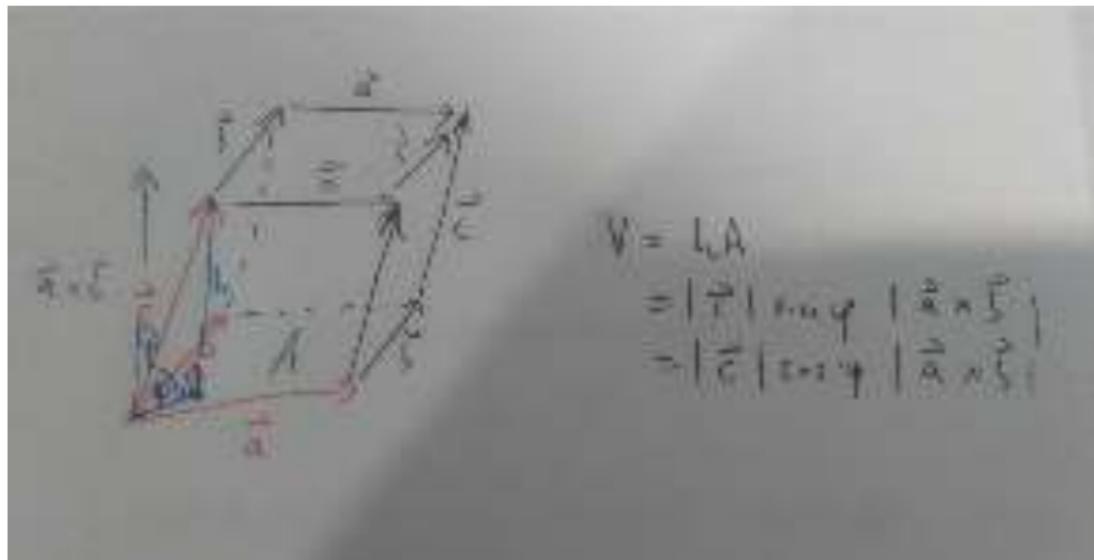


Figure: Volume of the parallelepiped spanned by $\mathbf{a}, \mathbf{b}, \mathbf{c}$

The picture illustrates the case $\psi \in [0, \pi/2]$. For $\psi \in (\pi/2, \pi]$ the formula $V = -|\mathbf{c}| \cos \psi |\mathbf{a} \times \mathbf{b}|$ is correct.

Example

We compute the volume of the tetrahedron T in \mathbb{R}^3 with vertices $\mathbf{v}_1 = (1, 0, 0)$, $\mathbf{v}_2 = (0, 1, 0)$, $\mathbf{v}_3 = (0, 0, 1)$, $\mathbf{v}_4 = (1, 1, 1)$.

First we verify that T is indeed a tetrahedron:

$$d(\mathbf{v}_1, \mathbf{v}_2) = |\mathbf{v}_1 - \mathbf{v}_2| = |(1, -1, 0)| = \sqrt{2}$$

$$d(\mathbf{v}_1, \mathbf{v}_3) = |(1, 0, -1)| = \sqrt{2}$$

$$d(\mathbf{v}_2, \mathbf{v}_3) = |(0, 1, -1)| = \sqrt{2}$$

$$d(\mathbf{v}_1, \mathbf{v}_4) = |(0, 1, 1)| = \sqrt{2}$$

$$d(\mathbf{v}_2, \mathbf{v}_4) = |(1, 0, 1)| = \sqrt{2}$$

$$d(\mathbf{v}_3, \mathbf{v}_4) = |(1, 1, 0)| = \sqrt{2}$$

$\implies T$ is a tetrahedron, whose edges have length $\sqrt{2}$.

The volume is obtained by moving one vertex, say \mathbf{v}_4 , to the origin and applying the usual formula for the volume of a pyramid:

$$\text{vol}(T) = \frac{1}{6} \text{vol } P(\mathbf{v}_1 - \mathbf{v}_4, \mathbf{v}_2 - \mathbf{v}_4, \mathbf{v}_3 - \mathbf{v}_4) = \frac{1}{6} \begin{vmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{vmatrix} = \frac{1}{3}.$$

Exercise

Show that the volume of a possibly irregular tetrahedron T (i.e., the faces need not be equilateral triangles) spanned by four arbitrary points $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \in \mathbb{R}^3$ can be defined by the same formula:

$$\text{vol}(T) = \frac{1}{6} |\det(\mathbf{v}_1 - \mathbf{v}_4 | \mathbf{v}_2 - \mathbf{v}_4 | \mathbf{v}_3 - \mathbf{v}_4)|. \quad (*)$$

Hint: The key point is to show that the right-hand side of $(*)$ doesn't depend on the order in which $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ are listed.

Returning to Cross Products

Algebraic Properties

A list of important properties of the cross product can be found in [Ste21], Theorem 11 on p. 859. We mention the following:

- ① $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c};$
- ② $\mathbf{a} \times (c\mathbf{b}) = c(\mathbf{a} \times \mathbf{b});$
- ③ $\mathbf{b} \times \mathbf{a} = -(\mathbf{a} \times \mathbf{b});$
- ④ $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} \neq \mathbf{a} \times (\mathbf{b} \times \mathbf{c}), \text{ except for special cases.}$

The first three properties imply that cross products of two vectors can be computed using the familiar rules for computing with real numbers (or dot products of vectors), except that the commutative law $ab = ba$ (resp., $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$) is replaced by the *anti-commutative law* (3). For cross products of three or more vectors, the analogy does no longer hold. A weak substitute for the associative law is *Jacobi's identity*

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} + (\mathbf{b} \times \mathbf{c}) \times \mathbf{a} + (\mathbf{c} \times \mathbf{a}) \times \mathbf{b} = \mathbf{0}; \text{ cf. [Ste21], Ex. 51, p. 863.}$$

Math 241
Calculus III

Thomas
Honold

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Introduction

2 Limits and Continuity

Limits of Sequences of Points

Limits and Continuity for Curves

Topology of Subsets of \mathbb{R}^n

3 Differentiation and Integration

4 Arc Length

Remarks on Uniform Continuity

Reparametrization

Smooth Parametric Curves

5 Curvature and Related Concepts

Parametrization with Respect to Arc Length

Curvature

Osculating Planes and Circles

Examples

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Today's Lecture: Vector Functions

Introduction

This and the next lectures discuss, supplement (and sometimes extend) material from [Ste21], Chapter 13.

Definition

A *vector function* or *parametric curve* is a map

$$f: I \rightarrow \mathbb{R}^n$$

with domain $I \subseteq \mathbb{R}$ and codomain \mathbb{R}^n for some fixed integer $n \geq 0$.

Notes

- You should already be familiar with the case $n = 2$ (*plane curves*) from Calculus II ([Ste21], Ch. 10). The current textbook chapter restricts attention to the case $n = 3$ (*space curves*). We will use the general setting $n \in \{1, 2, 3, 4, \dots\}$.
- Usually we require the domain I to be an interval, e.g., $[a, b]$, (a, b) , $[a, +\infty)$, \mathbb{R} , and f to be continuous. (In the literature these requirements are often part of the definition of a parametric curve.)
- Often the range $f(I) = \{f(t); t \in I\} \subseteq \mathbb{R}^n$ is referred to as a “curve” and f as a particular parametrization of this curve.

Notes con't

- In Physics parametric space curves $f: I \rightarrow \mathbb{R}^3$ are used to describe the motion of an object as a function of time. For this reason it is custom to denote the variable by “ t ”.
- A function $f: I \rightarrow \mathbb{R}^n$, $t \mapsto f(t)$ determines (and is determined by) n real-valued (“ordinary”) functions $f_i: I \rightarrow \mathbb{R}$ via

$$f(t) = \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{pmatrix} = f_1(t)\mathbf{e}_1 + f_2(t)\mathbf{e}_2 + \cdots + f_n(t)\mathbf{e}_n.$$

These are called *coordinate functions* of f .

In the case $n = 3$ we also write

$$\mathbf{r}(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} = x(t)\mathbf{e}_1 + y(t)\mathbf{e}_2 + z(t)\mathbf{e}_3 = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k},$$

to emphasize the physical meaning of a curve (position of the object at time t relative to the origin of the coordinate system).

Notes cont'd

- Of course we may also use row vector notation, e.g., $\mathbf{r}(t) = (x(t), y(t), z(t))$, whenever convenient.
- Since [Ste21] distinguishes 3-dimensional vectors $\mathbf{v} = \langle v_1, v_2, v_3 \rangle$ from points $P(v_1, v_2, v_3)$, it has to distinguish between vector functions $\mathbf{v}(t) = \langle v_1(t), v_2(t), v_3(t) \rangle$ and space curves $P(t) = P(v_1(t), v_2(t), v_3(t))$ as well. This complication doesn't arise in our view.
- However, choosing different coordinate systems/time scales for a real-world curve corresponds to a coordinate change/reparametrization of the corresponding mathematical curves, and hence a single real-world curve is represented by a whole equivalence class $[\gamma_0]$ of mathematical curves, whose members are of the form

$$\gamma(t) = \mathbf{A}\gamma_0(ct + d) + \mathbf{b}$$

with $\gamma_0: I \rightarrow \mathbb{R}^3$ a fixed representative curve, $\mathbf{x}' = \mathbf{Ax} + \mathbf{b}$ ($\mathbf{A} \in \mathbb{R}^{3 \times 3}$, $\mathbf{b} \in \mathbb{R}^3$) describing the coordinate change and $t' = ct + d$ ($c, d \in \mathbb{R}$) the reparametrization.

Notes cont'd

- For any unexplained terms (such as “map”, “domain”, “codomain”, “range”) you are referred to my Discrete Mathematics slides `math213_full_handout.pdf` or the textbook used for Math213:

[Ro13] Kenneth H. Rosen, Kamala Krithivasan, *Discrete Mathematics and Its Applications*, 7th adapted global edition, McGraw-Hill 2013

Chapters 1 and 2 of this book constitute an excellent introduction to the foundations of Mathematics. The terms mentioned above are explained in Section 2.3, “Functions”, which also contains an explanation of the immensely important concepts of injective/surjective/bijective maps. Some other chapters contain important foundational material as well, e.g., Chapter 5, “Induction and Recursion”, and Chapter 9, “Relations” (in particular Section 9.5., “Equivalence Relations”).

ME and CEE students are advised to study this material as well.

Examples

1 $f(t) = (t + 2, -t, 2t + 3)$, $t \in \mathbb{R}$

This can be written as

$$f(t) = \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix} + t \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix},$$

and shows that f parametrizes the line in \mathbb{R}^3 through the point $(2, 0, 3)$ with direction vector $(1, -1, 2)$.

2 $f(t) = (2 \cos t, 3, 2 \sin t)$ for $t \in [0, 2\pi]$

This can be written as

$$f(t) = (2 \cos t)\mathbf{e}_1 + (2 \sin t)\mathbf{e}_3 + (0, 3, 0),$$

and shows that f parametrizes a circle of radius 2 in the plane with equation $y = 3$ (a plane parallel to the (x, z) -plane) centered at $(0, 3, 0)$.

Examples (cont'd)

- ③ $f(t) = (\cos t, \sin t, t)$ for $t \in [0, +\infty)$

This is called a *helix* (“corkscrew curve”).

The projection of f to the (x, y) -plane is the unit circle (traversed infinitely many times). Since

$$\begin{aligned} f(t + 2\pi) &= (\cos(t + 2\pi), \sin(t + 2\pi), t + 2\pi) \\ &= f(t) + (0, 0, 2\pi), \end{aligned}$$

the helix is obtained by “glueing together” infinitely many pieces, all of which are vertical translates of a “unit piece” H_0 :

$$H_0 = f([0, 2\pi]),$$

$$H_1 = f([2\pi, 4\pi]) = H_0 + (0, 0, 2\pi),$$

$$H_2 = f([4\pi, 6\pi]) = H_0 + (0, 0, 4\pi), \quad \text{etc.}$$

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length
Curvature

Osculating Planes
and Circles

Examples

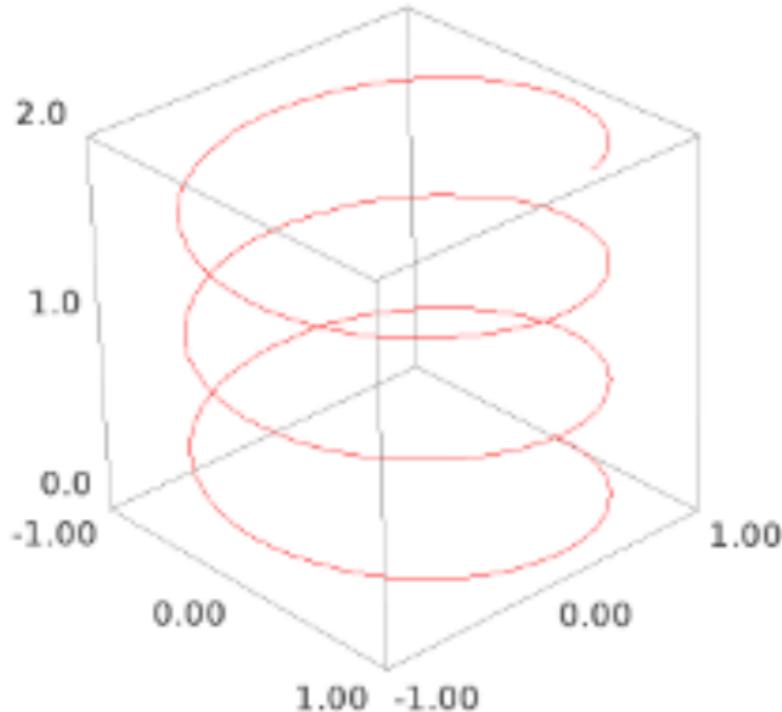
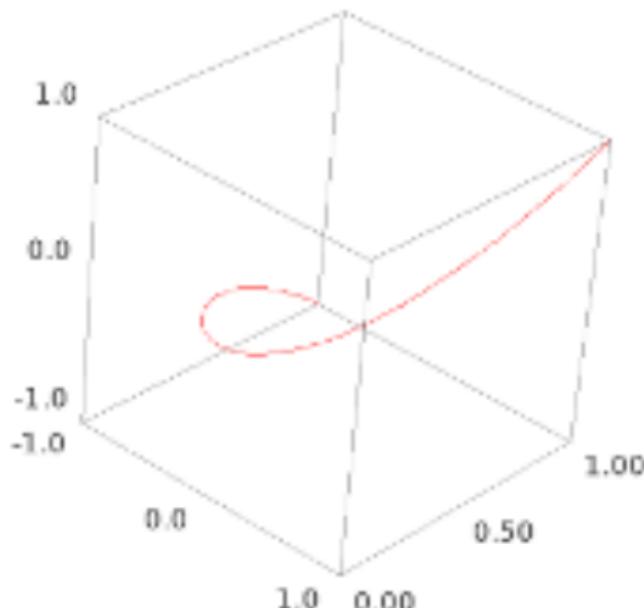


Figure: Helix with equation $f(t) = (\sin t, \cos t, t/10)$, $t \in [0, 20]$

Examples (cont'd)

4) $f(t) = (t, t^2, \dots, t^n), t \in \mathbb{R}$

This curve, which has codomain \mathbb{R}^n , is called a *normal rational curve*. In the special case $n = 3$, in which the defining equation is $f(t) = (t, t^2, t^3)$, it is called a *twisted cubic*.



Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

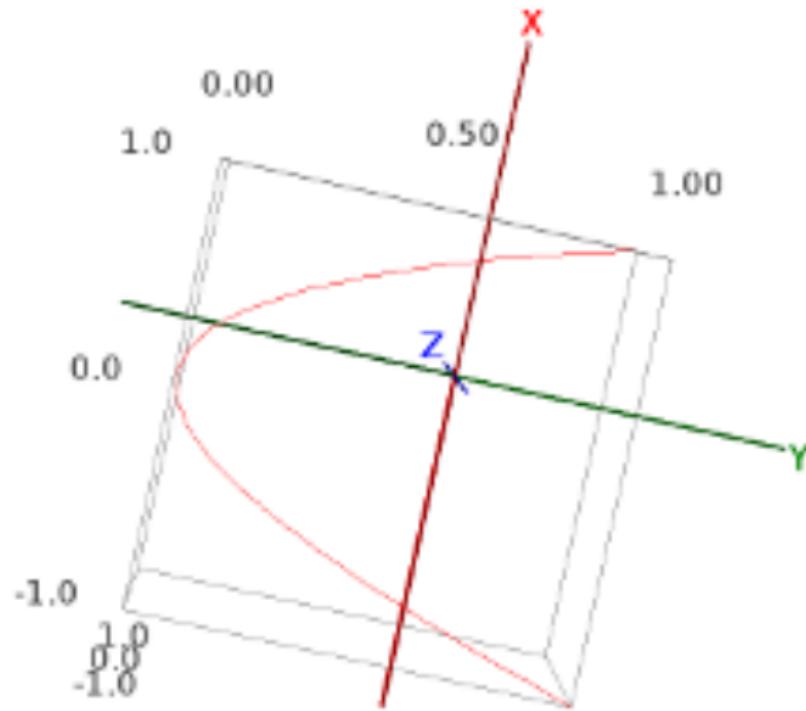


Figure: Twisted cubic projected to (x, y) -plane

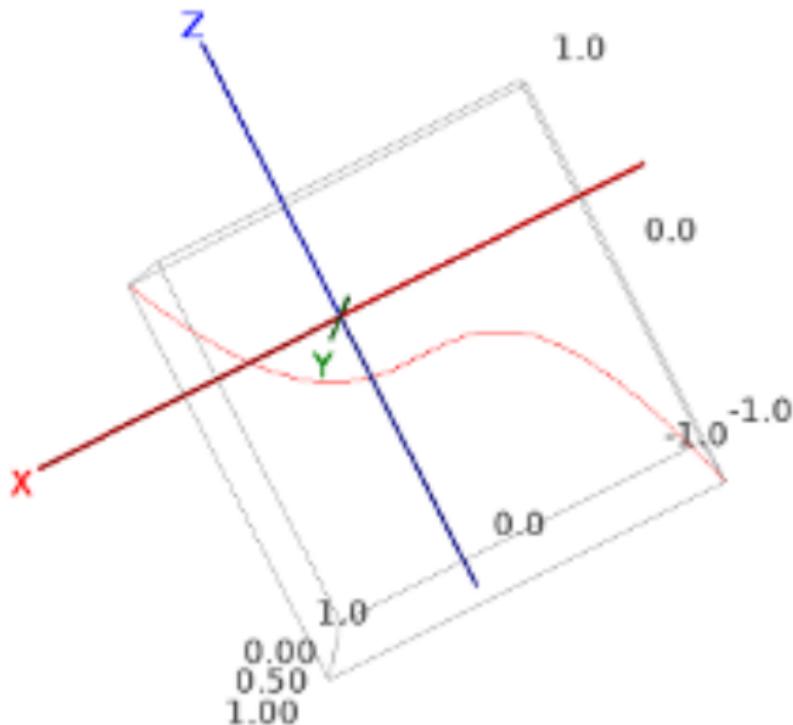


Figure: Twisted cubic projected to (x, z) -plane

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

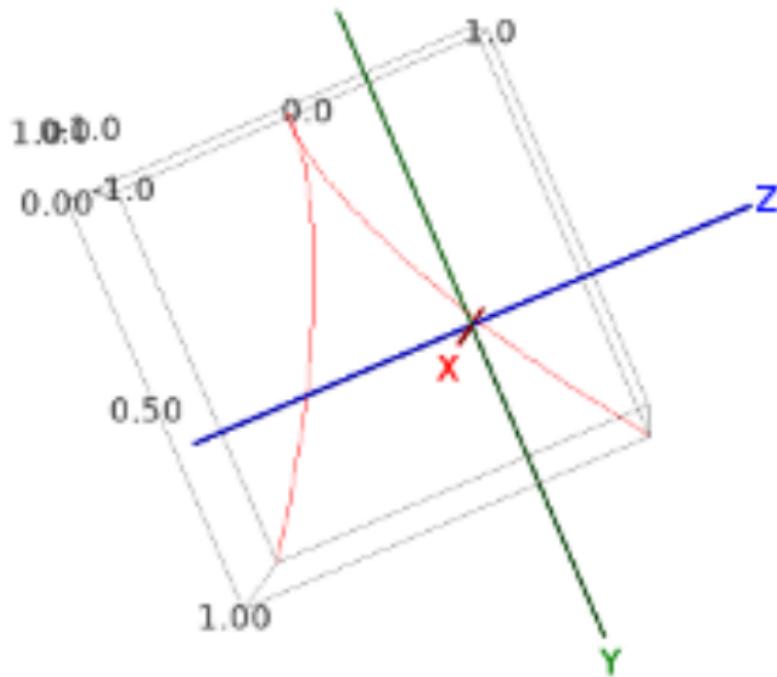
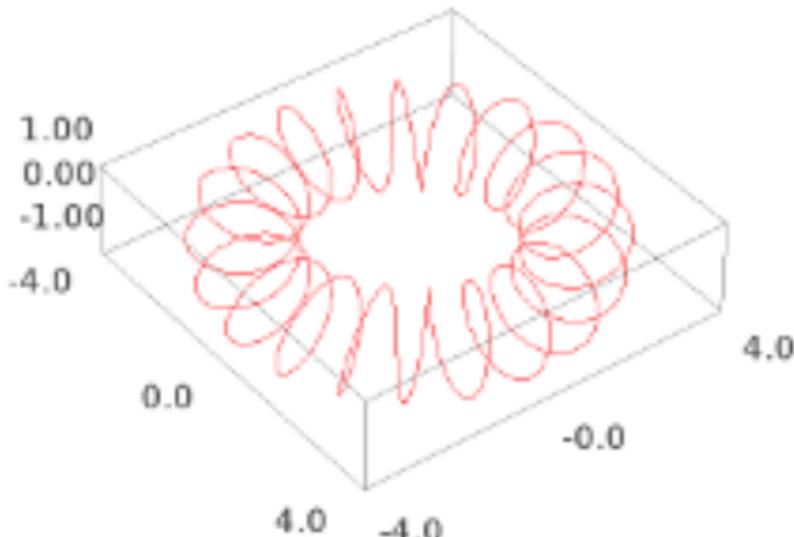


Figure: Twisted cubic projected to (y, z) -plane

Examples (cont'd)

- 5 $f(t) = ((3 + \cos(20t)) \cos t, (3 + \cos(20t)) \sin t, \sin(20t)),$
 $t \in [0, 2\pi]$

This space curve is called a *toroidal spiral*.



Example (cont'd)

Rewriting $f(t)$ as

$$f(t) = \begin{pmatrix} 3 \cos t \\ 3 \sin t \\ 0 \end{pmatrix} + \cos(20t) \begin{pmatrix} \cos t \\ \sin t \\ 0 \end{pmatrix} + \sin(20t) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

reveals why this describes a toroidal spiral:

f(t) is the superposition of a circle of radius 3 in the (x, y)-plane and a circle of smaller radius in the plane spanned by the z-axis and the radius vector of the large circle, which is a scalar multiple of $(\cos t, \sin t)$.

Period length of the large circle: 2π

Period length of the small circle: $2\pi/20$

The Range

Definition

Let $f: I \rightarrow \mathbb{R}^n$ be a parametric curve. The *range* of f is $f(I) = \{f(t); t \in I\}$

The range of f is a point set in \mathbb{R}^n . It is sometimes called a *non-parametric curve*, or just a *curve*.

Example

The range of the parametric twisted cubic $f: \mathbb{R} \rightarrow \mathbb{R}^3$, $t \mapsto (t, t^2, t^3)$ is

$$T = \{(x, y, z) \in \mathbb{R}^3; y = x^2 \wedge z = x^3\}.$$

T is referred to as *twisted cubic* as well. It can also be obtained as the intersection of the two surfaces

$$P = \{(x, y, z) \in \mathbb{R}^3; y = x^2\},$$

$$Q = \{(x, y, z) \in \mathbb{R}^3; z = x^3\}.$$

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

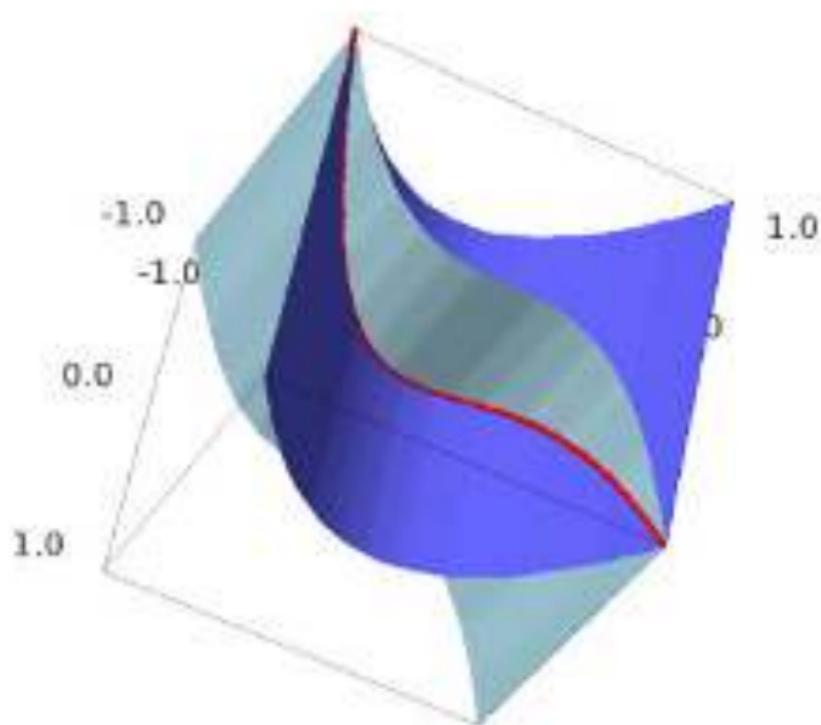


Figure: Twisted cubic as intersection of the surfaces P and Q

Conversely, we can turn a non-parametric curve into a parametric curve:

Example

Find a parametrization of the curve

$$C = \{(x, y, z) \in \mathbb{R}^3; x^2 + y^2 = 1 \wedge x + y + z = 1\},$$

which is the intersection of a cylinder and a plane.

Since $(x, y, z) \in C$ iff (x, y) is on the unit circle of \mathbb{R}^2 and $z = 1 - x - y$, a solution is

$$f(t) = \begin{pmatrix} \cos t \\ \sin t \\ 1 - \cos t - \sin t \end{pmatrix}, \quad t \in [0, 2\pi].$$

Question

Which type of geometric object does C represent?

Answer

We have

$$f(t) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + \cos t \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + \sin t \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

but we cannot conclude from this that C is a circle in the plane $E = (0, 0, 1) + \mathbb{R}(1, 0, -1) + \mathbb{R}(0, 1, -1)$ (which is just the parametric representation of $x + y + z = 1$). The problem is that the vectors $\mathbf{v}_1 = (1, 0, -1)$ and $\mathbf{v}_2 = (0, 1, -1)$ are not perpendicular.

Therefore, we let

$$\mathbf{u}'_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1 \\ -1/2 \end{pmatrix},$$

which is orthogonal to \mathbf{v}_1 ,

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \mathbf{u}_2 = \frac{\mathbf{u}'_2}{|\mathbf{u}'_2|} = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

Answer (con't)

which form a pair of orthogonal unit-length vectors, and express $f(t)$ in terms of \mathbf{u}_1 , \mathbf{u}_2 as follows:

$$\begin{aligned} f(t) &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + (\cos t)\mathbf{v}_1 + \sin t \left(\frac{1}{2}\mathbf{v}_1 + \mathbf{u}'_2 \right) \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + \left(\cos t + \frac{1}{2} \sin t \right) \mathbf{v}_1 + (\sin t) \mathbf{u}'_2 \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + \left(\sqrt{2} \cos t + \frac{\sin t}{\sqrt{2}} \right) \mathbf{u}_1 + \left(\frac{\sqrt{3} \sin t}{\sqrt{2}} \right) \mathbf{u}_2 \end{aligned}$$

With some effort it can be shown that

$$h(t) = \left(\sqrt{2} \cos t + \frac{\sin t}{\sqrt{2}}, \frac{\sqrt{3} \sin t}{\sqrt{2}} \right), \quad t \in [0, 2\pi]$$

parametrizes an ellipse with principal axes of lengths $a = \sqrt{3}$, $b = 1$, which arises from the standard ellipse with equation $x^2/3 + y^2 = 1$ by a 30° -degree rotation.

Exercise

With $h(t) = (h_1(t), h_2(t))^T$ as defined on the last slide, show that

$$e(t) = R(\pi/6)^{-1}h(t + \pi/4) = \begin{pmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{pmatrix}^{-1} \begin{pmatrix} h_1(t + \pi/4) \\ h_2(t + \pi/4) \end{pmatrix}$$

parametrizes the standard ellipse in the usual way.

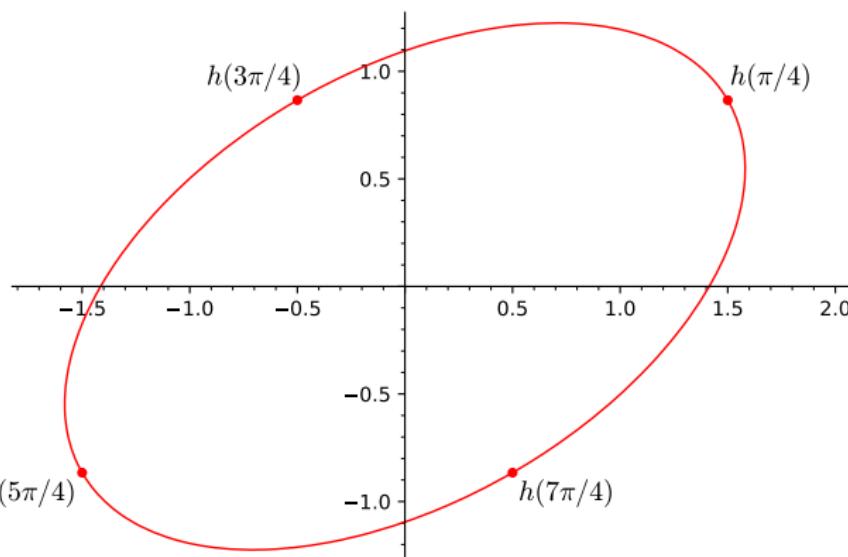


Figure: $h(t) = \left(\sqrt{2} \cos t + \sin t / \sqrt{2}, \sqrt{3} \sin t / \sqrt{2} \right)$, $t \in [0, 2\pi]$

Further Notes on the Range

- A given (non-parametric) curve admits many different parametrizations. For example, here are 7 different parametrizations of the unit circle in \mathbb{R}^2 :

$$f_1(t) = (\cos t, \sin t) \quad \text{for } t \in [0, 2\pi),$$

$$f_2(t) = (\cos t, \sin t) \quad \text{for } t \in [0, 2\pi],$$

$$f_3(t) = (\cos t, \sin t) \quad \text{for } t \in [-\pi, \pi],$$

$$f_4(t) = (\cos t, \sin t) \quad \text{for } t \in \mathbb{R},$$

$$f_5(t) = (\cos t, -\sin t) \quad \text{for } t \in [0, 2\pi],$$

$$f_6(t) = (-\sin t, \cos t) \quad \text{for } t \in [0, 2\pi],$$

$$f_7(t) = (\cos(t^2), \sin(t^2)) \quad \text{for } t \in [0, 2.6].$$

- A parametric curve $f: [a, b] \rightarrow \mathbb{R}^n$ is said to be *closed* if $f(a) = f(b)$, i.e., if the starting point and end point coincide. This property cannot be seen from the range $f([a, b])$ alone. In the previous note, the curves f_1, f_4, f_7 are not closed; the remaining curves are closed.

Limits and Continuity

These are defined as in Calculus I, except that the absolute value on \mathbb{R} is replaced by the Euclidean length $|\mathbf{x}| = \sqrt{x_1^2 + \cdots + x_n^2}$ on \mathbb{R}^n for $n > 1$. Recall that the (Euclidean) distance of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is $d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}|$.

Definition

Suppose $\mathbf{a} \in \mathbb{R}^n$ and $r \in \mathbb{R}^+ = \{x \in \mathbb{R}; x > 0\}$.

1 The subset

$$B_r(\mathbf{a}) = \{\mathbf{x} \in \mathbb{R}^n; d(\mathbf{x}, \mathbf{a}) < r\} \subset \mathbb{R}^n$$

is called (*Euclidean*) open ball with center \mathbf{a} and radius r .
Closed balls are defined in the same way by the condition
 $d(\mathbf{x}, \mathbf{a}) \leq r$ and denoted by $\overline{B_r(\mathbf{a})}$.

2 The subset

$$S_r(\mathbf{a}) = \{\mathbf{x} \in \mathbb{R}^n; d(\mathbf{x}, \mathbf{a}) = r\} \subset \mathbb{R}^n$$

is called (*Euclidean*) sphere with center \mathbf{a} and radius r .

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Notes

- Open balls are *open* and closed balls are *closed* in the sense of topology: For every $\mathbf{y} \in B_r(\mathbf{a})$ there exists $\epsilon > 0$ such that $B_\epsilon(\mathbf{y}) \subseteq B_r(\mathbf{a})$ (e.g., one can take $\epsilon = r - d(\mathbf{y}, \mathbf{a})$), and for every convergent sequence of points in $\overline{B_r(\mathbf{a})}$ the limit point is contained in $\overline{B_r(\mathbf{a})}$ as well.
- The sphere $S_r(\mathbf{a}) = \overline{B_r(\mathbf{a})} \setminus B_r(\mathbf{a})$ is equal to the *boundary* $\partial B_r(\mathbf{a})$ or $\overline{\partial B_r(\mathbf{a})}$ in the sense of Topology: Every neighborhood of a point on the sphere contains points of the corresponding ball as well as points outside the ball. This implies that the topological closure of $B_r(\mathbf{a})$ is $\overline{B_r(\mathbf{a})}$, justifying the “overbar” used to denote closed spheres.

Definition (Limit of a sequence of points)

Suppose $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ is an infinite sequence of points (vectors) in \mathbb{R}^n . The sequence $(\mathbf{x}^{(k)})$ is said to have the *limit* $\mathbf{a} \in \mathbb{R}^n$, notation $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{a}$, if for every $\epsilon > 0$ ("challenge") there exists an $N \in \mathbb{N}$ ("response") such that

$$\mathbf{x}^{(k)} \in B_\epsilon(\mathbf{a}) \quad \text{for all } k > N.$$

Notes

- In other words, we have $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{a}$ iff every ball around \mathbf{a} of positive radius contains all but finitely many terms of the sequence.
- $\mathbf{x}^{(k)} \in B_\epsilon(\mathbf{a})$ is equivalent to $d(\mathbf{x}^{(k)}, \mathbf{a}) = |\mathbf{x}^{(k)} - \mathbf{a}| < \epsilon$.
- As in the one-dimensional case, a sequence in \mathbb{R}^n need not converge, but if it converges then its limit is uniquely determined.
- Upper indexing for the terms of a sequence in \mathbb{R}^n avoids a conflict with the usual indexing of coordinates, i.e., we can write $\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$. The sequences $x_i^{(0)}, x_i^{(1)}, x_i^{(2)}, \dots$ (i.e., the coordinate i is fixed) are called *coordinate sequences* associated with $(\mathbf{x}^{(k)})$.

Example (rather stupid)

The sequence $\mathbf{x}^{(k)} = (\cos k, \sin k)$, $k = 0, 1, 2, \dots$, in the plane \mathbb{R}^2 doesn't converge (rather it moves, counterclock-wise, infinitely often around the unit circle with adjacent sequence terms being at distance 1, measured along the circle, apart).

On the other hand, the sequence $\mathbf{y}^{(k)} = (\cos \frac{1}{k}, \sin \frac{1}{k})$, $k = 1, 2, 3, \dots$, converges to the point $(1, 0)$. This follows from $\lim_{k \rightarrow \infty} \cos \frac{1}{k} = \cos 0 = 1$, $\lim_{k \rightarrow \infty} \sin \frac{1}{k} = \sin 0 = 0$; cf. the subsequent theorem and its remark.

Example (rather advanced)

For every complex number z of absolute value $|z| < 1$ we have

$$\sum_{k=0}^{\infty} z^k = 1 + z + z^2 + z^3 + \dots = \frac{1}{1-z}.$$

For this you need to know that complex numbers $z = (x, y)$ are just points in the plane, are added like vectors and multiplied according to $z_1 z_2 = (x_1, y_1)(x_2, y_2) := (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2)$, the complex number $w = \frac{1}{1-z}$ is the solution of $w(1-z) = 1$, and finally that series in \mathbb{R}^n are defined in terms of their sequence of partial sums just like in the one-dimensional case.

Definition (limits and continuity for curves)

Let $f: I \rightarrow \mathbb{R}^n$ be a parametric curve in \mathbb{R}^n , whose domain I is an interval, and $t_0 \in \bar{I} = I \cup \partial I$ (the closure of I , which may include $+\infty$ or $-\infty$ if I is unbounded).

- 1 f is said to have the *limit* $\mathbf{a} \in \mathbb{R}^n$ for $t \rightarrow t_0$, notation
 $\lim_{t \rightarrow t_0} f(t) = \mathbf{a}$, if for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$f(t) \in B_\epsilon(\mathbf{a}) \quad \text{whenever} \quad t \in I \wedge 0 < |t - t_0| < \delta.$$

- 2 f is said to be *continuous* in t_0 if $t_0 \in I$ and (1) holds with
 $\mathbf{a} = f(t_0) : \lim_{t \rightarrow t_0} f(t) = f(t_0)$.

Notes

- The condition, e.g. in (2), $f(t) \in B_\epsilon(f(t_0))$ will often be written as $|f(t) - f(t_0)| < \epsilon$; note however that here $f(t) - f(t_0)$ is an n -dimensional vector and $|f(t) - f(t_0)|$ refers to its length, in a way obscuring the geometry behind the definition.
- The closure of (a, b) , $(a, b]$, $[a, b)$ and $[a, b]$ itself is $[a, b]$.

Theorem

Suppose $f: I \rightarrow \mathbb{R}^n$ has coordinate functions f_1, \dots, f_n and $t_0 \in \bar{I}$.

- 1 The limit $\lim_{t \rightarrow t_0} f(t)$ exists if and only if all limits $\lim_{t \rightarrow t_0} f_i(t)$, $1 \leq i \leq n$, exist. If this is the case then

$$\lim_{t \rightarrow t_0} f(t) = \left(\lim_{t \rightarrow t_0} f_1(t), \dots, \lim_{t \rightarrow t_0} f_n(t) \right).$$

- 2 f is continuous at $t_0 \in I$ if and only if all coordinate functions f_i , $1 \leq i \leq n$, are continuous at t_0 .

Remark

Part (1) is also true for sequences: The limit $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)}$ of a sequence in \mathbb{R}^n exists iff the limits $\lim_{k \rightarrow \infty} x_i^{(k)}$, $1 \leq i \leq n$, of its n coordinate sequences exist. If this is the case then

$$\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \left(\lim_{k \rightarrow \infty} x_1^{(k)}, \dots, \lim_{k \rightarrow \infty} x_n^{(k)} \right).$$

The proof can be easily inferred from that of the theorem.

Example

Consider the two space curves $f(t) = (t, t^2, t^3)$ and $g(t) = e^{-t}(t, t^2, t^3) = (te^{-t}, t^2e^{-t}, t^3e^{-t})$, both with domain $I = \mathbb{R}$.

Both curves are continuous everywhere, since their coordinate functions are (known from Calculus I), and hence limits at $t_0 \in \mathbb{R}$ are obtained by evaluating the curves at t_0 ; for example,

$$\lim_{t \rightarrow 2} g(t) = g(2) = (2/e^2, 4/e^2, 8/e^2) = \frac{1}{e^2}(2, 4, 8).$$

$\lim_{t \rightarrow +\infty} f(t)$ does not exist, since $\lim_{t \rightarrow +\infty} t$ does not exist (and likewise $\lim_{t \rightarrow +\infty} t^2$, $\lim_{t \rightarrow +\infty} t^3$),
but $\lim_{t \rightarrow +\infty} g(t)$ exists:

$$\lim_{t \rightarrow +\infty} g(t) = \left(\lim_{t \rightarrow +\infty} te^{-t}, \lim_{t \rightarrow +\infty} t^2e^{-t}, \lim_{t \rightarrow +\infty} t^3e^{-t} \right) = (0, 0, 0),$$

the origin of \mathbb{R}^3 .

Proof of the theorem.

(2) follows from (1).

For the proof of (1) we use

$$\begin{aligned}|f(t) - \mathbf{a}| &= \sqrt{(f_1(t) - a_1)^2 + \cdots + (f_n(t) - a_n)^2} \\ &\leq \sqrt{n} \max_{1 \leq i \leq n} |f_i(t) - a_i|\end{aligned}$$

and the obvious $|f(t) - \mathbf{a}| \geq |f_i(t) - a_i|$. This shows:

If $\lim_{t \rightarrow t_0} f(t) = \mathbf{a}$, we can take the same $\delta = \delta(\epsilon)$ as a witness for $\lim_{t \rightarrow t_0} f_i(t) = a_i$.

Conversely, if $\lim_{t \rightarrow t_0} f_i(t) = a_i$ for all i and $\delta_i = \delta_i(\epsilon)$ are such that $|f_i(t) - a_i| < \epsilon$ whenever $|t - t_0| < \delta_i$, then

$$|f(t) - \mathbf{a}| < \sqrt{n} \cdot \epsilon \quad \text{if} \quad |t - t_0| < \min_{1 \leq i \leq n} \delta_i.$$

\implies We can take $\delta = \min_{1 \leq i \leq n} \delta_i (\epsilon / \sqrt{n})$ as a witness for “there exists $\delta > 0$ such that $|f(t) - \mathbf{a}| < \epsilon$ whenever $|t - t_0| < \delta$ ”.
Universal generalization over ϵ then proves $\lim_{t \rightarrow t_0} f(t) = \mathbf{a}$. □

Generalization

Providing half of the definition of limits and continuity for real-world curves

If $f: I \rightarrow \mathbb{R}^n$ is a curve, $\mathbf{v}_1, \dots, \mathbf{v}_n$ is a basis of \mathbb{R}^n ("new basis") and $\mathbf{b} \in \mathbb{R}^n$ ("new origin"), we have

$$f(t) = g_1(t)\mathbf{v}_1 + \cdots + g_n(t)\mathbf{v}_n + \mathbf{b}$$

for uniquely determined coordinate functions $g_i: I \rightarrow \mathbb{R}$.

(So far we have considered only the special case $\mathbf{v}_i = \mathbf{e}_i$, $\mathbf{b} = \mathbf{0}$, with notation f_i for the corresponding standard coordinate functions.)

Corollary

Suppose a curve $f: I \rightarrow \mathbb{R}^n$ is represented as above with respect to some basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of \mathbb{R}^n and $\mathbf{b} \in \mathbb{R}^n$, and let $t_0 \in \bar{I}$.

- 1 $\lim_{t \rightarrow t_0} f(t)$ exists if and only if all limits $\lim_{t \rightarrow t_0} g_i(t)$, $1 \leq i \leq n$, exist. If this is the case then

$$\lim_{t \rightarrow t_0} f(t) = \left(\lim_{t \rightarrow t_0} g_1(t) \right) \mathbf{v}_1 + \cdots + \left(\lim_{t \rightarrow t_0} g_n(t) \right) \mathbf{v}_n + \mathbf{b}.$$

- 2 f is continuous at $t_0 \in I$ if and only if all coordinate functions g_i , $1 \leq i \leq n$, are continuous at t_0 .

Proof for $\mathbf{b} = \mathbf{0}$.

Let $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n \times n}$ be the matrix with columns $\mathbf{v}_1, \dots, \mathbf{v}_n$. Since the \mathbf{v}_j form a basis of \mathbb{R}^n , the inverse matrix $\mathbf{B} = (b_{ij})$ of \mathbf{A} exists.

$$\Rightarrow f(t) = \mathbf{A} \begin{pmatrix} g_1(t) \\ \vdots \\ g_n(t) \end{pmatrix} = \begin{pmatrix} a_{11}g_1(t) + \cdots + a_{1n}g_n(t) \\ \vdots \\ a_{n1}g_1(t) + \cdots + a_{nn}g_n(t) \end{pmatrix}$$

From this we see that the standard coordinate functions f_i of f are linear combinations of g_i , viz. $f_i(t) = \sum_{j=1}^n a_{ij}g_j(t)$ for $1 \leq i \leq n$.

Similarly, we get $g_i(t) = \sum_{j=1}^n b_{ij}f_j(t)$ for $1 \leq i \leq n$.

\Rightarrow Using the laws for computing with limits of scalar functions (\rightarrow Calculus I), we obtain that the existence of all limits

$\lim_{t \rightarrow t_0} g_i(t)$ is equivalent to the existence of all limits $\lim_{t \rightarrow t_0} f_i(t)$, and

$$\begin{aligned} \lim_{t \rightarrow t_0} f(t) &= \begin{pmatrix} a_{11} \lim_{t \rightarrow t_0} g_1(t) + \cdots + a_{1n} \lim_{t \rightarrow t_0} g_n(t) \\ \vdots \\ a_{n1} \lim_{t \rightarrow t_0} g_1(t) + \cdots + a_{nn} \lim_{t \rightarrow t_0} g_n(t) \end{pmatrix} = \mathbf{A} \begin{pmatrix} \lim_{t \rightarrow t_0} g_1(t) \\ \vdots \\ \lim_{t \rightarrow t_0} g_n(t) \end{pmatrix} \\ &= \left(\lim_{t \rightarrow t_0} g_1(t) \right) \mathbf{v}_1 + \cdots + \left(\lim_{t \rightarrow t_0} g_n(t) \right) \mathbf{v}_n. \end{aligned}$$

□

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Example (loosely related)

Consider the maps

$$R: [0, 2\pi] \rightarrow \mathbb{R}^{2 \times 2}, \quad \phi \mapsto R(\phi) = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$$

and $S: [0, 2\pi] \rightarrow \mathbb{R}^{2 \times 2}, \phi \mapsto S(\phi)$; cf. Worksheet 3, Exercise W11.

Since $\mathbb{R}^{2 \times 2} \triangleq \mathbb{R}^4$, we can view R and S as (closed) curves in \mathbb{R}^4 . As such they are continuous, since their standard coordinate functions (e.g., $r_{11}(\phi) = \cos \phi$, $r_{12}(\phi) = -\sin \phi$, $r_{21}(\phi) = \sin \phi$, $r_{22}(\phi) = \cos \phi$) are continuous.

This says that a small change in the angle of a rotation/reflection effects only a small change in the rotated/reflected image of any given vector (e.g., look at the columns of $R(\phi)$, $S(\phi)$, which are the images of the standard basis of \mathbb{R}^2).

Topology of Subsets of \mathbb{R}^n

A glossary for later reference

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Suppose $\mathbf{a} \in \mathbb{R}^n$ and $D \subseteq \mathbb{R}^n$.

- 1 \mathbf{a} is called an *interior point* of D , and D a *neighborhood* of \mathbf{a} , if there exists $\epsilon > 0$ such that $B_\epsilon(\mathbf{a}) \subseteq D$. The set of interior points of D is called the *interior* of D and denoted by D° .
- 2 \mathbf{a} is called a *boundary point* of D if every ball $B_\epsilon(\mathbf{a})$ of positive radius contains a point in D and a point in the complementary set $\mathbb{R}^n \setminus D$. The set of boundary points of D is called the *boundary* of D and denoted by ∂D .
- 3 \mathbf{a} is called a *limit point* of D if there exists a sequence $(\mathbf{x}^{(k)})$ with terms $\mathbf{x}^{(k)} \in D$ that converges to \mathbf{a} . The set of limit points of D is called the *closure* of D and denoted by \overline{D} .
- 4 \mathbf{a} is called an *accumulation point* of D if every ball $B_\epsilon(\mathbf{a})$ of positive radius contains a point of D that is different from \mathbf{a} . The set of accumulation points of D is denoted by D' .

Math 241

Calculus III

Thomas
Honold

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

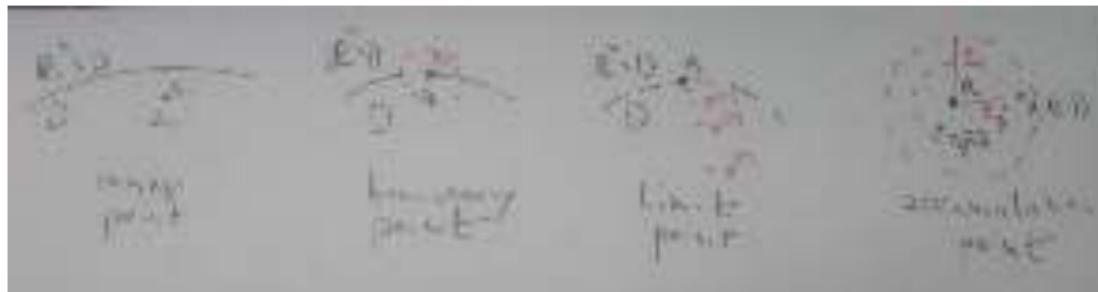
Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples



Topology of Subsets of \mathbb{R}^n Cont'd

Note that $D^\circ \subseteq D \subseteq \overline{D}$, since $\mathbf{a} \in B_\epsilon(\mathbf{a})$ (showing $D^\circ \subseteq D$) and a constant sequence $\mathbf{a}, \mathbf{a}, \mathbf{a}, \dots$ has limit \mathbf{a} (showing $D \subseteq \overline{D}$).

- 5 D is said to be *open* if $D^\circ = D$, i.e., D is a neighborhood of any of its points or, equivalently, D contains a ball of positive radius around any of its points.
- 6 D is said to be *closed* if $D = \overline{D}$, i.e., D contains all limits of sequences in D .

Notes

- D is open iff $D \cap \partial D = \emptyset$.
- D is closed iff $D \supseteq \partial D$.
- D is open iff its complementary set $\mathbb{R}^n \setminus D$ is closed (and vice versa, of course).
- The boundary of D and its complementary set are the same, $\partial D = \partial(\mathbb{R}^n \setminus D)$, and we have the disjoint decompositions $\overline{D} = D^\circ \cup \partial D$, $\mathbb{R}^n = D^\circ \cup \partial D \cup (\mathbb{R}^n \setminus D)^\circ$.

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Notes cont'd

- $\overline{D} = D \cup \partial D = D \cup D'$.
- There are two kinds of points in D :
 - (i) \mathbf{a} is an *isolated point* if there exists a ball $B_\epsilon(\mathbf{a})$ of positive radius containing no point of D besides \mathbf{a} .
 - (ii) \mathbf{a} is an accumulation point if every ball $B_\epsilon(\mathbf{a})$ of positive radius contains a point $\mathbf{x} \in D$ besides \mathbf{a} . In this case $B_\epsilon(\mathbf{a})$ must contain infinitely many points of D , because otherwise there would be a point $\mathbf{x}^* \in D \setminus \{\mathbf{a}\}$ minimizing the distance to \mathbf{a} , and the ball $B_{\epsilon^*}(\mathbf{a})$, $\epsilon^* = |\mathbf{x}^* - \mathbf{a}|$, wouldn't contain any point of D besides \mathbf{a} .

Examples (subsets of \mathbb{R}^2)

- 1 For the closed unit disk $D = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 1\}$ we have $\overline{D} = D$ (D is closed), $D^\circ = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 < 1\}$ (open unit disk), $\partial D = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 = 1\}$ (unit circle), and $D' = D$ (every point of D is an accumulation point).
- 2 For $D = \{(1/k, 0); k \in \mathbb{N}\} = \{(1, 0), (\frac{1}{2}, 0), (\frac{1}{3}, 0), \dots\}$ we have $\overline{D} = \partial D = D \cup \{(0, 0)\}$, $D^\circ = \emptyset$, $D' = \{(0, 0)\}$.
- 3 For $D = \mathbb{Q}^2$ (points in the plane with rational coordinates) we have $\overline{D} = \partial D = D' = \mathbb{R}^2$, $D^\circ = \emptyset$.

Differentiation of Curves

The definition of the derivative of a real-valued function generalizes almost verbatim to the case of parametric curves.

Definition

Let $f: I \rightarrow \mathbb{R}^n$ be a parametric curve.

- ① We say that f is *differentiable at $t_0 \in I$* if the limit

$$f'(t_0) = \lim_{h \rightarrow 0} \frac{1}{h} (f(t_0 + h) - f(t_0)) = \lim_{h \rightarrow 0} \begin{pmatrix} \frac{1}{h} (f_1(t_0 + h) - f_1(t_0)) \\ \vdots \\ \frac{1}{h} (f_n(t_0 + h) - f_n(t_0)) \end{pmatrix}$$

exists, and *differentiable (per se)* if this limit exists for all $t_0 \in I$.

- ② Let $D \subseteq I$ be the set of all t_0 at which f is differentiable according to (1). The parametric curve $D \rightarrow \mathbb{R}^n$, $t \mapsto f'(t)$ is called the *derivative* of f .

The statement in (2) requires our general definition of a curve. (D need not be an interval and $t \mapsto f'(t)$ need not be continuous.)

Remark

From the definition of $f'(t_0)$ and the theorem on the coordinate-wise evaluation of limits it follows that f is differentiable at t_0 iff all its coordinate functions f_1, \dots, f_n are, and if applicable that

$$f'(t_0) = \begin{pmatrix} f'_1(t_0) \\ \vdots \\ f'_n(t_0) \end{pmatrix}.$$

The same applies to the more general setting of coordinate functions with respect to an arbitrary basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of \mathbb{R}^n and an arbitrary “origin” $\mathbf{b} \in \mathbb{R}^n$ (the proof is the same as for limits):
 $\frac{d}{dt}(g_1(t)\mathbf{v}_1 + \cdots + g_n(t)\mathbf{v}_n + \mathbf{b}) = g'_1(t)\mathbf{v}_1 + \cdots + g'_n(t)\mathbf{v}_n.$

Examples

- 1** The twisted cubic $f(t) = (t, t^2, t^3)$, $t \in \mathbb{R}$, is differentiable with $f'(t) = (1, 2t, 3t^2)$.
- 2** The curves $\phi \mapsto R(\phi)$, $\phi \mapsto S(\phi)$ are differentiable with, e.g.,

$$R'(\phi) = \begin{pmatrix} -\sin \phi & -\cos \phi \\ \cos \phi & -\sin \phi \end{pmatrix} = R(\phi) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = R(\phi + \pi/2).$$

Tangents to a Curve

The derivative of f in t_0 (provided it exists) satisfies

$$\frac{1}{h} (f(t_0 + h) - f(t_0) - h f'(t_0)) \rightarrow \mathbf{0} \quad \text{for } h \rightarrow 0.$$

In the case $f'(t_0) \neq \mathbf{0}$ this says that the distance between the curve point $f(t_0 + h)$ and the point $f(t_0) + h f'(t_0)$ on the line $f(t_0) + \mathbb{R} f'(t_0)$ is asymptotically (for $h \rightarrow 0$) much smaller than h and hence much smaller than the vectors $f(t_0 + h) - f(t_0)$ and $h f'(t_0)$.

⇒ Geometrically speaking, the line $f(t_0) + \mathbb{R} f'(t_0)$ appears to touch the curve in $f(t_0)$.

Definition

If $f: I \rightarrow \mathbb{R}^n$ is differentiable at $t_0 \in I$ and $f'(t_0) \neq \mathbf{0}$, the line $f(t_0) + \mathbb{R} f'(t_0)$ (i.e., the line through $f(t_0)$ with direction vector $f'(t_0)$) is called the *tangent line* to f at t_0 .

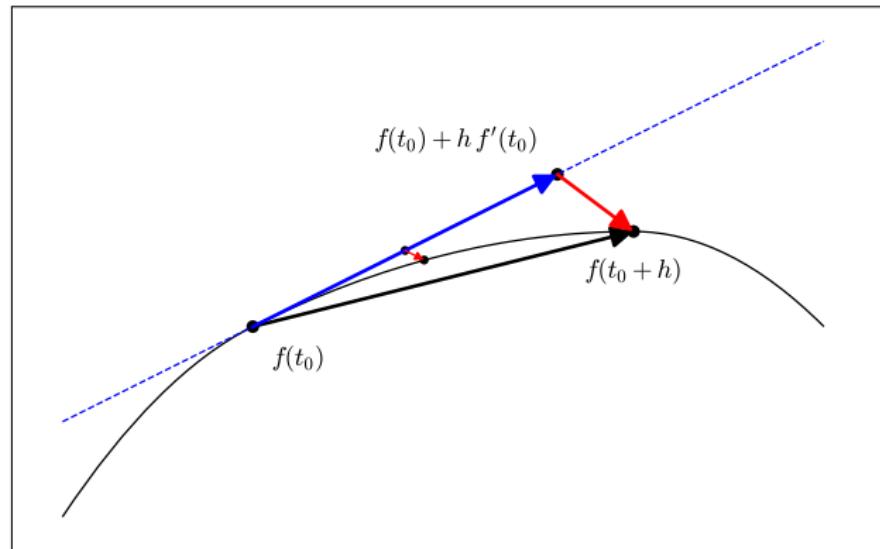


Figure: Illustration of the tangent line to f at t_0

The “error” of the *linear approximation* $f(t_0) + h f'(t_0)$ to f in t_0 , viz.

$$R(h) = f(t_0 + h) - f(t_0) - h f'(t_0),$$

satisfies $\lim_{h \rightarrow 0} \frac{|R(h)|}{|h|} = \lim_{h \rightarrow 0} \frac{|R(h)|}{|h f'(t_0)|} = \lim_{h \rightarrow 0} \frac{|R(h)|}{|f(t_0+h)-f(t_0)|} = 0$.

Afternote

The last sentence on the previous slide should be read as the statement “the three indicated limits are zero” (and hence a *a fortiori* equal). For the first limit this follows from the definition of the derivative, as discussed. The second limit, which differs from the first by a constant, must then be zero as well. Since

$$\frac{|R(h)|}{|f(t_0 + h) - f(t_0)|} = \frac{|R(h)|}{|h|} \frac{|h|}{|f(t_0 + h) - f(t_0)|},$$

the third limit is equal to the second, and hence zero.

For the triangles formed by the red, blue, and black vectors (one triangle for each value of h) this says that the red side decreases much faster than the blue and the black side when $h \rightarrow 0$. Equivalently, the angle at $f(t_0)$ tends to zero when $h \rightarrow 0$, accounting for the “touching effect”.

Example

The twisted cubic $f(t) = (t, t^2, t^3)$ passes through the points $P_0 = (0, 0, 0) = f(0)$ and $P_1 = (1, 1, 1) = f(1)$. Accordingly, we might say that the average direction of f for $t \in [0, 1]$ is $P_1 - P_0 = (1, 1, 1)$.

Question

Does there exist a parameter value $t_0 \in [0, 1]$ such that $f'(t_0) \in \mathbb{R}(1, 1, 1)$ (i.e., such that the average direction is realized by the derivative)?

Answer

The equation $f'(t) = (1, 2t, 3t^2) = c(1, 1, 1)$, with $t \in [0, 1]$ and $c \in \mathbb{R}$, is not solvable. Hence the answer is “No”.

Compare this with the Mean Value Theorem for real-valued functions and the following exercise.

Exercise

Suppose $f: [a, b] \rightarrow \mathbb{R}^2$ is a differentiable plane curve whose derivative is continuous and satisfies $f'(t) \neq \mathbf{0}$ for all $t \in [a, b]$. Then there exists $t_0 \in [a, b]$ and $c \in \mathbb{R}$ such that $f(b) - f(a) = c f'(t_0)$.

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Differentiation Rules for Curves

In order to distinguish vector-valued functions (i.e., curves) from scalar functions, we denote them by $\mathbf{u}(t) = (u_1(t), \dots, u_n(t))$, and for their derivatives we use the operator notation $\frac{d}{dt} \mathbf{u}(t) = \mathbf{u}'(t)$, which is sometimes more convenient.

Linearity

- ① $\frac{d}{dt} (\mathbf{u}(t) + \mathbf{v}(t)) = \frac{d}{dt} \mathbf{u}(t) + \frac{d}{dt} \mathbf{v}(t) = \mathbf{u}'(t) + \mathbf{v}'(t);$
- ② $\frac{d}{dt} (c \mathbf{u}(t)) = c \left(\frac{d}{dt} \mathbf{u}(t) \right) = c \mathbf{u}'(t) \text{ for } c \in \mathbb{R}.$

Products

- ③ $\frac{d}{dt} (c(t) \mathbf{u}(t)) = c'(t) \mathbf{u}(t) + c(t) \mathbf{u}'(t);$
- ④ $\frac{d}{dt} (\mathbf{u}(t) \cdot \mathbf{v}(t)) = \mathbf{u}'(t) \cdot \mathbf{v}(t) + \mathbf{u}(t) \cdot \mathbf{v}'(t);$
- ⑤ $\frac{d}{dt} (\mathbf{u}(t) \times \mathbf{v}(t)) = \mathbf{u}'(t) \times \mathbf{v}(t) + \mathbf{u}(t) \times \mathbf{v}'(t).$

Differentiation Rules Cont'd

Chain rule

$$6 \quad \frac{d}{dt} \mathbf{u}(c(t)) = c'(t) \mathbf{u}'(c(t)).$$

Of course, the functions involved need to satisfy certain requirements and the formulas need to hold only for certain arguments t . In the chain rule, for example, $c: J \rightarrow I$ should have range contained in the domain of $s \mapsto \mathbf{u}(s)$, and (6) holds whenever $c'(t)$ and $\mathbf{u}'(c(t))$ exist.

Proofs.

Only the proof of the rule for cross products will be given. Here the trick is to write, using bilinearity of the cross product,

$$\begin{aligned} & \frac{1}{h} (\mathbf{u}(t+h) \times \mathbf{v}(t+h) - \mathbf{u}(t) \times \mathbf{v}(t)) \\ &= \frac{1}{h} (\mathbf{u}(t+h) - \mathbf{u}(t)) \times \mathbf{v}(t+h) + \mathbf{u}(t) \times \frac{1}{h} (\mathbf{v}(t+h) - \mathbf{v}(t)) \end{aligned}$$

For $h \rightarrow 0$ this converges to $\mathbf{u}'(t) \times \mathbf{v}(t) + \mathbf{u}(t) \times \mathbf{v}'(t)$.

Proof cont'd.

Justification: We have used the fact that $\mathbf{x} \rightarrow \mathbf{a}$ and $\mathbf{y} \rightarrow \mathbf{b}$ implies $\mathbf{x} \times \mathbf{y} \rightarrow \mathbf{a} \times \mathbf{b}$, which is essentially the continuity of the cross product as a function of both factors.

This fact can be inferred from

$$\begin{aligned} |\mathbf{x} \times \mathbf{y} - \mathbf{a} \times \mathbf{b}| &= |(\mathbf{x} - \mathbf{a}) \times \mathbf{y} + \mathbf{a} \times (\mathbf{y} - \mathbf{b})| && \text{(same trick!)} \\ &\leq |\mathbf{x} - \mathbf{a}| \cdot |\mathbf{y}| + |\mathbf{a}| \cdot |\mathbf{y} - \mathbf{b}| \\ &\leq |\mathbf{x} - \mathbf{a}| \cdot (|\mathbf{b}| + 1) + |\mathbf{a}| \cdot |\mathbf{y} - \mathbf{b}| && \text{(if } |\mathbf{y} - \mathbf{b}| \leq 1\text{)} \end{aligned}$$

with some experience. □

Note

The stated differentiation rules all have Calculus I proofs. For example, one may alternatively use the explicit formula for the cross product and the theorem on the coordinate-wise computation of the derivative of a curve to reduce the product rule for cross products to the product rule for differentiation of scalar functions; cf. exercises.

Examples

Rule 3 Consider again the curve $g(t) = e^{-t}(t, t^2, t^3)$.

In this case $c(t) = e^{-t}$, $\mathbf{u}(t) = (t, t^2, t^3)$, and Rule 3 gives

$$\begin{aligned} g'(t) &= -e^{-t}(t, t^2, t^3) + e^{-t}(1, 2t, 3t^2) \\ &= e^{-t}(1-t, 2t-t^2, 3t^2-t^3). \end{aligned}$$

Another application of Rule 3 is a proof of the formula

$$\frac{d}{dt} \sum_{i=1}^n g_i(t) \mathbf{v}_i = \sum_{i=1}^n g'_i(t) \mathbf{v}_i$$

mentioned earlier:

$$\frac{d}{dt} (g_i(t) \mathbf{v}_i) = g'_i(t) \mathbf{v}_i + g_i(t) \mathbf{v}'_i = g'_i(t) \mathbf{v}_i, \text{ since } \mathbf{v}_i \text{ is constant.}$$

Rule 4 For a curve with radius vector $\mathbf{r}(t)$, the scalar function $t \mapsto |\mathbf{r}(t)|^2$ gives the squared distance at time t between the curve point and the origin. Its derivative is conveniently computed using Rule 4:

$$\frac{d}{dt} |\mathbf{r}(t)|^2 = \frac{d}{dt} (\mathbf{r}(t) \cdot \mathbf{r}(t)) = \mathbf{r}'(t) \cdot \mathbf{r}(t) + \mathbf{r}(t) \cdot \mathbf{r}'(t) = 2 \mathbf{r}(t) \cdot \mathbf{r}'(t)$$

There is also a more direct (and more cumbersome) Calculus I derivation of this formula using, e.g.,

$$|\mathbf{r}(t)|^2 = x(t)^2 + y(t)^2 + z(t)^2 \text{ for } n = 3.$$

Examples (cont'd)

Rule 6 Consider again the plane curve $f(t) = (\cos(t^2), \sin(t^2))$.

This is of the form $f(t) = \mathbf{u}(c(t))$ with $c(t) = t^2$,
 $\mathbf{u}(s) = (\cos s, \sin s)$ and may be called a
“reparametrization” of the unit circle.

Rule 6 gives

$$f'(t) = 2t \begin{pmatrix} -\sin(t^2) \\ \cos(t^2) \end{pmatrix},$$

which is a scalar multiple of the derivative of the standard parametrization of the unit circle, viz. $\mathbf{u}(s)$, at the same point (corresponding to $s = t^2$).

Consequence of Rule 6

Tangents to parametric curves are preserved by reparametrizations $s = c(t)$, provided only that $c'(t) \neq 0$.

Thus they form a property of the corresponding non-parametric curve (the range of the parametric curve).

Integration of Curves

Definition

Curves are *integrated coordinate-wise*, i.e., if

$f(t) = (f_1(t), f_2(t), \dots, f_n(t))$ and the f_i are integrable over $[a, b]$ then

$$\int_a^b f(t) dt = \left(\int_a^b f_1(t) dt, \int_a^b f_2(t) dt, \dots, \int_a^b f_n(t) dt \right).$$

Notes

- Similar to differentiation, linearity of the integral implies that $\int_a^b (\sum_{i=1}^n g_i(t) \mathbf{v}_i) dt = \sum_{i=1}^n \left(\int_a^b g_i(t) dt \right) \mathbf{v}_i$ for any basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of \mathbb{R}^n (provided g_i are integrable over $[a, b]$).
- The Fundamental Theorem of Calculus continues to hold in the n -dimensional setting:

$$\int_a^b f(t) dt = F(b) - F(a) \quad \text{if } F'(t) = f(t) \text{ for } t \in [a, b]$$

and f is continuous on $[a, b]$.

Arc Length

First recall the following fundamental property of \mathbb{R} .

Supremum

If a non-empty subset X of \mathbb{R} has an upper bound s (i.e., $x \leq s$ for all $x \in X$, then it has a *least upper bound* s_0 (i.e., $s_0 \leq s$ for all upper bounds s). The number $s_0 \in \mathbb{R}$, which is uniquely determined, is called *supremum* of X and denoted by $\sup X$.

Dually, if $X \neq \emptyset$ is bounded from below, there exists a *greatest lower bound* $\inf X$, the *infimum* of X .

Observation 1

Given a curve $f: I \rightarrow \mathbb{R}^n$ and a subdivision $T = \{t_0, t_1, \dots, t_N\}$ of its parameter interval I (i.e., $t_i \in I$ and $t_0 < t_1 < \dots < t_N$), the *polygonal path* with vertices $P_0 = f(t_0)$, $P_1 = f(t_1)$, \dots , $P_N = f(t_N)$ provides an approximation to f and hence its length

$$L(T, f) = \sum_{i=1}^N |P_i - P_{i-1}|$$

an approximation to the arc length of the curve f .

Observation 2

If additional points are added to the path (i.e., T is replaced by some $T' \supseteq T$), the length $L(T, f)$ can only increase.

For the case of one additional point $P = f(t)$, $t \in (t_{i-1}, t_i)$, the triangle inequality shows this:

$$\begin{aligned} L(T', f) &= \sum_{j=1}^{i-1} |P_j - P_{j-1}| + |P - P_{i-1}| + |P_i - P| + \sum_{j=i+1}^N |P_j - P_{j-1}| \\ &\geq \sum_{j=1}^{i-1} |P_j - P_{j-1}| + |P_i - P_{i-1}| + \sum_{j=i+1}^N |P_j - P_{j-1}| = L(T, f). \end{aligned}$$

This motivates the following

Definition

A parametric curve $f: I \rightarrow \mathbb{R}^n$ (where I is an interval in \mathbb{R}) is said to be *rectifiable* if there exists $s \in \mathbb{R}$ such that $L(T, f) \leq s$ for all subdivisions T of I . If this is the case,

$$L(f) = \sup\{L(T, f); T \text{ a subdivision of } I\}$$

is called *(arc) length* of f .

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

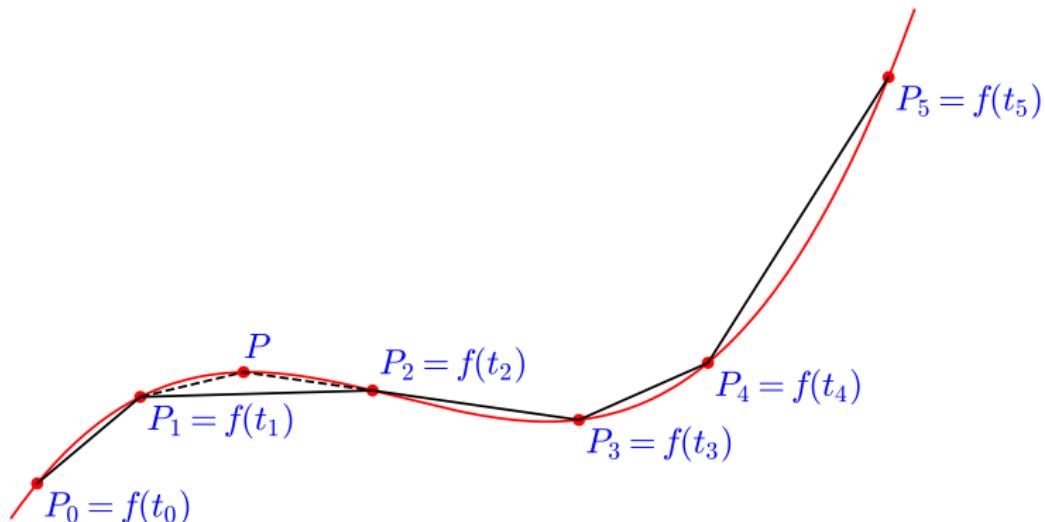


Figure: Polygonal path for f with $T = \{t_0, t_1, t_2, t_3, t_4, t_5\}$ and an extension to a polygonal path with $T' = T \cup \{t\}$, $t \in (t_1, t_2)$

Another Definition of Arc Length

Outlined in our textbook

This definition somewhat resembles that of the Riemann integral by means of Riemann sums. It requires $I = [a, b]$ to be compact (i.e., closed and bounded), and restricts attention to subdivisions $a = t_0 < t_1 < \dots < t_N = b$ (so-called *partitions* of $[a, b]$).

The *mesh size* of such a partition T is $\mu(T) = \max\{t_i - t_{i-1}; 1 \leq i \leq N\}$, the maximum length of the N intervals $[a, t_1], [t_1, t_2], \dots, [t_{N-1}, b]$.

Second Definition

A parametric curve $f: [a, b] \rightarrow \mathbb{R}^n$ is said to be *rectifiable* if there exists $L \in \mathbb{R}$ such that for every $\epsilon > 0$ there exists $\delta > 0$ such that $|L(T, f) - L| < \epsilon$ for all partitions T of $[a, b]$ of mesh size $\mu(T) < \delta$. If this is the case, L is called *(arc) length* of f .

Notes

- The 2nd definition appears to be stronger, but in fact both definitions are equivalent for continuous curves.
- The 1st definition is more general (it applies to all intervals $I \subseteq \mathbb{R}$) and easier to work with, since only an apparently weaker property (boundedness of $L(T, f)$) needs to be checked.

Equivalence of Both Definitions

In the cases where the 2nd definition applies

For the “limit” in the 2nd definition we write succinctly

$$L = \lim_{\mu(T) \rightarrow 0} L(T, f).$$

Theorem

Let $f: [a, b] \rightarrow \mathbb{R}^n$ be a parametric curve with compact (i.e., closed and bounded) parameter interval.

- ① If $L = \lim_{\mu(T) \rightarrow 0} L(T, f)$ exists, then $L(f) = \sup_T L(T, f) = L$.
- ② If $L(f) = \sup_T L(T, f)$ exists in \mathbb{R} and f is continuous, then $\lim_{\mu(T) \rightarrow 0} L(T, f) = L(f)$.

Note that it does not matter whether we take the supremum over all subdivisions of $[a, b]$ or only those subdivisions which contain the endpoints a, b .

The assumption in (2) that f be continuous imposes no restriction, since arc length makes sense only for continuous curves. As an example consider the discontinuous plane curve $f(t) = (t, \lfloor t \rfloor)$ for $t \in [0, 1]$, which has $L(f) = 2$. Draw the curve and see what happens.

Uniform Continuity

Needed for the proof of the theorem

Recall that $f: I \rightarrow \mathbb{R}$ (or \mathbb{R}^n) is continuous if for $t_0 \in I$ and $\epsilon > 0$ there is a “response” $\delta = \delta(t_0, \epsilon)$ (i.e., δ may depend on t_0 and ϵ) such that $|f(t) - f(t_0)| < \epsilon$ whenever $t \in I$ and $|t - t_0| < \delta$.

Definition

f is *uniformly continuous* if for every $\epsilon > 0$ a response $\delta = \delta(\epsilon)$ can be found that works simultaneously for all $t_0 \in I$, i.e.,

$$|f(t) - f(t')| < \epsilon \quad \text{whenever } t, t' \in I \text{ and } |t - t'| < \delta.$$

Uniform continuity in general is stronger than continuity (per se), but we have the following

Lemma

If $f: I \rightarrow \mathbb{R}$ (or \mathbb{R}^n) is continuous and $I = [a, b]$ is a compact interval, then f is uniformly continuous.

A proof of the lemma will be outlined afterwards.

Proof of the theorem.

(1) Clearly $L \leq \sup_T L(T, f)$ and it suffices to show that $L \geq L(T, f)$ for all partitions T .

Assume $L < L(T, f)$ for some T and set $\epsilon = L(T, f) - L$. There exists $\delta > 0$ such that $|L(T', f) - L| < \epsilon$ whenever $\mu(T') < \delta$.
 $\Rightarrow L(T', f) < L(T, f)$ whenever $\mu(T') < \delta$.

But the common refinement $T'' = T \cup T'$ of T and one such T' has $L(T'', f) \geq L(T, f)$ and still $\mu(T'') < \delta$; contradiction.

(2) Given $\epsilon > 0$, we can choose a partition T with $L(T, f) > L(f) - \epsilon/2$.

Now consider an arbitrary partition T' of $[a, b]$ with $\mu(T') < \delta$, where $\delta > 0$ has yet to be determined.

The common refinement $T'' = T \cup T'$ has $L(T'', f) \geq L(T, f)$ and hence also $L(T'', f) > L(f) - \epsilon/2$. If $T = \{t_0, t_1, \dots, t_N\}$, at most $N - 1$ line segments of the inscribed polygonal path corresponding to T' need to be replaced by a pair of new line segments to produce the polygonal path corresponding to T'' . Denoting by ℓ an upper bound for the lengths of the new line segments, we conclude $L(T'', f) \leq L(T', f) + 2(N - 1)\ell$.

Proof cont'd.

By construction, the new line segments have the form $[f(t''_{i-1}), f(t''_i)]$ with $0 < t''_i - t''_{i-1} < \delta$. Since f is uniformly continuous on $[a, b]$, there exists for every $\ell > 0$ a corresponding choice $\delta > 0$ such that $|f(t''_{i-1}) - f(t''_i)| < \ell$ for all new line segments, provided only that T' satisfies $\mu(T') < \delta$.

\Rightarrow By setting $2(N-1)\ell = \epsilon/2$, i.e., $\ell = \frac{\epsilon}{4(N-1)}$, we can achieve $L(T'', f) \leq L(T', f) + \epsilon/2$.

$$\Rightarrow L(T', f) \geq L(T'', f) - \frac{\epsilon}{2} > L(f) - \frac{\epsilon}{2} - \frac{\epsilon}{2} = L(f) - \epsilon$$

for all partitions T' with $\mu(T') < \delta$.

This proves $\lim_{\mu(T) \rightarrow 0} L(T, f) = L(f)$. □

Introduction

Limits and
Continuity

Limits of Sequences

of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Notes on the proof

- Since $\mu(T') < \delta$ implies $\mu(T'') < \delta$, all line segments of T'' (not only the newly inserted ones) satisfy the condition $|f(t''_{i-1}) - f(t''_i)| < \ell$.
- The geometric meaning of uniform continuity used in the proof, which may help in understanding the abstract concept, deserves to be highlighted:

A parametric curve $f: [a, b] \rightarrow \mathbb{R}^n$ is uniformly continuous iff for every $\epsilon > 0$ there exists $\delta > 0$ such that for any partition T of $[a, b]$ with $\mu(T) < \delta$ all line segments of the corresponding polygonal path have length $< \epsilon$.

A Handy Formula for Computing Arc Lengths

Theorem

If $f: [a, b] \rightarrow \mathbb{R}^n$ is a C^1 -curve (i.e., f' exists and is continuous on $[a, b]$) with compact parameter interval, then f is rectifiable and

$$L(f) = \int_a^b |f'(t)| dt.$$

Proof.

Clearly in the definition of $L(f)$ we may restrict attention to subdivisions T with $t_0 = a$, $t_N = b$. For such a subdivision we have

$$\int_a^b |f'(t)| dt = \sum_{i=1}^N \int_{t_{i-1}}^{t_i} |f'(t)| dt,$$

$$L(T, f) = \sum_{i=1}^N |f(t_i) - f(t_{i-1})| = \sum_{i=1}^N \left| \int_{t_{i-1}}^{t_i} f'(t) dt \right|$$

Proof cont'd.

Since $\left| \int_{t_{i-1}}^{t_i} f'(t) dt \right| \leq \int_{t_{i-1}}^{t_i} |f'(t)| dt$ (a nontrivial exercise, cf.

Homework 3, H17), we get $L(T, f) \leq \int_a^b |f'(t)| dt$ for all subdivisions T and hence that f is rectifiable with $L(f) \leq \int_a^b |f'(t)| dt$.

It remains to show that for each $\epsilon > 0$ there exists a subdivision T with $L(T, f) \geq \int_a^b |f'(t)| dt - \epsilon$.

Since $t \mapsto f'(t)$ is continuous on $[a, b]$, it is uniformly continuous (cf. the earlier lemma), and hence there exists $\delta > 0$ such that $|f'(t_1) - f'(t_2)| < \epsilon$ for all $t_1, t_2 \in [a, b]$ with $|t_1 - t_2| < \delta$.

Now choose a subdivision $T = \{t_0, \dots, t_N\}$ satisfying

$$\Delta t_i = t_i - t_{i-1} < \delta \text{ for } 1 \leq i \leq N.$$

$$\implies |f'(t)| \leq |f'(t_i)| + \epsilon \text{ for } t \in [t_{i-1}, t_i]$$

Proof con't.

$$\begin{aligned}
 & \implies \int_{t_{i-1}}^{t_i} |f'(t)| dt \leq |f'(t_i)| \Delta t_i + \epsilon \Delta t_i \\
 & = \left| \int_{t_{i-1}}^{t_i} (f'(t) + f'(t_i) - f'(t)) dt \right| + \epsilon \Delta t_i \\
 & \leq \left| \int_{t_{i-1}}^{t_i} f'(t) dt \right| + \left| \int_{t_{i-1}}^{t_i} (f'(t_i) - f'(t)) dt \right| + \epsilon \Delta t_i \\
 & \leq |f(t_i) - f(t_{i-1})| + 2\epsilon \Delta t_i
 \end{aligned}$$

Adding these N inequalities gives

$$\int_a^b |f'(t)| dt \leq L(T, f) + 2\epsilon(b-a), \quad \text{i.e.}$$

$$L(T, f) \geq \int_a^b |f'(t)| dt - 2\epsilon(b-a).$$

Replacing ϵ by $\frac{\epsilon}{2(b-a)}$ then finishes the proof. □

Example

We compute the arc length of the helix $f(t) = (\cos t, \sin t, t)$, restricted to $t \in [0, 2\pi]$.

$$\begin{aligned} L &= \int_0^{2\pi} \left| \begin{pmatrix} -\sin t \\ \cos t \\ 1 \end{pmatrix} \right| dt = \int_0^{2\pi} \sqrt{\sin^2 t + \cos^2 t + 1} dt \\ &= \int_0^{2\pi} \sqrt{2} dt = 2\sqrt{2}\pi. \end{aligned}$$

Compare with the arc length of its projection onto the (x, y) -plane, the unit circle, which is

$$\int_0^{2\pi} \left| \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix} \right| dt = \int_0^{2\pi} \sqrt{\sin^2 t + \cos^2 t} dt = \int_0^{2\pi} 1 dt = 2\pi.$$

Remark

Like tangent lines, arc length is invariant under (continuous, bijective) reparametrizations of a curve and thus a property of the non-parametric curve (the range); cf. subsequent discussion. This is why we can speak of “arc length of the unit circle”.

Piece-wise C^1 -Curves

A curve $f: [a, b] \rightarrow \mathbb{R}^n$ is called a *piece-wise C^1 -curve* if there exists a subdivision $a = t_0 < t_1 < \dots < t_r = b$ such that the restriction of f to $[t_{i-1}, t_i]$ is a C^1 -curve for all $i \in \{1, \dots, r\}$.

The arc length formula generalizes to piece-wise C^1 -curves, as is easily seen using

$$L(f) = \int_a^b |f'(t)| dt = \sum_{i=1}^r \int_{t_{i-1}}^{t_i} |f'(t)| dt.$$

Example

The plane curve

$$f(t) = \begin{cases} (\cos(\pi t), \sin(\pi t)) & \text{for } 0 \leq t \leq 1, \\ (2t - 3, 0) & \text{for } 1 \leq t \leq 2, \end{cases}$$

parametrizes a semi-circle, and its arc length is given by

$$L(f) = \int_0^1 \sqrt{(-\pi \sin(\pi t))^2 + (\pi \cos(\pi t))^2} dt + \int_1^2 \sqrt{2^2 + 0^2} dt = \pi + 2.$$

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

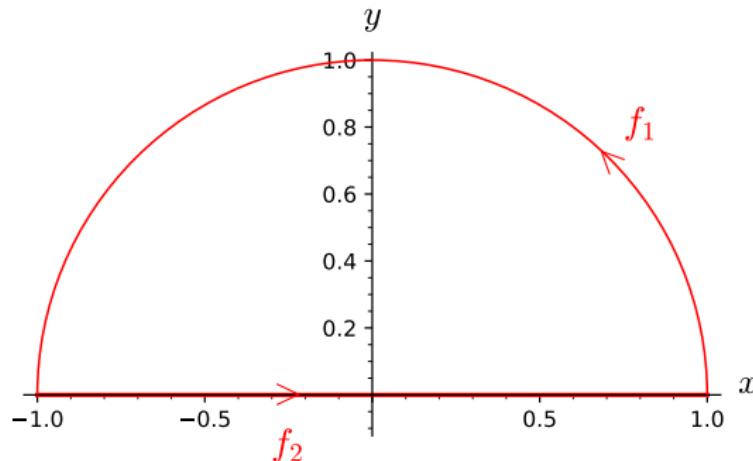


Figure: The two smooth pieces f_1, f_2 of the curve f

$$\begin{aligned}f_1(t) &= (\cos(\pi t), \sin(\pi t)) && \text{for } 0 \leq t \leq 1, \\f_2(t) &= (2t - 3, 0) && \text{for } 1 \leq t \leq 2.\end{aligned}$$

Note

That f is piece-wise C^1 entails in particular that f is continuous and for $i \in \{1, \dots, r\}$ the one-sided derivatives

$$f'_+(t_{i-1}) = \lim_{t \downarrow t_{i-1}} \frac{f(t) - f(t_{i-1})}{t - t_{i-1}}, \quad f'_-(t_i) = \lim_{t \uparrow t_i} \frac{f(t) - f(t_i)}{t - t_i}$$

exist and that $\lim_{t \downarrow t_{i-1}} f'(t) = f'_+(t_{i-1})$, $\lim_{t \uparrow t_i} f'(t) = f'_-(t_i)$.

Example

The curve $f: [0, \frac{1}{\pi}] \rightarrow \mathbb{R}^2$ defined by

$$f(t) = \begin{cases} (t, t \cos(1/t)) & \text{if } 0 < t \leq \frac{1}{\pi}, \\ (0, 0) & \text{if } t = 0, \end{cases}$$

is continuous in $[0, \frac{1}{\pi}]$ and differentiable in $(0, \frac{1}{\pi}]$, but not rectifiable.

Setting $t_k = \frac{1}{k\pi}$ for $k = 1, 2, \dots$, we have $f(t_k) = \frac{1}{k\pi}(1, (-1)^k)$, $|f(t_k) - f(t_{k-1})| \geq \frac{2}{k\pi}$, and hence for $T_N = \{t_1, \dots, t_N\}$ the estimate $L(T_N, f) \geq \frac{2}{\pi} \sum_{k=2}^N \frac{1}{k}$, which on account of the divergence of the harmonic series shows that $L(T_N, f)$ is unbounded.

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

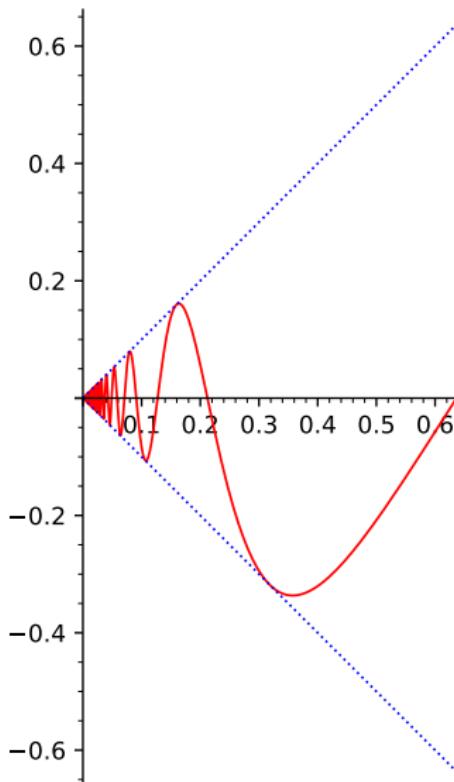


Figure: The graph of $F(x) = x \cos(1/x)$, $x > 0$, extended
continuously to $[0, \infty)$ by $F(0) = 0$

Uniform Continuity Continued

Here are a few examples/counterexamples for uniform continuity of real valued functions:

- 1 The function $f(x) = x^2$, $x \in \mathbb{R}$ is not uniformly continuous, since, e.g.,

$$|x^2 - x_0^2| = |x + x_0| |x - x_0| < 1$$

cannot be achieved by merely requiring $|x - x_0|$ to be small. More precisely, for any positive integer n the system

$x - x_0 = 1/n \wedge x + x_0 = n$ can be solved for x, x_0 to yield
 $|x^2 - x_0^2| = 1$.

- 2 The function $g(x) = \sqrt{x}$, $x \in [0, \infty)$ is uniformly continuous, although its derivative for $x \downarrow 0$ is unbounded. This follows from

$$|\sqrt{x} - \sqrt{x_0}|^2 \leq |\sqrt{x} - \sqrt{x_0}| (\sqrt{x} + \sqrt{x_0}) = |x - x_0|,$$

which shows that $\delta = \epsilon^2$ provides a uniform response.

- ③ Any C^1 -function $f: [a, b] \rightarrow \mathbb{R}$ on a compact interval is uniformly continuous. This follows from the Mean Value Theorem of Calculus I, viz.

$$f(x) - f(x_0) = f'(\xi)(x - x_0)$$

for some ξ between x and x_0 , together with the fact that f' is bounded. If $|f'| \leq M$ on $[a, b]$ then $\delta = \epsilon/M$ provides the required uniform response.

Now we supply a proof of the earlier lemma that a continuous function $f: [a, b] \rightarrow \mathbb{R}$ (or curve $f: [a, b] \rightarrow \mathbb{R}^n$) is necessarily uniformly continuous. The proof is based on the Bolzano-Weierstrass Theorem, which is arguably the most important tool in a rigorous development of Calculus.

Theorem (Bolzano-Weierstrass)

Every bounded sequence (x_n) in \mathbb{R} has a convergent subsequence, i.e., there exist positive integers $n_1 < n_2 < n_3 < \dots$ and $x \in \mathbb{R}$ such that $\lim_{k \rightarrow \infty} x_{n_k} = x$.

Proof.

By rescaling, if necessary, we may assume that $x_n \in [0, 1]$ for all $n \in \mathbb{N}$. Set $n_0 = 0$ and $I_0 = [0, 1]$.

Split I_0 into two closed intervals of equal length, viz. $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$. At least one of these intervals, call it I_1 , must contain infinitely many sequence terms x_n . Thus it is possible to choose $n_1 > n_0$ such that $x_{n_1} \in I_1$. Then split I_1 into two closed intervals of equal length; call one of these, which contains infinitely many sequence terms, I_2 , and select $n_2 > n_1$ with $x_{n_2} \in I_2$; etc.

By induction, we obtain a nested sequence $I_0 \supset I_1 \supset I_2 \supset \dots$ of closed intervals I_k , with I_k of length 2^{-k} , and a subsequence (x_{n_k}) of (x_n) satisfying $x_{n_k} \in I_k$.

Writing $I_k = [a_k, b_k]$, it is then easy to see that

$$\lim_{k \rightarrow \infty} x_{n_k} = \lim_{k \rightarrow \infty} a_k = \sup\{a_k; k \in \mathbb{N}\},$$

the unique point contained in $\bigcap_{k=1}^{\infty} I_k$. □

Proof of the lemma.

Assume that $f: [a, b] \rightarrow \mathbb{R}$ is continuous but not uniformly continuous.

Then there exists $\epsilon_0 > 0$ and for each $n \in \mathbb{Z}^+$ a pair $t_n, t'_n \in [a, b]$ satisfying

$$|t_n - t'_n| < 1/n \quad \text{and} \quad |f(t_n) - f(t'_n)| \geq \epsilon_0.$$

Reason: No $\delta = 1/n$ works as a response for ϵ_0 .

By the Bolzano-Weierstrass Theorem, (t_n) has a convergent subsequence (t_{n_k}) ; let $t = \lim_{k \rightarrow \infty} t_{n_k}$.

$[a, b]$ is closed $\Rightarrow t \in [a, b]$

$|t_n - t'_n| < 1/n \Rightarrow t'_{n_k} \rightarrow t$ for $k \rightarrow \infty$ as well

Continuity then implies that

$$\lim_{k \rightarrow \infty} f(t_{n_k}) = f(t) = \lim_{k \rightarrow \infty} f(t'_{n_k}).$$

$$\Rightarrow f(t_{n_k}) - f(t'_{n_k}) \rightarrow 0 \text{ for } k \rightarrow \infty.$$

But this contradicts $|f(t_{n_k}) - f(t'_{n_k})| \geq \epsilon_0$. □

Exercise

Suppose $f, g: I \rightarrow \mathbb{R}^n$ are related by a coordinate change, $g(t) = \mathbf{A}f(t) + \mathbf{b}$ with $\mathbf{A} \in \mathbb{R}^{n \times n}$ invertible and $\mathbf{b} \in \mathbb{R}^n$. Show:

- ① f is rectifiable iff g is rectifiable.
- ② If $\mathbf{A} = \lambda \mathbf{I}_n$, $\lambda \in \mathbb{R} \setminus \{0\}$, then $L(g) = |\lambda| L(f)$ (provided one, and hence both of f, g are rectifiable).
- ③ If \mathbf{A} is orthogonal then $L(g) = L(f)$ (provided one, and hence both of f, g are rectifiable).

Exercise

Show that the circumference c of the ellipse with equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (a \geq b > 0) \text{ is given by}$$

$$c = 4a \int_0^{\pi/2} \sqrt{1 - e^2 \sin^2 t} dt, \quad e = \sqrt{1 - \frac{b^2}{a^2}}.$$

The quantity $e \in [0, 1)$ is called (*numerical*) eccentricity of the ellipse, and $E(e) = \int_0^{\pi/2} \sqrt{1 - e^2 \sin^2 t} dt$ complete elliptic integral of the second kind.

Reparametrization

Suppose $I, J \subseteq \mathbb{R}$ are intervals, $f: I \rightarrow \mathbb{R}^n$ is a curve and $t = t(s): J \rightarrow I$ is an order-preserving or order-reversing bijection (i.e., surjective and either strictly increasing or strictly decreasing).

Definition

The curve $g: J \rightarrow \mathbb{R}^n$, $s \mapsto f(t(s))$ is called a *reparametrization* of f .

Notes

- All reparametrizations g of a rectifiable curve f have the same arc length as f . This follows directly from the definition of arc length, since subdivisions of J and I are in one-to-one correspondence via $\{s_0, s_1, \dots, s_N\} \mapsto \{t(s_0), t(s_1), \dots, t(s_N)\}$ if $s \mapsto t(s)$ is increasing, respectively,
 $\{s_0, s_1, \dots, s_N\} \mapsto \{t(s_N), t(s_{N-1}), \dots, t(s_0)\}$ if $s \mapsto t(s)$ is decreasing.
- If $[\alpha, \beta] \rightarrow [a, b]$, $s \mapsto t(s)$ is a (surjective) C^1 -function with $t'(s) \neq 0$ for $s \in [\alpha, \beta]$, then it must be an order-preserving or order-reversing bijection, since by continuity either $t'(s) > 0$ for all $s \in [\alpha, \beta]$ or $t'(s) < 0$ for all $s \in [\alpha, \beta]$.

Notes cont'd

In this case and if f is a C^1 -curve, we can also apply the arc length formula to conclude that $L(f) = L(g)$ as follows:

Case 1: $t'(s) > 0$.

The substitution rule of Calculus I gives

$$L(f) = \int_a^b |f'(t)| dt = \int_{\alpha}^{\beta} |f'(t(s))| t'(s) ds$$

By the chain rule, $|g'(s)| = |f'(t(s))t'(s)| = |f'(t(s))| t'(s)$, and hence $L(f) = \int_{\alpha}^{\beta} |g'(s)| ds = L(g)$.

Case 2: $t'(s) < 0$.

Here we have $t(\alpha) = b$, $t(\beta) = a$, and $|t'(s)| = -t'(s)$.

\implies The substitution rule gives again

$$L(f) = \int_{\beta}^{\alpha} -|f'(t(s))t'(s)| ds = \int_{\alpha}^{\beta} |f'(t(s))t'(s)| ds = L(g).$$

Example

Consider the parametric curves

$$f(t) = (\cos t, \sin t), \quad t \in [0, 2\pi] \quad \text{and}$$

$$g(s) = (\cos(2s), \sin(2s)), \quad s \in [0, \pi].$$

Both provide a parametrization of the unit circle;

g is obtained from f using the reparametrization $t(s) = 2s$, $s \in [0, \pi]$, which maps $[0, \pi]$ bijectively onto $[0, 2\pi]$. This implies $L(g) = L(f) = 2\pi$.

Physically speaking, g moves once around the unit circle in the same direction as f , but at twice the speed and in half the time.

Example

The curve $g(t) = (\cos(t^2), \sin(t^2))$, $t \in [0, 1]$ is a reparametrization of $f(t) = (\cos t, \sin t)$, $t \in [0, 1]$.

$\Rightarrow L(g) = L(f) = \text{guess what?}$

For $h(t) = (\cos(1 - t^2), \sin(1 - t^2))$, $t \in [0, 1]$ the same is true, except that now the reparametrization is order-reversing.

$\Rightarrow L(h) = L(f) = \dots$ as well.

Smooth Parametric Curves

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

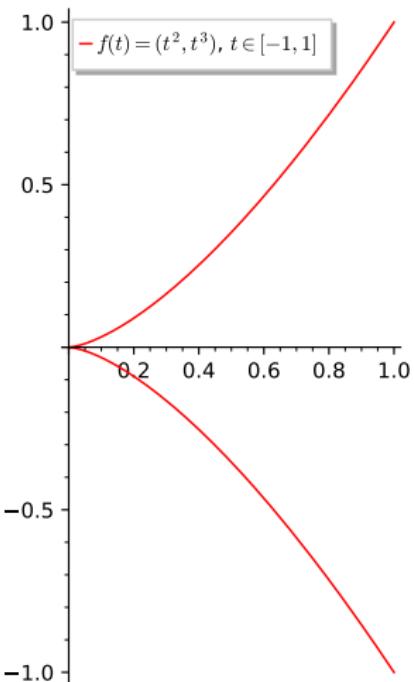
Definition

A parametric curve $f: I \rightarrow \mathbb{R}^n$ is said to be *smooth* if f is a C^1 -curve and $f'(t) \neq \mathbf{0}$ for all $t \in I$.

A smooth parametric curve has a tangent line at each curve point $f(t)$, viz. the line $L = f(t) + \mathbb{R}f'(t)$. (But if f is not one-to-one, the same curve point $f(t_1) = f(t_2)$, $t_1 < t_2$, may have different tangents, since $f'(t_1) \neq f'(t_2)$ is possible.)

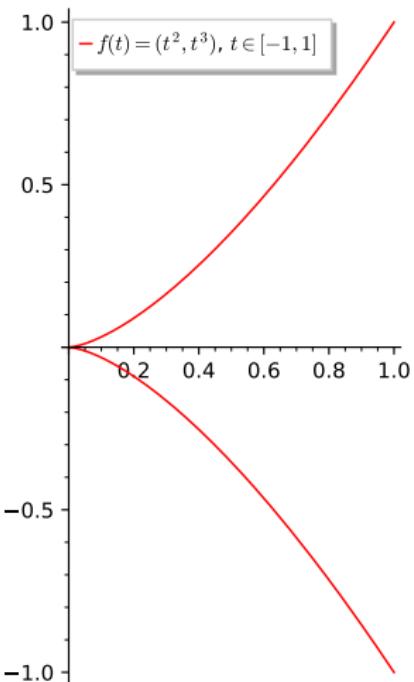
Example (A non-smooth C^1 -curve)

Neile's semicubic parabola $f(t) = (t^2, t^3)$, $t \in \mathbb{R}$ has a cusp in $f(0) = (0, 0)$. The tangent direction makes a U-turn in $t = 0$, which can only happen if the derivative at $t = 0$ vanishes. Indeed, we have $f'(t) = (2t, 3t^2)$ and hence $f'(0) = (0, 0)$.



Example (A smooth curve with a double point)

The plane curve $f(t) = (t^2 - 1, t^3 - t)$, $t \in \mathbb{R}$ is smooth, since $f'(t) = (2t, 3t^2 - 1) \neq (0, 0)$ for all $t \in \mathbb{R}$. The curve passes the origin twice (for $t = \pm 1$), with different tangent directions $f'(\pm 1) = (\pm 2, 2)$.



Parametrization with Respect to Arc Length

Definition

Suppose $f: [a, b] \rightarrow \mathbb{R}^n$ is smooth. Then

$$s(t) = \int_a^t |f'(\tau)| d\tau, \quad t \in [a, b]$$

is called the *arc length function* and $g(s) = f(t(s))$, where $t(s)$ is the inverse function of $s(t)$, *parametrization of f with respect to arc length*.

(The symbol ' τ ' has been used to denote the integration variable, because τ is the Greek letter for t .)

Notes

- Since f is smooth, we have $s'(t) = |f'(t)| > 0$ for $t \in [a, b]$, and hence $s: [a, b] \rightarrow [0, L(f)]$ is an order-preserving bijection; in particular, the inverse $t(s)$ is well-defined.
- The curve g has domain $[0, L(f)]$ and is also smooth.
- The definition generalizes to the case in which the domain of f is an arbitrary interval I , if one picks $a \in I$ first.
- A curve is parametrized with respect to arc length iff its derivative has constant length 1. We check this for the curve $g(s) = f(t(s))$ defined above:

By the Fundamental Theorem of Calculus, the arc length function has derivative $s'(t) = \frac{d}{dt} \int_a^t |f'(\tau)| d\tau = |f'(t)|$.

$$\begin{aligned} \implies g'(s) &= \frac{d}{ds} f(t(s)) = f'(t(s)) t'(s) = f'(t(s)) \frac{1}{s'(t(s))} \\ &= \frac{f'(t(s))}{|f'(t(s))|}, \end{aligned}$$

a vector of length 1. (Strictly speaking, we should write $\frac{1}{|\mathbf{x}|} \mathbf{x}$ for such a vector and not $\frac{\mathbf{x}}{|\mathbf{x}|}$.)

Curvature

Definition

Let $f: I \rightarrow \mathbb{R}^3$ be a smooth space curve.

- ① The *unit tangent vector* of f at “time” $t \in I$ is defined as

$$\mathbf{T}(t) = \frac{\mathbf{f}'(t)}{|\mathbf{f}'(t)|}.$$

- ② If f is a C^2 -curve, the *curvature* of f at $t \in I$ is defined as

$$\kappa(t) = \frac{|\mathbf{T}'(t)|}{|\mathbf{f}'(t)|}.$$

- ③ If f is a C^2 -curve and $\kappa(t) \neq 0$ (i.e., $t \mapsto \mathbf{T}(t)$ is smooth), the *unit normal vector* and *binormal vector* of f at $t \in I$ are

$$\mathbf{N}(t) = \frac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|}, \quad \text{resp.}, \quad \mathbf{B}(t) = \mathbf{T}(t) \times \mathbf{N}(t).$$

Introduction

Limits and Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation and Integration

Arc Length

Remarks on Uniform
Continuity

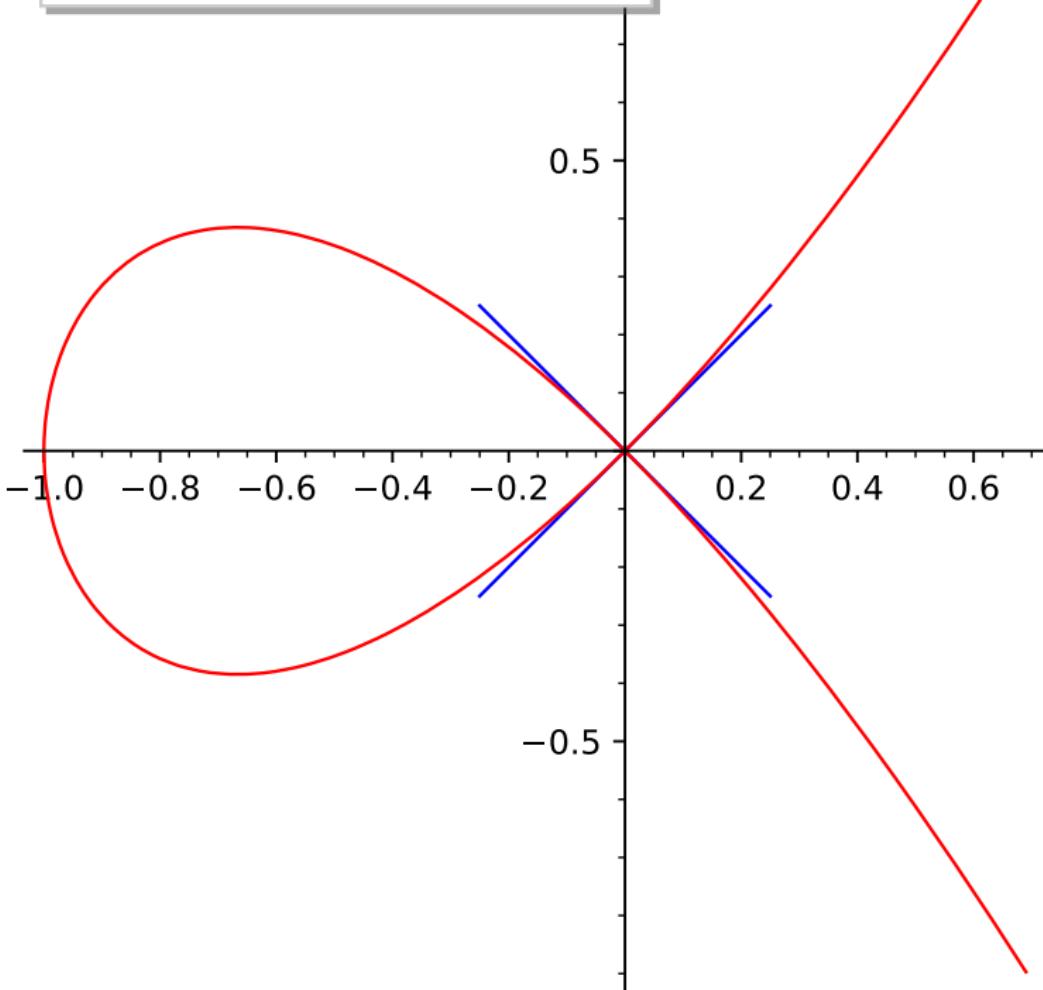
Reparametrization
Smooth Parametric
Curves

Curvature and Related Concepts

Parametrization with
Respect to Arc
Length

Curvature

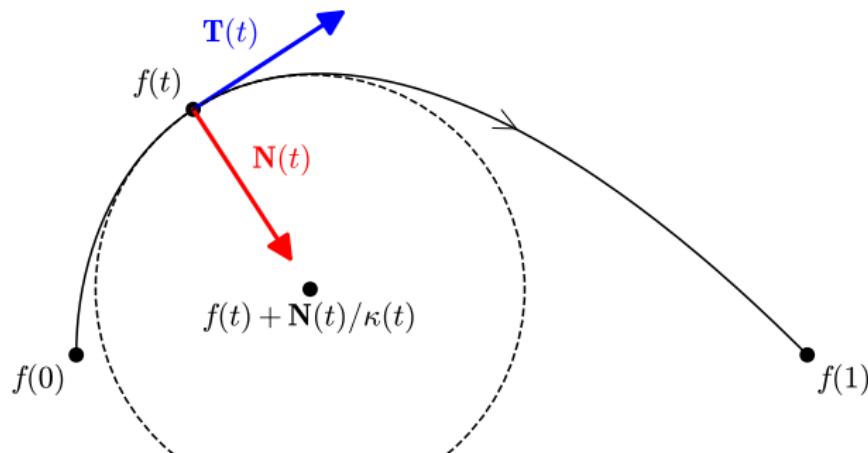
Osculating Planes
and Circles
Examples



If you want to know how this plot was generated—here is it again with coordinate axes. The curve is

$$f(t) = 4(t^2 - 1, t - t^3), \quad t \in [0, 1],$$

a slight variant of that with the double point shown earlier.



Since $\kappa(0.4) \approx 0.852$, the radius of the circle is slightly larger than 1. The maximum of κ is at $t^* \approx 0.42$ with value $\kappa(t^*) \approx 0.856$.

Notes on the definition

- Unit tangent vectors are invariant under orientation-preserving reparametrizations $g(s) = f(t(s))$ and point into the opposite direction for orientation-reversing reparametrizations, since

$$\mathbf{T}_g(s) = \frac{g'(s)}{|g'(s)|} = \frac{f'(t(s))t'(s)}{|f'(t(s))| |t'(s)|} = \pm \mathbf{T}_f(t), \quad t = t(s).$$

- The curvature $\kappa(t)$ measures how quickly the curve f changes direction at the point $f(t)$. It is invariant under both types of reparametrizations. If f is parametrized with respect to arc length s , we have $\kappa(s) = |\mathbf{T}'(s)|$.
- The vectors $\mathbf{T}(t)$, $\mathbf{N}(t)$, $\mathbf{B}(t)$ are mutually orthogonal vectors of length 1 and with positive orientation. This follows from

$$\mathbf{T}(t) \cdot \mathbf{T}'(t) = \frac{1}{2} \frac{d}{dt} |\mathbf{T}(t)|^2 = 0.$$

They form the so-called *Frenet-Serret frame (TNB frame)* of f .

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Notes cont'd

- $\mathbf{T}(t)$, $\mathbf{N}(t)$ can also be obtained by *orthonormalizing* $f'(t)$, $f''(t)$ in the usual way:

$$\mathbf{u}_1 = \frac{f'(t)}{|f'(t)|} = \mathbf{T}(t),$$

$$\mathbf{u}'_2 = f''(t) - \frac{f''(t) \cdot f'(t)}{|f'(t) \cdot f'(t)} f'(t)$$

$$= |f'(t)| \left(\frac{f''(t)}{|f'(t)|} - \frac{f'(t) \cdot f''(t)}{|f'(t)|^3} f'(t) \right)$$

$$= |f'(t)| \mathbf{T}'(t) \quad (\text{to see this, compute } \mathbf{T}'(t) = \frac{d}{dt} \frac{f'(t)}{|f'(t)|})$$

$$\mathbf{u}_2 = \frac{\mathbf{u}'_2}{|\mathbf{u}'_2|} = \frac{\mathbf{T}'(t)}{|\mathbf{T}'(t)|} = \mathbf{N}(t).$$

- *Frenet-Serret formulas* for C^3 -curves parametrized w.r.t. arc length:

$$\mathbf{T}'(s) = \kappa(s) \mathbf{N}(s),$$

$$\mathbf{N}'(s) = -\kappa(s) \mathbf{T}(s) + \tau(s) \mathbf{B}(s),$$

$$\mathbf{B}'(s) = -\tau(s) \mathbf{N}(s).$$

Notes cont'd

- The curvature of a smooth C^2 -curve f is given in terms of $f'(t)$, $f''(t)$ as

$$\kappa(t) = \frac{|f'(t) \times f''(t)|}{|f'(t)|^3}.$$

For the proof we use the preceding expression for $\mathbf{T}'(t)$, Pythagoras' Theorem in the form $|\mathbf{x} - \mathbf{y}|^2 = |\mathbf{x}|^2 - |\mathbf{y}|^2$ if $\mathbf{x} - \mathbf{y} \perp \mathbf{y}$ (see the previous computation of \mathbf{u}_2'), and $|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2$ for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$.

$$\begin{aligned}\kappa(t)^2 &= \frac{|\mathbf{T}'(t)|^2}{|f'(t)|^2} = \left| \frac{f''(t)}{|f'(t)|^2} - \frac{f'(t) \cdot f''(t)}{|f'(t)|^4} f'(t) \right|^2 \\ &= \frac{|f''(t)|^2}{|f'(t)|^4} - \frac{(f'(t) \cdot f''(t))^2}{|f'(t)|^6} \quad (\text{since } \mathbf{u}_1 \perp \mathbf{u}_2') \\ &= \frac{|f''(t)|^2 |f'(t)|^2 - (f'(t) \cdot f''(t))^2}{|f'(t)|^6} \\ &= \frac{|f'(t) \times f''(t)|^2}{|f'(t)|^6}.\end{aligned}$$

Introduction

Limits and
ContinuityLimits of Sequences
of PointsLimits and Continuity
for CurvesTopology of Subsets
of \mathbb{R}^n Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
CurvesCurvature and
Related
ConceptsParametrization with
Respect to Arc
Length
CurvatureOsculating Planes
and Circles

Examples

Osculating Planes and Circles

Taylor's Formula continues to hold for curves, except for the Lagrange remainder term (which would involve different values in $(t_0, t_0 + h)$ for different coordinate functions):

Taylor's Formula for Curves

If $f: I \rightarrow \mathbb{R}^n$ is a C^r -curve and $t_0 \in I$, we have

$$f(t_0 + h) = \sum_{i=0}^r \frac{h^i}{i!} f^{(i)}(t_0) + o(h^r) \quad \text{for } h \rightarrow 0.$$

(Little-o notation is explained on the next slide.)

We will use the quadratic approximation ($r = 2$)

$$f(t_0 + h) = f(t_0) + hf'(t_0) + \frac{h^2}{2} f''(t_0) + o(h^2) \quad \text{for } h \rightarrow 0.$$

Big-O and little-o Notation

Introduced by P. BACHMANN and E. LANDAU

First we state the definition for real-valued functions.

Definition

Suppose $f, g: I \rightarrow \mathbb{R}$ are functions and $x_0 \in I$.

- (i) We say that $f(x) = O(g(x))$ for $x \rightarrow x_0$ (read “ $f(x)$ is big-oh of $g(x)$ ”) if there are positive constants C and δ such that

$$|f(x)| \leq C |g(x)| \quad \text{whenever } x \in I \text{ and } 0 < |x - x_0| < \delta.$$

- (ii) We say that $f(x) = o(g(x))$ for $x \rightarrow x_0$ (read “ $f(x)$ is little-oh of $g(x)$ ”) if for every $\epsilon > 0$ there exists $\delta = \delta(\epsilon) > 0$ such that

$$|f(x)| \leq \epsilon |g(x)| \quad \text{whenever } x \in I \text{ and } 0 < |x - x_0| < \delta.$$

Notes

- The definition is extended to curves in the usual way, i.e., by changing x to t and interpreting $|f(t)|$ as the Euclidean length of the vector $f(t)$.
- The quadratic Taylor formula for curves can also be stated as

$$\lim_{h \rightarrow 0} \frac{1}{h^2} \left(f(t_0 + h) - f(t_0) - hf'(t_0) - \frac{h^2}{2} f''(t_0) \right) = 0$$

and says that $f(t_0 + h)$ is approximated by the curve $f(t_0) + hf'(t_0) + \frac{h^2}{2} f''(t_0)$ with an error that is substantially smaller than h^2 .

- In Discrete Mathematics related asymptotic notions for functions with domain \mathbb{N} are considered. These are used to analyze the complexity of algorithms.

Definition

Suppose $f: I \rightarrow \mathbb{R}^n$ is a C^2 -curve and $t_0 \in I$ is such that $f'(t_0), f''(t_0)$ are linearly independent; equivalently, $f'(t_0) \neq \mathbf{0}$ and $\kappa(t_0) \neq 0$.

- The plane

$$f(t_0) + \mathbb{R} f'(t_0) + \mathbb{R} f''(t_0) = f(t_0) + \mathbb{R} \mathbf{T}(t_0) + \mathbb{R} \mathbf{N}(t_0),$$

which has $\mathbf{B}(t_0)$ as normal vector, is called *osculating plane* of f at t_0 . (The planes through $f(t_0)$ with normal vectors $\mathbf{T}(t_0)$ and $\mathbf{N}(t_0)$ are called *normal plane*, resp., *rectifying plane*.)

- The circle in the osculating plane with center $f(t_0) + \frac{1}{\kappa(t_0)} \mathbf{N}(t_0)$ and radius $\frac{1}{\kappa(t_0)}$ is called *osculating circle* of f at t_0 .
- Suppose that $f'(t), f''(t)$ are linearly independent for all $t \in I$. The curve $e(t) = f(t) + \frac{1}{\kappa(t)} \mathbf{N}(t)$, $t \in I$, which describes the movement of the center of the osculating circle of f , is called *evolute* of f .

Introduction

Limits and
ContinuityLimits of Sequences
of PointsLimits and Continuity
for CurvesTopology of Subsets
of \mathbb{R}^n Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization

Smooth Parametric
CurvesCurvature and
Related
ConceptsParametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

The name “osculating plane” for
 $E = f(t_0) + \mathbb{R}f'(t_0) + \mathbb{R}f''(t_0)$ is justified by

$$f(t_0 + h) = f(t_0) + \underbrace{hf'(t_0) + \frac{h^2}{2}f''(t_0)}_{\in E} + o(h^2) \quad \text{for } h \rightarrow 0.$$

The name “osculating circle” is justified by the following

Theorem

Under the assumptions of the definition, if both f and its osculating circle c at t_0 are (re-)parametrized with respect to arc length measured from $f(t_0)$, i.e., $g(s) = f(t(s))$ with $s(t) = \int_{t_0}^t |f'(\tau)| d\tau$ and similarly for c , we have

$$g(s) = c(s) + o(s^2) \text{ for } s \downarrow 0.$$

Rewriting this as $\lim_{s \downarrow 0} \frac{|g(s) - c(s)|}{s^2} = 0$ shows that the points at distance s (measured along the arc) from $f(t_0)$ on the curve and on its osculating circle have distance $\ll s^2$ when $s \downarrow 0$.

Proof cont'd.

The reparametrization of f is given by

$$\begin{aligned}g(s) &= f(t(s)) = g(0) + s g'(0) + \frac{s^2}{2} g''(0) + o(s^2) \\&= f(t_0) + s \mathbf{T}(t_0) + \frac{s^2}{2} \frac{\mathbf{T}'(t_0)}{|f'(t_0)|} + o(s^2) \\&= f(t_0) + s \mathbf{T}(t_0) + \frac{s^2}{2} \kappa(t_0) \mathbf{N}(t_0) + o(s^2)\end{aligned}$$

The arc-length parametrization of a circle of radius r is

$s \mapsto (r \cos(s/r), r \sin(s/r))$ (to see this, check the length of the derivative!).

$$\implies c(s) = f(t_0) + \frac{1}{\kappa(t_0)} \mathbf{N}(t_0) - \frac{\cos(\kappa(t_0)s)}{\kappa(t_0)} \mathbf{N}(t_0) + \frac{\sin(\kappa(t_0)s)}{\kappa(t_0)} \mathbf{T}(t_0)$$

Computing the derivatives up to order 2 shows $c(0) = f(t_0)$, $c'(0) = \mathbf{T}(t_0)$, $c''(0) = \kappa(t_0) \mathbf{N}(t_0)$, and hence that the quadratic Taylor polynomials of g and c are the same. □

Example

For the standard helix $f(t) = (\cos t, \sin t, t)$ we have

$$\mathbf{T}(t) = \frac{1}{\sqrt{2}} \begin{pmatrix} -\sin t \\ \cos t \\ 1 \end{pmatrix}, \quad \mathbf{N}(t) = \begin{pmatrix} -\cos t \\ -\sin t \\ 0 \end{pmatrix}, \quad \mathbf{B}(t) = \frac{1}{\sqrt{2}} \begin{pmatrix} \sin t \\ -\cos t \\ 1 \end{pmatrix}$$

The osculating plane at $t = 0$ has equation $-y + z = 0$ and parametric form $\mathbb{R}(0, 1, 1) + \mathbb{R}(1, 0, 0)$ (since it passes through the origin).

The curvature of the standard helix is

$$\kappa(t) = \frac{|\mathbf{T}'(t)|}{|f'(t)|} = \frac{\left| \frac{1}{\sqrt{2}}(-\cos t, -\sin t, 0) \right|}{\sqrt{2}} = \frac{1}{2},$$

and thus is constant.

The osculating circle at $t = 0$ has center

$f(0) + 2\mathbf{N}(0) = (-1, 0, 0)$, radius $\frac{1}{\kappa(0)} = 2$ and lies in the plane $y = z$. A suitable parametrization is

$$\begin{aligned} \mathbf{c}(t) &= (-1, 0, 0) - 2 \cos t \mathbf{N}(0) + 2 \sin t \mathbf{T}(0) \\ &= (-1 + 2 \cos t, \sqrt{2} \sin t, \sqrt{2} \sin t), \quad t \in [0, 2\pi]. \end{aligned}$$

Example

We determine the osculating circle of the plane parabola with equation $y = x^2$ in $(0, 0)$.

Viewing the parabola as graph of the function $p(x) = x^2$ and parametrizing it as the space curve $f(x) = (x, p(x), 0)$, we obtain

$$f'(x) = \begin{pmatrix} 1 \\ p'(x) \\ 0 \end{pmatrix}, \quad f''(x) = \begin{pmatrix} 0 \\ p''(x) \\ 0 \end{pmatrix}, \quad f'(x) \times f''(x) = \begin{pmatrix} 0 \\ 0 \\ p''(x) \end{pmatrix},$$

and hence

$$\kappa(x) = \frac{|f'(x) \times f''(x)|}{|f'(x)|^3} = \frac{|p''(x)|}{(1 + (p'(x))^2)^{3/2}}.$$

This formula holds for graphs of arbitrary C^2 -functions.

In the case under consideration we have $p'(0) = 0$, $p''(0) = 2$.
 $\Rightarrow \kappa(0) = 2$ and the osculating circle in $(0, 0)$ has equation

$$x^2 + (y - \frac{1}{2})^2 = \frac{1}{4}.$$

Finally we determine the evolute of the parabola:

$$\kappa(x) = \frac{2}{\sqrt{1+4x^2}^3},$$

$$\mathbf{T}(x) = \frac{1}{\sqrt{1+4x^2}} \begin{pmatrix} 1 \\ 2x \end{pmatrix},$$

$$\mathbf{N}(x) = \frac{1}{\sqrt{1+4x^2}} \begin{pmatrix} -2x \\ 1 \end{pmatrix}, \quad (\text{since } x \mapsto x^2 \text{ is convex})$$

$$E(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix} + \frac{1+4x^2}{2} \begin{pmatrix} -2x \\ 1 \end{pmatrix} = \begin{pmatrix} -4x^3 \\ \frac{1}{2} + 3x^2 \end{pmatrix}.$$

The substitution $x' = \sqrt{3}x$ shows that $E(x)$ has range

$$\left\{ \left(ax'^3, \frac{1}{2} + x'^2 \right); x' \in \mathbb{R} \right\} \quad \text{with } a = -4/\sqrt{3}^3.$$

Scaling by a^2 and setting $t = ax'$ reveals that the non-parametric evolute is a rotated, stretched and shifted version of Neile's semicubic parabola.

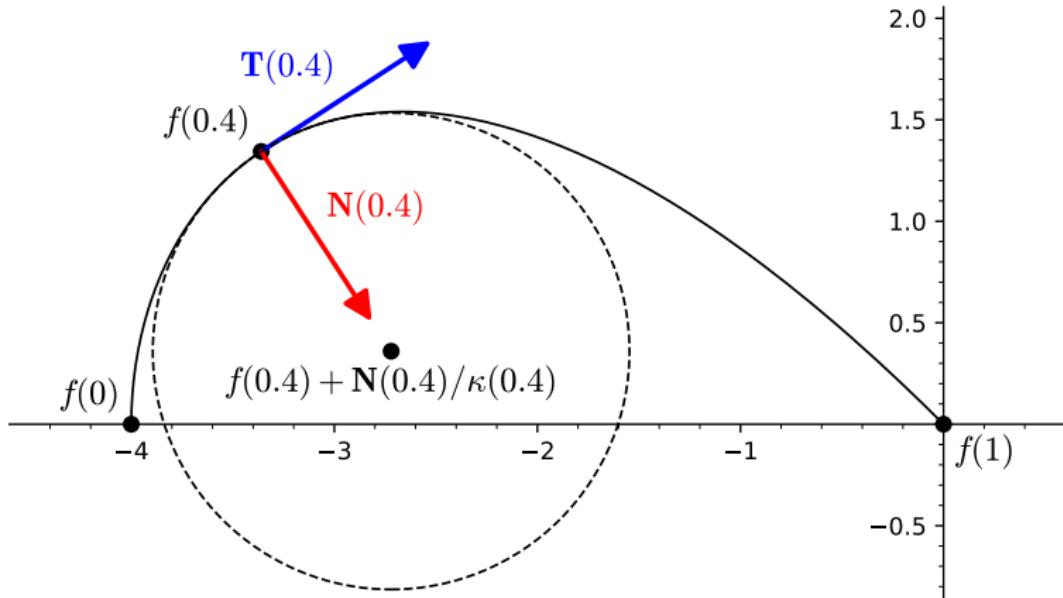


Figure: The parabola $y = x^2$, its evolute, and the osculating circles to the parabola in $(0, 0)$ and $(\frac{1}{2}, \frac{1}{4})$

Introduction

Limits and
Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation
and
Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and
Related
Concepts

Parametrization with
Respect to Arc
Length

Curvature

Osculating Planes
and Circles

Examples

Question

How to decide whether a space curve $f: I \rightarrow \mathbb{R}^3$ is planar (i.e., contained in a plane $E = \mathbf{a} + U$)?

Answer

If $f(t) \in E$ for all $t \in I$ then $\frac{1}{h}(f(t+h) - f(t)) \in U$ for all t, h such that $t, t+h \in I$.

$\Rightarrow f'(t) \in U$ for all $t \in I$, since U is closed (assuming that f is differentiable).

Iterating the argument, we obtain $f''(t) \in U$ for all $t \in I$ (and similarly for all higher derivatives).

$\Rightarrow E$ is the osculating plane of f at every point $f(t)$ with $\kappa(t) \neq 0$.

Conversely, it is easy to see that a curve with the same osculating plane at every point must be contained in that plane.

Introduction

Limits and Continuity

Limits of Sequences
of Points

Limits and Continuity
for Curves

Topology of Subsets
of \mathbb{R}^n

Differentiation and Integration

Arc Length

Remarks on Uniform
Continuity

Reparametrization
Smooth Parametric
Curves

Curvature and Related Concepts

Parametrization with
Respect to Arc
Length

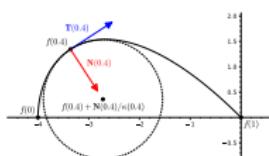
Curvature

Osculating Planes
and Circles

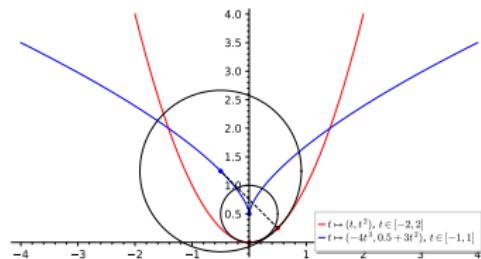
Examples

Question

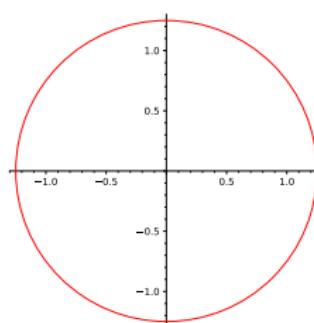
Where is the curvature largest?



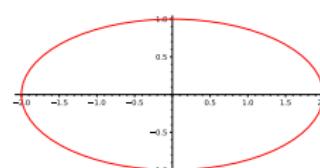
(a) circle



(b) ellipse



(c) parabola



(d) hyperbola

Answer

For a circle (constant curvature) nowhere, for an ellipse at the two vertices on the semi-major axis, and for a parabola/hyperbola at the vertex (intersection point with the axis).

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

Today's Lecture: Applications to Physics

The Physics of Space Curves

Model

The motion of an object (particle, celestial body, etc.) in space is described, relative to some coordinate system and after fixing the units of measurement, by a space curve $I \rightarrow \mathbb{R}^3$, $t \mapsto \mathbf{r}(t)$. The vector $\mathbf{r}(t)$ can be thought of as *position vector* (or *radius vector*) of the object relative to the origin at time t .

Units of measurement

A (scalar) physical quantity is described by a value and a corresponding unit of measurement (e.g., a velocity of 10 ms^{-1}).

Changing between different units of measurement amounts to rescaling the values by a constant (e.g., $10 \text{ ms}^{-1} = \frac{10 \cdot 3600}{1000} = \text{km h}^{-1} = 36 \text{ km h}^{-1}$) and is compatible with differentiation/integration. For this reason, the units of measurement are usually ignored in Mathematics (e.g., we speak of “the distance $d \in \mathbb{R}$ traveled per unit time”).

Coordinate systems

Here the situation is more subtle. A model sufficient for us is the following: Admissible coordinate systems in \mathbb{R}^3 (relative to the standard coordinate system) consist of a distinguished point \mathbf{b} ("origin") and 3 pairwise orthogonal unit vectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ (*cartesian coordinate system*). The coordinate change is afforded by the linear map

$$\mathbb{R}^3 \rightarrow \mathbb{R}^3, \mathbf{x}' \mapsto \mathbf{U}\mathbf{x}' + \mathbf{b} = \mathbf{x},$$

where $\mathbf{U} \in \mathbb{R}^{3 \times 3}$ denotes the matrix with columns $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$.

Definition

A matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ is said to be *orthogonal* if $\mathbf{U}^T \mathbf{U} = \mathbf{I}_n$.

Since the entries of $\mathbf{U}^T \mathbf{U}$ are the pairwise dot products of the columns $\mathbf{u}_1, \dots, \mathbf{u}_n$ of \mathbf{U} , this is equivalent to

$$|\mathbf{u}_j| = 1 \text{ for } 1 \leq j \leq n \quad \text{and} \quad \mathbf{u}_i \cdot \mathbf{u}_j = 0 \text{ for } i \neq j.$$

Such vectors are said to form an *orthonormal basis* of \mathbb{R}^n .

Examples

Examples of 3×3 orthogonal matrices are \mathbf{I}_3 ,

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \mp \sin \phi \\ 0 & \pm \sin \phi & \cos \phi \end{pmatrix}, \quad \frac{1}{3} \begin{pmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{pmatrix}.$$

Properties

Let $\mathbf{U} \in \mathbb{R}^{n \times n}$ be orthogonal.

① $|\mathbf{U}\mathbf{x}| = |\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^n$

As a consequence, the coordinate change $\mathbf{x}' = \mathbf{U}\mathbf{x} + \mathbf{b}$ preserves distances (and also orthogonality and angles):

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= |\mathbf{x} - \mathbf{y}| = |\mathbf{U}\mathbf{x}' + \mathbf{b} - (\mathbf{U}\mathbf{y}' + \mathbf{b})| \\ &= |\mathbf{U}(\mathbf{x}' - \mathbf{y}')| = |\mathbf{x}' - \mathbf{y}'| = d(\mathbf{x}', \mathbf{y}'), \end{aligned}$$

② $\mathbf{U}\mathbf{U}^T = \mathbf{I}_n$

As a consequence, the rows of an orthogonal matrix form an orthonormal basis of \mathbb{R}^n as well (and $\mathbf{U}^T = \mathbf{U}^{-1}$ is orthogonal).

Properties cont'd

③ $\det(\mathbf{U}) = \pm 1$

Proofs.

(1) We use that the dot product of column vectors can be written as $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$.

$$\implies |\mathbf{Ux}|^2 = (\mathbf{Ux})^T \mathbf{Ux} = \mathbf{x}^T \mathbf{U}^T \mathbf{Ux} = \mathbf{x}^T \mathbf{x} = |\mathbf{x}|^2,$$

using $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$ for matrices \mathbf{A}, \mathbf{B} such that \mathbf{AB} exists.

(2) follows from the fact that $\mathbf{AB} = \mathbf{I}_n$ for $n \times n$ matrices \mathbf{A}, \mathbf{B} implies $\mathbf{BA} = \mathbf{I}_n$ and $\mathbf{B} = \mathbf{A}^{-1}$.

(3) is obtained from

$$1 = \det(\mathbf{I}_n) = \det(\mathbf{U}^T \mathbf{U}) = \det(\mathbf{U}^T) \det(\mathbf{U}) = \det(\mathbf{U})^2,$$

which employs the multiplicativity $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$ of the determinant for $n \times n$ matrices. □

Further Notes

- It can be shown that every *isometry* (i.e., distance-preserving map) $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ has the form $f(\mathbf{x}) = \mathbf{U}\mathbf{x} + \mathbf{b}$ for some orthogonal matrix $\mathbf{U} \in \mathbb{R}^{n \times n}$ and some vector $\mathbf{b} \in \mathbb{R}^n$.
- The orthogonal matrices in $\mathbb{R}^{2 \times 2}$ are precisely the rotation and reflection matrices $R(\phi)$, $S(\phi)$ of Worksheet 3, Exercise W11.
- The orthogonal matrices in $\mathbb{R}^{3 \times 3}$ determine either rotations with axes through the origin, or reflections at a plane containing the origin, or some simple combinations thereof.
- It is possible to extend the orthonormalization process for two vectors, that we have described, to any basis of \mathbb{R}^n (*Gram-Schmidt orthogonalization*). Writing the starting basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ as a matrix $\mathbf{A} = (\mathbf{v}_1 | \dots | \mathbf{v}_n)$ and the computed orthonormal basis $\mathbf{u}_1, \dots, \mathbf{u}_n$ as $\mathbf{Q} = (\mathbf{u}_1 | \dots | \mathbf{u}_n)$, this corresponds to a factorization

$$\mathbf{A} = \mathbf{QR} \quad \text{with } \mathbf{Q} \text{ orthogonal, } \mathbf{R} \text{ upper triangular,}$$

and such that the elements r_{ii} on the main diagonal of \mathbf{R} are positive (*QR-factorization of \mathbf{A}*).

Classical Mechanics

Suppose $I \rightarrow \mathbb{R}^3$, $t \mapsto \mathbf{r}(t)$ describes the motion of an object.

(vectorial) velocity $\mathbf{v}(t) = \lim_{h \rightarrow 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h} = \mathbf{r}'(t)$
speed (scalar velocity) $v(t) = |\mathbf{v}(t)| = |\mathbf{r}'(t)|$
acceleration $\mathbf{a}(t) = \mathbf{v}'(t) = \mathbf{r}''(t)$

provided, of course, that these quantities are defined.

Notes

- $\frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h}$ is the average displacement per unit time in $[t, t + h]$ and $\mathbf{v}(t)$ its limit for $h \rightarrow 0$.
- $\frac{|\mathbf{r}(t+h) - \mathbf{r}(t)|}{h}$ is the average distance per unit time traveled in $[t, t + h]$ and $v(t)$ its limit for $h \rightarrow 0$.
- The arc length formula $s(t) = \int_{t_0}^t |\mathbf{r}'(\tau)| d\tau = \int_{t_0}^t v(\tau) d\tau$ gives the distance traveled in $[t_0, t]$ as an integral over the speed, and conversely the speed as $v(t) = s'(t)$ (or, succinctly, $v = ds/dt$).

Notes cont'd

The Fundamental Theorem of Calculus (vectorial form) gives the

- Displacement/velocity in terms of velocity/acceleration:

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \int_{t_0}^t \mathbf{v}(\tau) d\tau, \quad \mathbf{v}(t) = \mathbf{v}(t_0) + \int_{t_0}^t \mathbf{a}(\tau) d\tau$$

- The acceleration is usually given by *Newton's Second Law of Motion*:

$$\mathbf{F}(t) = m \mathbf{a}(t) \quad (\text{force} = \text{mass} \times \text{acceleration})$$

Example

An object with mass m moves on a circular path with radius a and constant angular speed ω . Find the “driving force”.

In an appropriate coordinate system the motion is described by the planar curve

$$\mathbf{r}(t) = \begin{pmatrix} a \cos(\omega t) \\ a \sin(\omega t) \end{pmatrix}$$

Differentiating twice gives

$$\mathbf{v}(t) = \begin{pmatrix} -a\omega \sin(\omega t) \\ a\omega \cos(\omega t) \end{pmatrix}, \quad \mathbf{a}(t) = \begin{pmatrix} -a\omega^2 \cos(\omega t) \\ -a\omega^2 \sin(\omega t) \end{pmatrix},$$

and Newton's 2nd Law shows

$$\mathbf{F}(t) = m\mathbf{a}(t) = \begin{pmatrix} -am\omega^2 \cos(\omega t) \\ -am\omega^2 \sin(\omega t) \end{pmatrix} = -m\omega^2 \mathbf{r}(t).$$

\implies The acting force is directed towards the center of the circle (*centripetal force*).

Example

A ballistic rocket is fired with angle of elevation α and muzzle speed v_0 . Compute the orbit of the rocket and determine the angle α that maximizes the range.

We assume that the earth's surface is flat and the only acting force is the gravitational force (no air resistance). Then the orbit is contained in the vertical plane determined by the launch pad and its direction. Choosing appropriate coordinates in E gives

$$\mathbf{F} = m\mathbf{a} = mg(0, -1), \quad (g \approx 9.81 \text{ ms}^{-2})$$

$$\mathbf{a} = g(0, -1) = (0, -g),$$

$$\mathbf{v}(t) = \mathbf{v}_0 + \int_0^t (0, -g) d\tau = \mathbf{v}_0 + (0, -gt),$$

$$\mathbf{r}(t) = \mathbf{r}_0 + \int_0^t \mathbf{v}_0 + (0, -gt) d\tau = \mathbf{r}_0 + t\mathbf{v}_0 + (0, -\frac{1}{2}gt^2).$$

Substituting $\mathbf{v}_0 = (v_0 \cos \alpha, v_0 \sin \alpha)$ and $\mathbf{r}_0 = (0, 0)$ gives the position of the rocket as

$$\mathbf{r}(t) = ((v_0 \cos \alpha)t, (v_0 \sin \alpha)t - \frac{1}{2}gt^2).$$

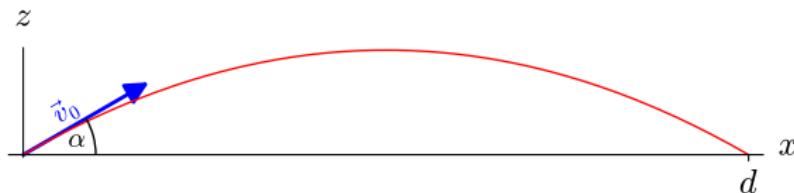


Figure: Illustration of ballistic rocket

Example (con't)

The rocket touches down at time t_1 satisfying $(v_0 \sin \alpha)t_1 - \frac{1}{2}gt_1^2 = 0$, i.e., $t_1 = 2v_0 \sin \alpha / g$, and at distance

$$d = (v_0 \cos \alpha)t_1 = \frac{2v_0^2 \sin \alpha \cos \alpha}{g} = \frac{v_0^2 \sin(2\alpha)}{g}$$

to the launch pad.

The range d is maximized by choosing $\alpha = \pi/4$, and $d_{\max} = v_0^2/g$.

T, N, κ Reexamined

In what follows the parameter t is omitted, e.g., we write $v = v(t)$.
The unit tangent vector has the following physical interpretation:

$$\mathbf{T} = \frac{\mathbf{r}'}{|\mathbf{r}'|} = \frac{\mathbf{v}}{v} \iff \mathbf{v} = v\mathbf{T}$$

Applying the product rule, this gives

$$\mathbf{a} = \mathbf{v}' = v'\mathbf{T} + v\mathbf{T}' = v'\mathbf{T} + v|\mathbf{T}'|\mathbf{N} = v'\mathbf{T} + \kappa v^2\mathbf{N},$$

since $\kappa = |\mathbf{T}'| / |\mathbf{r}'| = |\mathbf{T}'| / v$.

\implies The coordinate vector of \mathbf{a} with respect to the TNB frame $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is $(v', \kappa v^2, 0)$.

Caution

Since everything depends on t , this should rather be written as

$$\mathbf{a}(t) = v'(t)\mathbf{T}(t) + \kappa(t)v(t)^2\mathbf{N}(t) + 0\mathbf{B}(t).$$

Observation

The acceleration $\mathbf{a} = \mathbf{a}(t)$ always lies in the direction space of the osculating plane of the object's curve. (If you view \mathbf{a} as attached to the curve, i.e. $\mathbf{r} + \mathbf{a}$, it lies in the osculating plane.)

Sometimes it is convenient to have formulas for the coordinates of \mathbf{a} with respect to \mathbf{T}, \mathbf{N} solely in terms of vectorial quantities:

$$v' = \mathbf{a} \cdot \mathbf{T} = \frac{\mathbf{a} \cdot \mathbf{v}}{|\mathbf{v}|} = \frac{\mathbf{r}' \cdot \mathbf{r}''}{|\mathbf{r}'|},$$

$$\kappa v^2 = \frac{|\mathbf{v} \times \mathbf{a}|}{v^3} \cdot v^2 = \frac{|\mathbf{v} \times \mathbf{a}|}{v} = \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|}.$$

$$\Rightarrow \mathbf{a} = \frac{\mathbf{r}' \cdot \mathbf{r}''}{|\mathbf{r}'|} \mathbf{T} + \frac{|\mathbf{r}' \times \mathbf{r}''|}{|\mathbf{r}'|} \mathbf{N}$$

KEPLER's Laws of Planetary Motion

- ① A planet revolves around the sun in an elliptical orbit with the sun at one focus.
- ② The line joining the sun to a planet sweeps out equal areas in equal times.
- ③ The square of the period of revolution of a planet is proportional to the cube of the length of the major axis of its orbit.

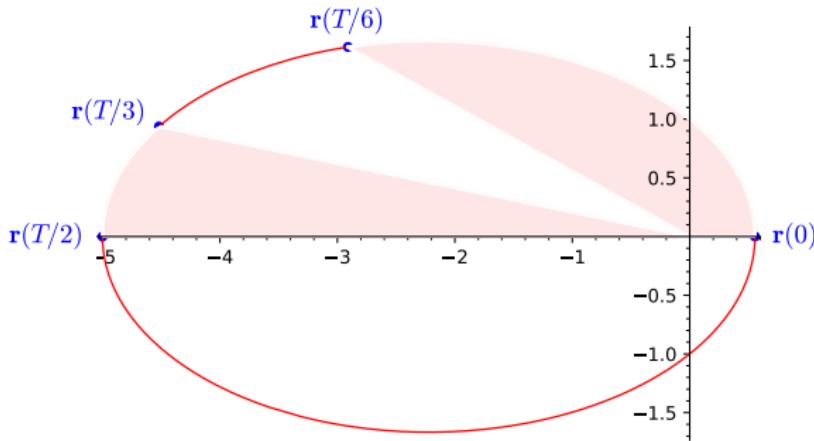


Figure: Illustration of Kepler's 2nd Law

Theorem (NEWTON 1687)

Kepler's Laws are consequences of

2nd Law of Motion $\mathbf{F} = m\mathbf{a}$

Law of Gravitation $\mathbf{F} = -\frac{GMm}{r^3}\mathbf{r} = -\frac{GMm}{r^2}\frac{\mathbf{r}}{r}$

Polar form of Conics

A *conic* (or *conic section*) is the solution set of an (inhomogeneous) quadratic equation in \mathbb{R}^2 , i.e.,

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \quad (A, B, C) \neq (0, 0, 0).$$

Apart from degenerate cases, every conic can be transformed into one of three canonical forms, viz.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (a \geq b > 0) \quad \text{if} \quad B^2 - 4AC < 0 \quad (\text{ellipse})$$

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \quad (a, b > 0) \quad \text{if} \quad B^2 - 4AC > 0 \quad (\text{hyperbola})$$

$$y^2 + 4cx = 0 \quad (a > 0) \quad \text{if} \quad B^2 - 4AC = 0 \quad (\text{parabola})$$

Here a, b denote the semi-axes in the case of an ellipse or hyperbola, and $c > 0$ denotes the focal length (distance between vertex and focus) in the case of a parabola. A circle is a special case of an ellipse (the case $a = b$, in which the two semi-axes reduce to the radius of the circle).

Matrix form

The equation of a conic can be written as

$$\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} D \\ E \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix} + F = 0,$$

i.e., as $\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{b}^T \mathbf{x} + c = 0$, where now $\mathbf{x} = (x_1, x_2)^T$,
 $\mathbf{A} = \begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} = \mathbf{A}^T$, $\mathbf{b} = (D, E)^T$, $c = F$.

Choosing \mathbf{A} as a symmetric matrix makes the representation unique and facilitates the transformation into canonical form.

A comprehensive description of conics and their properties can be found in [Ste21], Sections 10.5 and 10.6, and in the Wikipedia http://en.wikipedia.org/wiki/Conic_section; cf. also the subsequent exercise and Homework 4, Exercise H 25.

Non-degenerate conics can be defined as the set of points in the plane \mathbb{R}^2 , for which the ratio of the distances to a point P (*focus*) and a line L (*directrix*) with $P \notin L$ is a fixed constant $e > 0$.

Further parameters

eccentricity e The constant ratio $\frac{d(\mathbf{x}, P)}{d(\mathbf{x}, L)}$, where P denotes a focus and L the directrix of the conic.

focal parameter p The distance between P and L .

semi-latus rectum l The distance between P and one of the two intersection points of the conic and the line through P parallel to L (note that $e = l/p$).

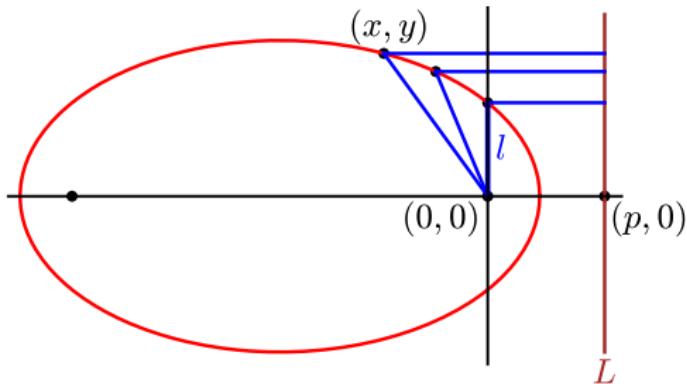


Figure: Ellipse with $P = (0, 0)$ and $e = 0.8$

Remark

There are several—sometimes conflicting—notations in use for the parameters of a non-degenerate conic. For example, the distance from the center to a focus ($\sqrt{a^2 - b^2}$ for an ellipse, $\sqrt{a^2 + b^2}$ for a hyperbola) is sometimes referred to as (“linear”) “eccentricity” and the term “numerical eccentricity” used for what we call eccentricity. The relation between the two quantities is given by

$$e = \frac{\sqrt{a^2 \mp b^2}}{a} = \sqrt{1 \mp \frac{b^2}{a^2}}.$$

The semilatus rectum is sometimes called “form parameter” of the ellipse/hyperbola and denoted by p , conflicting with our notation for the focal parameter. In terms of a, b the semilatus rectum is given as $p = b^2/a$. It is also equal to the radius of the smallest osculating circle of the ellipse/hyperbola, which is obtained in the perihelion/aphelion; cf. Exercise 51 on Worksheet 6. Finally, the focal parameter in terms of a, b is given as

$$p = \frac{b^2}{ae} = \frac{b^2}{\sqrt{a^2 \mp b^2}}.$$

All these properties are easily deduced from the formulas expressing a, b in terms of e, p ; cf. the subsequent exercise.

Proposition

The curve given in polar coordinates as

$$r = \frac{l}{1 + e \cos \theta} \quad \text{with } l, e > 0$$

parametrizes an ellipse if $e < 1$, a branch of a hyperbola if $e > 1$, and a parabola if $e = 1$. Moreover, $(0, 0)$ is a focus of the conic, e gives the eccentricity of the conic and l its semi-latus rectum.

Proof.

We use a coordinate system, in which the focus is at $(0, 0)$ and the directrix is the vertical line $x = p > 0$; see picture.

⇒ The equation of the conic is $\sqrt{x^2 + y^2} = e|p - x|$,
in polar coordinates

$$r = \pm e(p - r \cos \theta) \iff r = \frac{\pm l}{1 \pm e \cos \theta}.$$

Case 1: $e < 1$. Since $|e \cos \theta| < 1$, the signs must be $+/-$ and the equation parametrizes an ellipse with perihelion in $\left(\frac{l}{1+e}, 0\right)$ and aphelion in $\left(\frac{l}{1-e}, 0\right)$.

Proof cont'd.

Case 2: $e = 1$. The signs are again $+/+$ but $\theta = \pi$ is excluded.

This describes a parabola with vertex in $\left(\frac{l}{1+e}, 0\right)$ and open on the left side.

Case 3: $e > 1$. Here both sign combinations $+/+$ and $-/-$ are possible. The equation $r = \frac{l}{1+e \cos \theta}$ yields a curve that is defined for $\cos \theta > -1/e$ (i.e., $|\theta| < \arccos(-1/e)$), which must be the left branch of a hyperbola with asymptotes of slope $\pm\sqrt{e^2 - 1}$ (using $e = \sqrt{1 + \frac{b^2}{a^2}}$ and solving for b/a). The corresponding right branch has equation $r = \frac{-l}{1-e \cos \theta}$ and is defined for $\cos \theta > 1/e$ (i.e., $|\theta| < \arccos(1/e)$).

Since $r(\theta) = \frac{l}{1+e \cos \theta}$ satisfies $r(\pi/2) = l$, the semi-latus rectum equals l in all three cases. □

Exercise

- ① Show that the canonical form of a parabola with focal parameter p is $y^2 + 2px = 0$; the vertex and focus of the canonical parabola are at $(0, 0)$ and $(-p/2, 0)$, respectively.
- ② Show that an ellipse with eccentricity e and focal parameter p has semi-axes $a = \frac{ep}{1-e^2}$, $b = \frac{ep}{\sqrt{1-e^2}}$; in the focus-directrix form the center is at $\left(-\frac{e^2 p}{1-e^2}, 0\right)$.
- ③ Show that a hyperbola with eccentricity e and focal parameter p has semi-axes $a = \frac{ep}{e^2-1}$, $b = \frac{ep}{\sqrt{e^2-1}}$; in the focus-directrix form the center is at $\left(\frac{e^2 p}{e^2-1}, 0\right)$ (and thus in particular to the right of the directrix).
- ④ Describe the geometric meaning of the rectangle with vertices $(\pm a, \pm b)$ for an ellipse/hyperbola in canonical form. Which eccentricities correspond to the square case $a = b$?
- ⑤ Describe the shapes of the ellipses/hyperbolas in the limiting cases $e \downarrow 0$, $e \uparrow 1$, $e \downarrow 1$, $e \rightarrow \infty$.

Hint: In (1)–(3) start with the focus-directrix form $x^2 + y^2 = e^2(x - p)^2$ and eliminate the term linear in x .

Proof of the 1st Kepler Law.

First we show that the orbit must be planar.

Arrange coordinates in such a way that the center of the sun is located in the origin $(0, 0, 0)$. The position vector of the planet is denoted by $\mathbf{r} = \mathbf{r}(t)$.

Newton's Laws imply

$$\mathbf{a} = -\frac{GM}{r^3} \mathbf{r} \implies \mathbf{r} \times \mathbf{a} = \mathbf{0}$$

$$\implies \frac{d}{dt}(\mathbf{r} \times \mathbf{v}) = \mathbf{v} \times \mathbf{v} + \mathbf{r} \times \mathbf{a} = \mathbf{0} \implies \mathbf{r} \times \mathbf{v} = \mathbf{h} \text{ is a constant}$$

(This is an instance of the *law of conservation of angular momentum*: If the *torque* $\mathbf{r} \times \mathbf{F} = \mathbf{r} \times (m\mathbf{a}) = m(\mathbf{r} \times \mathbf{a})$ of a particle of constant mass m is identically zero, its *angular momentum* $\mathbf{r} \times (m\mathbf{v}) = m(\mathbf{r} \times \mathbf{v})$ must be constant.)

For planets we have $\mathbf{h} \neq \mathbf{0}$, i.e., \mathbf{r} and \mathbf{v} are linearly independent. (Otherwise the planet's motion would be restricted to a line, viz.

$\mathbb{R}\mathbf{r}(t_0)$ with t_0 chosen arbitrarily.)

$\implies t \mapsto \mathbf{r}(t)$ lies entirely in the plane E through $\mathbf{0}$ with normal vector \mathbf{h} .

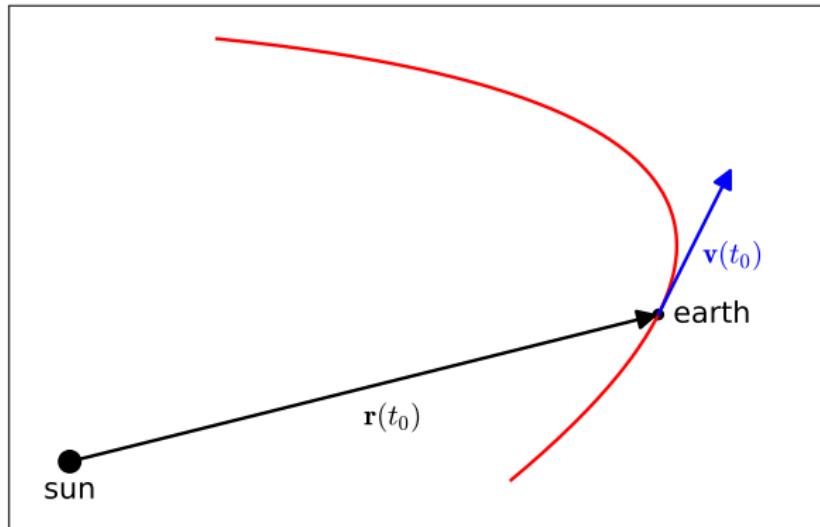


Figure: The basic setting of the Kepler Laws

If you are interested how this picture was drawn: I took $\mathbf{r}(t_0) = (4, 1)$, $\mathbf{v}(t_0) = \mathbf{r}'(t_0) = (0.5, 1)$, and the orbit as

$$\mathbf{r}(t) = \mathbf{r}(0) + t \mathbf{r}'(0) + \frac{t^2}{2} \mathbf{w}, \quad t \in [a, b],$$

with \mathbf{w} and $[a, b]$ chosen in such a way that the picture looks “nice”.

Proof cont'd.

The vector $\mathbf{u}_3 = \frac{\mathbf{h}}{|\mathbf{h}|}$ will serve as the 3rd unit vector of our coordinate system, with $\mathbf{u}_1, \mathbf{u}_2 \in E$ yet to be determined.

Further we have, writing $\mathbf{u} = \mathbf{r}/|\mathbf{r}| = \mathbf{r}/r$,

$$\begin{aligned}\mathbf{h} &= \mathbf{r} \times \mathbf{v} = \mathbf{r} \times \mathbf{r}' = r\mathbf{u} \times (r\mathbf{u})' \\ &= r\mathbf{u} \times r\mathbf{u}' + r\mathbf{u} \times r'\mathbf{u} = r^2(\mathbf{u} \times \mathbf{u}')\end{aligned}$$

$$\begin{aligned}\mathbf{a} \times \mathbf{h} &= \left(\frac{-GM}{r^2} \mathbf{u} \right) \times (r^2(\mathbf{u} \times \mathbf{u}')) = -GM\mathbf{u} \times (\mathbf{u} \times \mathbf{u}') \\ &= -GM((\mathbf{u} \cdot \mathbf{u}')\mathbf{u} - (\mathbf{u} \cdot \mathbf{u})\mathbf{u}') \\ &= GM\mathbf{u}'\end{aligned}$$

$$(\mathbf{v} \times \mathbf{h})' = \mathbf{v}' \times \mathbf{h} = \mathbf{a} \times \mathbf{h} = GM\mathbf{u}'$$

$$\implies \mathbf{v} \times \mathbf{h} = GM\mathbf{u} + \mathbf{c} \quad \text{with } \mathbf{c} \text{ a constant vector in the plane } E = \mathbf{h}^\perp.$$

Proof cont'd.

Case 1: $\mathbf{c} = \mathbf{0}$.

In this case $\mathbf{r} = r\mathbf{u}$ is orthogonal to \mathbf{v} , which implies

$$\frac{d}{dt} |\mathbf{r}|^2 = 2\mathbf{r} \cdot \mathbf{v} = 0.$$

\Rightarrow The planet moves on the circle in E of (constant) radius r around the origin.

Case 2: $\mathbf{c} \neq \mathbf{0}$.

In this case the vectors $\mathbf{u}_1 = \frac{\mathbf{c}}{|\mathbf{c}|}$, $\mathbf{u}_2 = \mathbf{u}_3 \times \mathbf{u}_1 = \frac{\mathbf{h} \times \mathbf{c}}{|\mathbf{h}||\mathbf{c}|}$, and \mathbf{u}_3 form a positively oriented orthonormal basis of \mathbb{R}^3 , and we can use polar coordinates in the plane $E = \mathbb{R}\mathbf{r} + \mathbb{R}\mathbf{v} = \mathbb{R}\mathbf{u}_1 + \mathbb{R}\mathbf{u}_2$ to describe the planet's motion:

$$\mathbf{r}(t) = r(t) \cos \theta(t) \mathbf{u}_1 + r(t) \sin \theta(t) \mathbf{u}_2.$$

Then $\cos \theta = \frac{\mathbf{r} \cdot \mathbf{c}}{rc}$ with $c = |\mathbf{c}|$, but not necessarily $\theta \in [0, \pi]$ as in the definition of the angle between two vectors.

The angle $\theta(t)$ can be taken as a continuous function, which doesn't jump from 2π to 0 as $\mathbf{r}(t)$ crosses the x -axis, but instead varies from $2(k - 1)\pi$ to $2k\pi$ during the k -th revolution of the planet.

Proof cont'd.

Now we compute $\mathbf{r} \cdot (\mathbf{v} \times \mathbf{h})$ in two ways (writing $h = |\mathbf{h}|$):

$$\mathbf{r} \cdot (\mathbf{v} \times \mathbf{h}) = (\mathbf{r} \times \mathbf{v}) \cdot \mathbf{h} = \mathbf{h} \cdot \mathbf{h} = h^2,$$

$$\begin{aligned}\mathbf{r} \cdot (\mathbf{v} \times \mathbf{h}) &= \mathbf{r} \cdot (GM\mathbf{u} + \mathbf{c}) = GM\mathbf{r} \cdot \mathbf{u} + \mathbf{r} \cdot \mathbf{c} \\ &= GMr\mathbf{u} \cdot \mathbf{u} + |\mathbf{r}| |\mathbf{c}| \cos \theta \\ &= GMr + rc \cos \theta.\end{aligned}$$

$$\Rightarrow r = \frac{h^2}{GM + c \cos \theta} = \frac{h^2/GM}{1 + (c/GM) \cos \theta},$$

which is the polar equation of a conic with eccentricity $e = c/(GM)$ and semi-latus rectum $l = h^2/(GM)$.

Since the orbit of a planet is a closed curve, it must be an ellipse.

This completes the proof of Kepler's 1st Law.

The 2nd and 3rd of Kepler's Laws are considered in Homework 5, Exercise H28.

Notes

- ① The preceding proof doesn't reveal the fact that the parametric space curve describing the motion of a planet is periodic.

But at least we can conclude from $rv \geq |\mathbf{r} \times \mathbf{v}| = h$ and $r \leq \frac{l}{1-e}$ that $v \geq v_0 = \frac{h(1-e)}{l}$ and (because the arc length of an ellipse is finite) that $\mathbf{r}(t)$ must revolve around the origin in finite time.

In order to prove periodicity, one also needs that the vectorial velocity $\mathbf{v}(t_0 + T)$ after a full revolution of period T is equal to $\mathbf{v}(t_0)$. This can be concluded from the law of conservation of energy, which implies $v(t_0 + T) = v(t_0)$. An alternative proof of periodicity is implicit in the proof of Kepler's 2nd Law.

- ② The proof of Kepler's 1st Law shows that, more generally, the orbit of any object subject to a central force field (and no other forces) is contained in a non-degenerate conic of the plane $\mathbb{R}\mathbf{r} + \mathbb{R}\mathbf{v}$ (except for the case $\mathbf{h} = \mathbf{0}$). This applies in particular to space flights from the earth. An important question is to determine the type of orbit (ellipse, parabola, or hyperbola) from the initial position $\mathbf{r}_0 = \mathbf{r}(t_0)$ and velocity $\mathbf{v}_0 = \mathbf{v}(t_0)$.

Notes cont'd

② (cont'd)

The type of orbit is determined by the eccentricity $e = c/GM$.

$$\begin{aligned}\mathbf{c} &= \mathbf{v} \times \mathbf{h} - GM\mathbf{u} = \mathbf{v} \times (\mathbf{r} \times \mathbf{v}) - GM \frac{\mathbf{r}}{r} \\ &= v^2 \mathbf{r} - (\mathbf{r} \cdot \mathbf{v})\mathbf{v} - GM \frac{\mathbf{r}}{r} = \left(v^2 - \frac{GM}{r}\right) \mathbf{r} - (\mathbf{r} \cdot \mathbf{v})\mathbf{v}\end{aligned}$$

From this we see already that the orbit is circular iff

$$v_0 = \sqrt{GM/r_0} \text{ and } \mathbf{r}_0 \perp \mathbf{v}_0.$$

In general we have, denoting the angle between \mathbf{r} and \mathbf{v} by ϕ ,

$$\begin{aligned}c^2 &= \left(v^2 - \frac{GM}{r}\right)^2 r^2 - 2 \left(v^2 - \frac{GM}{r}\right) (\mathbf{r} \cdot \mathbf{v})^2 + (\mathbf{r} \cdot \mathbf{v})^2 v^2 \\ &= (rv^2 - GM)^2 - \left(v^2 - \frac{2GM}{r}\right) r^2 v^2 \cos^2 \phi \\ &= r^2 v^2 \left(v^2 - \frac{2GM}{r}\right) \sin^2 \phi + (GM)^2 \\ &\leqslant (GM)^2 \iff v^2 \leqslant \frac{2GM}{r}.\end{aligned}$$

Notes cont'd

② (cont'd)

From this we see that

$$e \begin{cases} < 1 & \text{if } v_0 < \sqrt{2GM/r_0}, \\ = 1 & \text{if } v_0 = \sqrt{2GM/r_0}, \\ > 1 & \text{if } v_0 > \sqrt{2GM/r_0}. \end{cases}$$

The quantity $\sqrt{2GM/r_0}$ is called the *parabolic velocity* (or *escape velocity*, since for $e \geq 1$ the orbit is unbounded and the object won't return to its initial position).

On the surface of the earth the escape velocity is about

$$\sqrt{\frac{2(6.67 \times 10^{-11} \text{ m}^3\text{kg}^{-1}\text{s}^{-2})(5.98 \times 10^{24} \text{ kg})}{6370 \text{ km}}} \approx 11,2 \text{ km/s.}$$

Exercise

In this exercise we suppose that the angle $\phi \in (0, \pi)$ between \mathbf{r}_0 and \mathbf{v}_0 is fixed.

- ① Show that the eccentricity e of the orbit is minimized by choosing v_0 as the “circular velocity” $\sqrt{GM/r_0}$ and that $e = \cos \phi$ in this case.
- ② Show that for any e satisfying $\cos \phi < e < 1$ there are precisely two initial velocities $v_0^{(1)} < \sqrt{GM/r_0} < v_0^{(2)}$ for which the resulting orbit is an ellipse with eccentricity e . Explain this phenomenon.
- ③ For $v \downarrow 0$ our formula for c^2 implies $e \rightarrow 1$. Does the orbit approach a parabola in this case?

Math 241
Calculus III

Thomas
Honold

Functions of
Several
Variables

Basic Terminology
Examples
Complex Functions
Further Rules

Limit
Computations

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Functions of Several Variables

Basic Terminology

Examples

Complex Functions

Further Rules

2 Limit Computations

Functions of
Several
Variables

Basic Terminology

Examples

Complex Functions

Further Rules

Limit
Computations

Today's Lecture: Functions of Several Variables

Similarities

The concepts of limit and continuity almost verbatim carry over to functions of several variables $f: D \rightarrow \mathbb{R}^m$, $\mathbf{x} \mapsto f(\mathbf{x}) = \mathbf{y}$ with $D \subseteq \mathbb{R}^n$, except that we now use the Euclidean length to measure distances in both the domain and range of f .

Recall that such a map f is said to have *domain* D , *codomain* \mathbb{R}^m , and *range* $f(D) = \{f(\mathbf{x}); \mathbf{x} \in D\}$.

Notes

- For $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, the notational variants

$$f(\mathbf{x}) = f((x_1, x_2, \dots, x_n)) = f(x_1, x_2, \dots, x_n)$$

and even column vector notation for \mathbf{x} and $\mathbf{y} = f(\mathbf{x})$ will be used synonymously.

- As in the case of curves, a “vector function” f with domain $D \subseteq \mathbb{R}^n$ and codomain \mathbb{R}^m determines m “scalar functions” f_1, f_2, \dots, f_m with codomain \mathbb{R} and the same domain D as f via

$$f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})) \quad \text{for } \mathbf{x} \in D.$$

Examples

Consider the three functions

$$f(x, y) = x^2 + y^2, \quad (x, y) \in \mathbb{R}^2,$$

$$g(x_1, x_2) = \begin{pmatrix} x_1 + x_2 \\ 2x_1 - 3x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad (x_1, x_2) \in \mathbb{R}^2,$$

$$h(x, y) = (x^2 - y^2, 2xy), \quad (x, y) \in \mathbb{R}^2.$$

All three functions have domain \mathbb{R}^2 , but f is scalar-valued (codomain \mathbb{R}) while g and h are vector-valued (codomain \mathbb{R}^2).

The range of f is

$$f(\mathbb{R}^2) = \{x^2 + y^2; (x, y) \in \mathbb{R}^2\} = \mathbb{R}_0^+$$

(the non-negative real numbers), since these are exactly the numbers admitting a representation as a sum of two squares (even as a single square).

The range of g , which is a linear map, consists of all $(y_1, y_2) \in \mathbb{R}^2$ for which the system $x_1 + x_2 = y_1 \wedge 2x_1 - 3x_2 = y_2$ is solvable for x_1, x_2 . Since $\text{rk } \mathbf{A} = 2$, every such system is (uniquely) solvable. Hence $g(\mathbb{R}^2) = \mathbb{R}^2$, i.e., g is surjective (even bijective).

Examples (cont'd)

The map h is the real representation of the complex function $\mathbb{C} \rightarrow \mathbb{C}$, $z \mapsto z^2$ ("squaring map"), since

$$(x, y)^2 = (x + y i)^2 = x^2 - y^2 + 2xy i = (x^2 - y^2, 2xy),$$

where as usual a complex number $z = x + y i \in \mathbb{C}$ is identified with the point $(x, y) \in \mathbb{R}^2$.

Since every complex number w has at least one square root (two square roots $\pm \sqrt{r} e^{i\phi/2}$ if $w = r e^{i\phi} \neq 0$), the map h is also surjective.

You should also note the different notations used when specifying f, g, h . For f we have used x, y as variable names, and one would then denote the dependent variable by z , i.e., write $z = f(x, y)$.

For g we have used x_1, x_2 , and one can continue as

$\begin{pmatrix} x_3 \\ x_4 \end{pmatrix} = g(x_1, x_2)$ or, as we have done, as $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = g(x_1, x_2)$.

Moreover, both row and column vector notation has been used in the specification of g, h (in the case of g even a mix of the two notations).

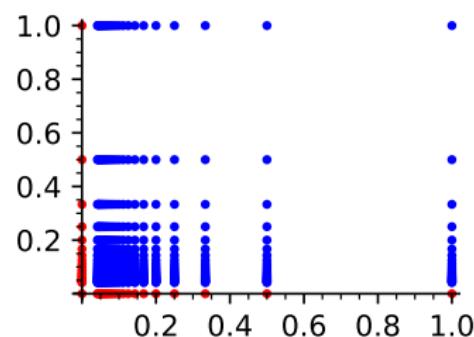
The concept of a limit for functions of several variables (strictly speaking, also for functions of one variable) requires the concept of an “accumulation point”, which was defined earlier:

Definition (recalled)

A point $\mathbf{x} \in \mathbb{R}^n$ is called an *accumulation point* of a set $D \subseteq \mathbb{R}^n$, if every ball around \mathbf{x} (of positive radius) contains a point in $D \setminus \{\mathbf{x}\}$. The set of all accumulation points of D is denoted by D' .

Examples

- The set $D = \{(1/k, 1/l); k, l \in \mathbb{Z}^+\}$
has $D' = \{(0, 0)\} \cup \{(1/k, 0); k \in \mathbb{Z}^+\}$
 $\cup \{(0, 1/l); l \in \mathbb{Z}^+\}$; cf. picture.
For example, the origin is an accumulation point of D , since any ball (“disk”) $B_\epsilon(0, 0) = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 < \epsilon^2\}$ contains a point of D , e.g., $([2/\epsilon]^{-1}, [2/\epsilon]^{-1})$.



- Open balls $B_r(\mathbf{a}) = \{\mathbf{x} \in \mathbb{R}^n; |\mathbf{x} - \mathbf{a}| < r\}$ have
 $B_r(\mathbf{a})' = \{\mathbf{x} \in \mathbb{R}^n; |\mathbf{x} - \mathbf{a}| \leq r\} = \overline{B_r(\mathbf{a})}$.

Definition (Limits and Continuity)

Suppose $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is a function and $\mathbf{x}_0 \in \mathbb{R}^n$.

- 1 f is said to have the limit $\mathbf{a} \in \mathbb{R}^m$ for $\mathbf{x} \rightarrow \mathbf{x}_0$, notation $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = \mathbf{a}$, if $\mathbf{x}_0 \in D'$ and for every $\epsilon > 0$ there exists $\delta > 0$ such that $\mathbf{x} \in D \wedge 0 < |\mathbf{x} - \mathbf{x}_0| < \delta$ implies $|f(\mathbf{x}) - \mathbf{a}| < \epsilon$.
- 2 f is said to be continuous at \mathbf{x}_0 if $\mathbf{x}_0 \in D$ and for every $\epsilon > 0$ there exists $\delta > 0$ such that $\mathbf{x} \in D \wedge |\mathbf{x} - \mathbf{x}_0| < \delta$ implies $|f(\mathbf{x}) - f(\mathbf{x}_0)| < \epsilon$; equivalently, $f(B_\delta(\mathbf{x}_0) \cap D) \subseteq B_\epsilon(f(\mathbf{x}_0))$.

Notes

- If $\mathbf{x}_0 \notin D'$ then there exists $\delta > 0$ such that the “punctured ball” $\{\mathbf{x} \in \mathbb{R}^n; 0 < |\mathbf{x} - \mathbf{x}_0| < \delta\}$ contains no point of D .
 $\Rightarrow \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = \mathbf{a}$ would be true for any $\mathbf{a} \in \mathbb{R}^m$.
 Restricting the definition to $\mathbf{x}_0 \in D'$ ensures that a function can have at most one limit for $\mathbf{x} \rightarrow \mathbf{x}_0$ (easy exercise).
- If $f = (f_1, \dots, f_m)$ with $f_i: D \rightarrow \mathbb{R}$ and $\mathbf{x}_0 \in D'$ then $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x})$ exists iff all limits $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f_i(\mathbf{x})$, $1 \leq i \leq m$ exist; if this is the case, we have
 $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = (\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f_1(\mathbf{x}), \dots, \lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f_m(\mathbf{x}))$.

Notes cont'd

- As in Calculus I, the definition of limits is extended to include certain improper limits, e.g., $\lim_{x \rightarrow x_0} f(x) = \pm\infty$ in the case $m = 1$ (defined in the obvious way) and, particularly important, $\lim_{|x| \rightarrow \infty} f(x) = a$, if for every $\epsilon > 0$ there exists $R > 0$ such that $f(x) \in B_\epsilon(a)$ for all $x \in D$ with $|x| > R$.
Question: Which condition on D serves as replacement for $x_0 \in D'$ in this case?
- Continuity can be expressed in terms of limits:
If $x_0 \in D \setminus D'$ (x_0 is an *isolated point* of D) then f is continuous at x_0 ; if $x_0 \in D \cap D'$ then f is continuous at x_0 iff $\lim_{x \rightarrow x_0} f(x)$ exists and is equal to $f(x_0)$.

Afternote

The continuity of functions in *isolated points* x_0 of their domains is not of any importance. If you have difficulties to understand it, note that we can choose the response δ in such a way that $B_\delta(x_0)$ contains no point of D except x_0 . Then in the implication " $x \in D \wedge |x - x_0| < \delta$ implies $|f(x) - f(x_0)| < \epsilon$ " the premise is true only for $x = x_0$, in which case the conclusion is also true: $|f(x_0) - f(x_0)| = 0 < \epsilon$. Thus the implication is true for all x (and ϵ).

Graphs

Definition

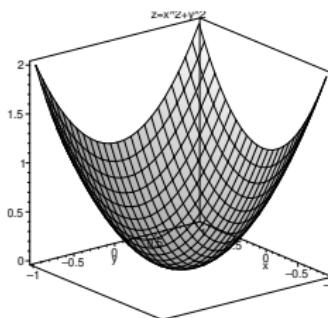
Let $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, be a function. The *graph* of f is the point set

$$G_f = \{(\mathbf{x}, f(\mathbf{x})); \mathbf{x} \in D\} \subseteq \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}.$$

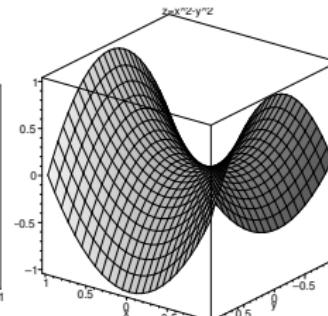
Alternatively, writing $f = (f_1, \dots, f_m)$, the graph of f is the set of points $(x_1, \dots, x_{m+n}) \in \mathbb{R}^{m+n}$ satisfying

$$(x_1, \dots, x_n) \in D \quad \text{and} \quad x_{n+i} = f_i(x_1, \dots, x_n) \text{ for } 1 \leq i \leq m.$$

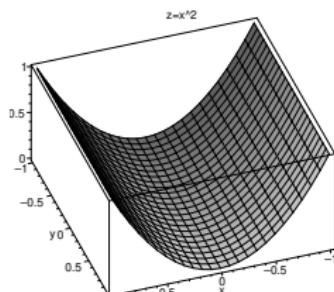
Examples



(a) $z = x^2 + y^2$



(b) $z = x^2 - y^2$



(c) $z = x^2$

Notes

- Graphs of multivariate functions can be visualized only in one case, $(n, m) = (2, 1)$, in which $n + m = 3$.
- Graphs are also defined for parametric curves (univariate vector-valued functions). Graphs of parametric curves can be visualized in the case $(n, m) = (1, 2)$, i.e., plane curves. The graph of a parametric plane curve is a non-parametric space curve, which encodes the plane curve together with its parametrization. For example, the graph of the standard parametrization $f(t) = (\cos t, \sin t)$, $t \in [0, 2\pi]$, of the unit circle is

$$G_f = \{(t, \cos t, \sin t); t \in [0, 2\pi]\},$$

i.e., a piece of a helix around the x -axis. Different parametrizations of the unit circle give different graphs, e.g., changing the parameter interval to \mathbb{R} produces the full helix.

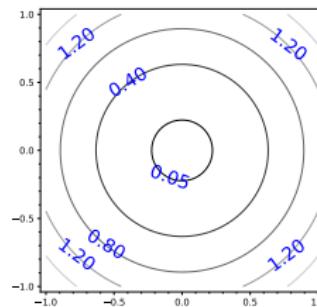
Level Sets

Definition

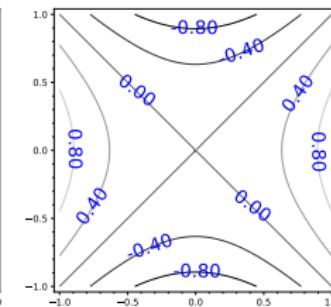
Let $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$ be a function with 1-dimensional codomain and $k \in \mathbb{R}$. The *level set* (or *contour*) of f corresponding to the level k is the point set

$$N_f(k) = f^{-1}(k) = \{\mathbf{x} \in D; f(\mathbf{x}) = k\} \subseteq D.$$

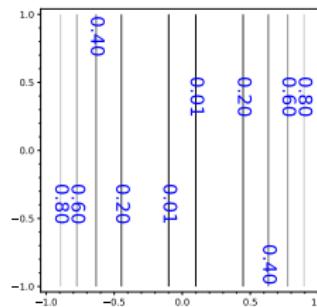
Examples



(a) $z = x^2 + y^2$



(b) $z = x^2 - y^2$



(c) $z = x^2$

Examples

Example (the length function)

The length function $\mathbb{R}^n \rightarrow \mathbb{R}^1$, $\mathbf{x} \mapsto |\mathbf{x}| = \sqrt{x_1^2 + \cdots + x_n^2}$ has domain \mathbb{R}^n , codomain \mathbb{R} and range $\mathbb{R}_0^+ = \{r \in \mathbb{R}; r \geq 0\}$. It is (uniformly) continuous everywhere, since

$$|\mathbf{x}| = |(\mathbf{x} - \mathbf{y}) + \mathbf{y}| \leq |\mathbf{x} - \mathbf{y}| + |\mathbf{y}|$$

and similarly for $|\mathbf{y}|$, which implies

$$||\mathbf{x}| - |\mathbf{y}|| = \pm(|\mathbf{x}| - |\mathbf{y}|) \leq |\mathbf{x} - \mathbf{y}|.$$

$\implies \delta = \epsilon$ works in the definition of continuity at $\mathbf{x}_0 = \mathbf{y}$ (independently of the particular choice of $\mathbf{x}_0 \in \mathbb{R}^n$).

One could also argue that $\mathbf{x} \mapsto |\mathbf{x}|$ is the composition of two continuous functions (a polynomial and the square root function), and hence itself continuous; cf. Slide “Further Rules”.

Example (cont'd)

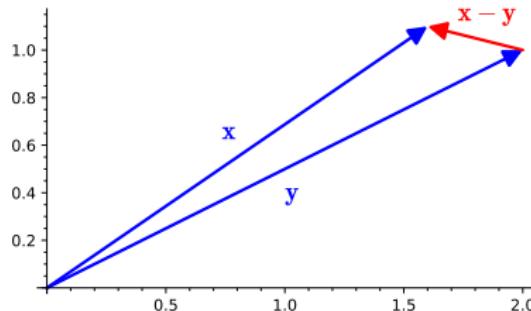


Figure: Continuity of the length function

If the points (vectors) \mathbf{x}, \mathbf{y} are close then their lengths $|\mathbf{x}|, |\mathbf{y}|$ are close; more precisely, the difference of their lengths is at most the length of their difference.

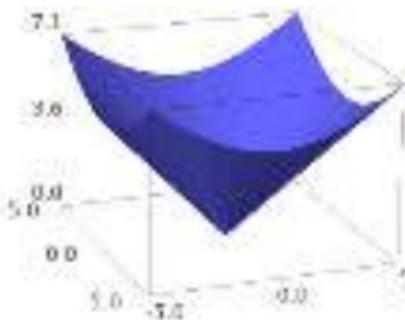
Example (cont'd)

For $n = 2$ the graph of the length function is

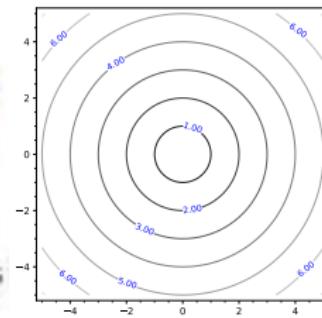
$$G = \left\{ (x, y, z) \in \mathbb{R}^3; z = \sqrt{x^2 + y^2} \right\} \subset \mathbb{R}^3$$

or, in parametric form, $G = \left\{ (x, y, \sqrt{x^2 + y^2}); x, y \in \mathbb{R} \right\}.$

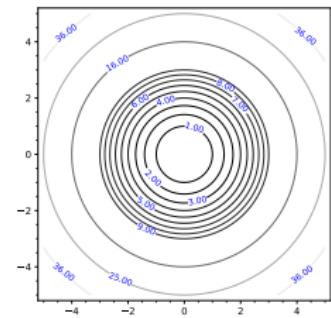
It has a cusp at the origin, as a 3-dimensional plot reveals.



(a) Graph of $|z|$



(b) Contours of $|z|$



(c) Contours of $z = x^2 + y^2$

Note the different spacing of its contours (the circles $x^2 + y^2 = k^2$) and those of $(x, y) \mapsto x^2 + y^2$ (the circles $x^2 + y^2 = k$).

Example (polynomial functions)

A polynomial function in n variables has the form

$$a(\mathbf{x}) = a(x_1, x_2, \dots, x_n) = \sum_{(i_1, \dots, i_n) \in I} a_{i_1, i_2, \dots, i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n},$$

where $I \subset \mathbb{N}^n$ is a finite set.

If $a \neq 0$ (i.e., a has at least one nonzero coefficient a_{i_1, \dots, i_n}) then the non-negative integer

$\deg(a) = \max\{i_1 + i_2 + \cdots + i_n; a_{i_1, \dots, i_n} \neq 0\}$ is called *degree* of a .

Polynomial functions have domain \mathbb{R}^n (unless restricted otherwise) and are continuous everywhere.

This follows from the fact that the coordinate functions

$\pi_i: \mathbb{R}^n \rightarrow \mathbb{R}, \mathbf{x} \mapsto x_i$ are continuous, since

$$|\pi_i(\mathbf{x}) - \pi_i(\mathbf{y})| = |x_i - y_i| \leq \sqrt{\sum_{j=1}^n (x_j - y_j)^2} = |\mathbf{x} - \mathbf{y}|$$

(cf. the discussion of $\mathbf{x} \mapsto |\mathbf{x}|$), and the fact that (finite) sums and products of continuous functions, as well as all constant functions, are continuous (analogous to the case of one-variable functions, cf. Calculus I, and with similar proofs).

Example (rational functions)

A rational function in n variables is a quotient

$$f(\mathbf{x}) = f(x_1, \dots, x_n) = \frac{a(x_1, \dots, x_n)}{b(x_1, \dots, x_n)}$$

of polynomials a and $b \neq 0$.

If one assumes that $\gcd(a, b) = 1$ (i.e., a and b have no common polynomial factor of degree ≥ 1) then a and b uniquely determined by f up to a factor $\lambda \in \mathbb{R} \setminus \{0\}$.

The (maximal) domain of f is then $D = \{\mathbf{x} \in \mathbb{R}^n; b(\mathbf{x}) \neq 0\}$, and f is continuous in D . This follows from the continuity of a, b and the fact that quotients of continuous functions are continuous wherever they are defined.

Examples are $f(x, y) = \frac{2x}{x^2 + y^2 + 1}$, $g(x, y) = \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}$.

Both f and g have domain \mathbb{R}^2 . The rational functions $1/f$, $1/g$, $1/(fg)$ have domains $D_1 = \{(x, y) \in \mathbb{R}^2; x \neq 0\}$ (" \mathbb{R}^2 without the y -axis"), $D_2 = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \neq 1\}$ (" \mathbb{R}^2 without the unit circle"), and $D_3 = D_1 \cap D_2$ (" \mathbb{R}^2 without both y -axis and unit circle"), respectively.

An Example with a Nasty Domain

Example

Consider the function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^2$, defined by

$$f(x, y) = \frac{1}{1 - xy} + \sqrt{y^2 - x^3 + x},$$

where $D = \{(x, y) \in \mathbb{R}^2; xy \neq 1 \wedge y^2 \geq x^3 - x\}$ is the maximal domain for which the expression on the right-hand side is defined.

The set D is disconnected and consists of the 3 connected components

$$D_1 = \{(x, y) \in \mathbb{R}^2; y < 1/x < 0\},$$

$$D_2 = \{(x, y) \in \mathbb{R}^2; y > 1/x > 0 \wedge y^2 \geq x^3 - x\},$$

$$D_3 = \{(x, y) \in \mathbb{R}^2; xy < 1 \wedge y^2 \geq x^3 - x\}.$$

The last set D_3 has a “hole”, viz. the “egg-like” set

$$D_4 = \{(x, y) \in \mathbb{R}^2; -1 < x < 0 \wedge y^2 \leq x^3 - x\}.$$

Since compositions of continuous functions are continuous (cf. a subsequent slide), f is continuous in its whole domain D .

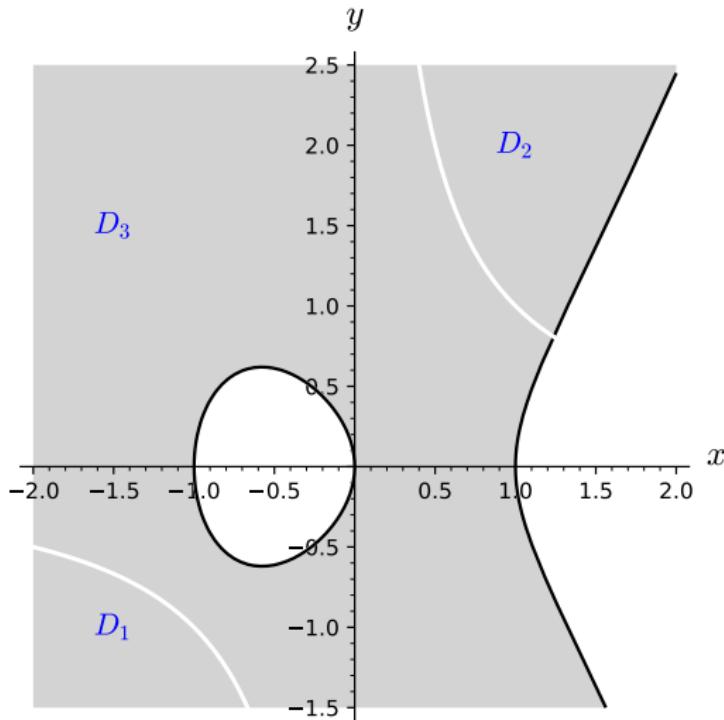


Figure: Maximal domain of $(x, y) \mapsto \frac{1}{1-xy} + \sqrt{y^2 - x^3 + x}$
(the black boundary curves are included in the domain)

Examples with $m > 1$

Example (polar coordinate map)

Consider $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2, (r, \phi) \mapsto (r \cos \phi, r \sin \phi)$.

This function is continuous everywhere, since it is composed of continuous one-variable functions, coordinate projections and addition/multiplication.

If we restrict the domain to the “strip” $S = (0, +\infty) \times [0, 2\pi)$, we obtain a bijection from S onto the “punctured plane” $\mathbb{R}^2 \setminus \{(0, 0)\}$.

Example

Consider $g: D \rightarrow \mathbb{R}^2$ defined by

$$g(b, c) = \left(\frac{1}{2} \left(-b - \sqrt{b^2 - 4c} \right), \frac{1}{2} \left(-b + \sqrt{b^2 - 4c} \right) \right),$$

where $D = \{(b, c) \in \mathbb{R}^2; b^2 \geq 4c\}$ is the corresponding maximal domain.

The map g is continuous (on D) and inverts

$f(x_1, x_2) = (-x_1 - x_2, x_1 x_2)$; think of solving the quadratic $x^2 + bx + c = (x - x_1)(x - x_2) = 0$.

Example (cont'd)

It is worth to look at the meaning of continuity in this case:

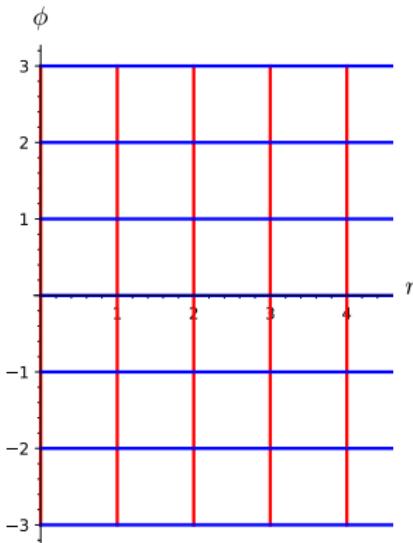
Continuity of g in $(b, c) \in D$ means that for any given $\epsilon > 0$ we can find a response $\delta > 0$ such that the roots x'_1, x'_2 of any polynomial $x^2 + b'x + c'$ with $|b' - b| < \delta$ and $|c' - c| < \delta$ satisfy $|x'_1 - x_1| < \epsilon$ and $|x'_2 - x_2| < \epsilon$; cf. a later exercise. In other words, small changes of the coefficients of a quadratic polynomial result in small changes of its roots.

This statement is slightly (but only slightly!) inaccurate, since it doesn't take into account the possibility $(b', c') \notin D$. For example, if $b^2 - 4c = 0$ then there are choices (b', c') arbitrarily close to (b, c) satisfying $b'^2 - 4c' < 0$, and hence such that $x^2 + b'x + c' = 0$ has no solution.

Also, the statement doesn't indicate how δ varies with ϵ . For example, if $b^2 - 4c$ is close to zero then $\delta \ll \epsilon$ (owing to the fact that the derivative of $x \mapsto \sqrt{x}$ is unbounded near $x = 0$), and hence in this case for a prescribed accuracy level of the roots x_1, x_2 one needs to know the coefficients b, c to a much higher accuracy. (A refined error analysis of this kind will be discussed later using the machinery of differentiation.)

Visualization of Maps $\mathbb{R}^2 \rightarrow \mathbb{R}^2$

Draw images of coordinate lines



f

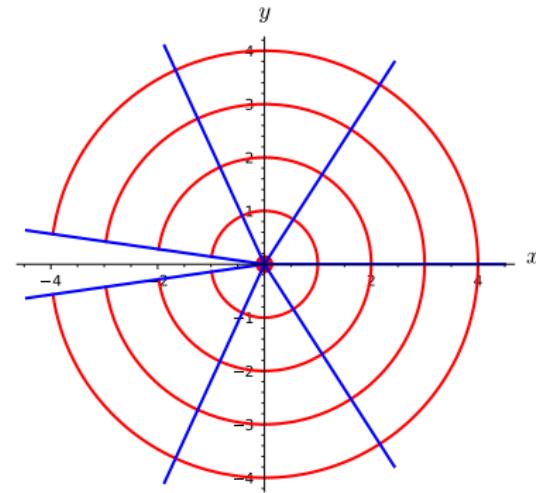


Figure: Images of the coordinate lines $r = \text{const}$ (red) and $\phi = \text{const}$ (blue) under the polar coordinate map $f(r, \phi) = (r \cos \phi, r \sin \phi)$

Can you imagine how f “deforms” the grid shown on the left into the structure on the right?

Visualization of Maps $\mathbb{R}^2 \rightarrow \mathbb{R}^2$

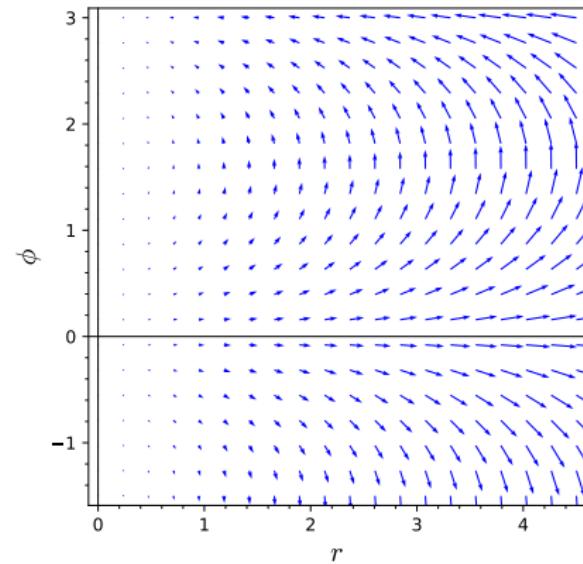
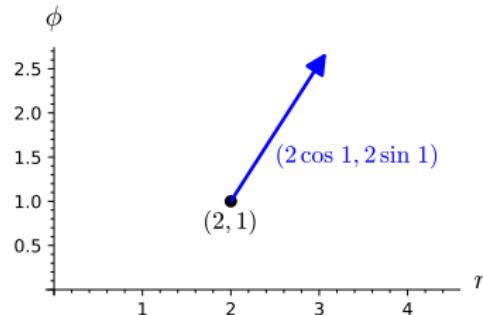
The vector field view

Functions of
Several
Variables

Basic Terminology
Examples

Complex Functions
Further Rules

Limit
Computations



(a) The vector
 $f(r, \phi) = (r \cos \phi, r \sin \phi)$ is
attached to the point (r, ϕ) ...

(b) ... resulting in a “vector field”.
(Lengths of vectors are scaled
by a fixed constant.)

Example (Complex functions)

The complex numbers \mathbb{C} are nothing but the vectors $(a, b) \in \mathbb{R}^2$ multiplied in a fancy way:

$$(a, b)(c, d) = (ac - bd, ad + bc).$$

Hence complex functions $f: D \rightarrow \mathbb{C}$, $D \subseteq \mathbb{C}$, correspond to functions $f: D \rightarrow \mathbb{R}^2$, $D \subseteq \mathbb{R}^2$, providing many interesting examples of such functions. Before discussing examples, we provide a brief review of the most basic properties of complex numbers.

The standard representation $(a, b) = a + bi$:

This representation is obtained by the identification $a \triangleq (a, 0)$, i.e., real numbers are considered as points on the x -axis (“real axis”), and by setting $i = (0, 1)$ (the so-called “imaginary unit”). Indeed,

$$i^2 = (0, 1)(0, 1) = (-1, 0) \triangleq -1,$$

$$a + bi \triangleq (a, 0) + (b, 0)(0, 1) = (a, 0) + (0, b) = (a, b).$$

a is called *real part* of $z = a + bi$ (notation $a = \operatorname{Re} z$), and b is called *imaginary part* of z (notation $b = \operatorname{Im} z$).

Example (Complex functions cont'd)

The field structure:

The complex numbers form a *field* (cf. Lecture 1) with respect to vector addition $(a, b) + (c, d) = (a + c, b + d)$ and multiplication $(a, b)(c, d) = (ac - bd, ad + bc)$. This field is commonly denoted by \mathbb{C} . The two distinguished elements are $0 = (0, 0)$ and $1 = (1, 0)$, and the multiplicative inverse of $a + bi \neq 0$ is $\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2} i$. Memorize this as follows:

$$\frac{1}{a+bi} = \frac{a-bi}{(a+bi)(a-bi)} = \frac{a-bi}{a^2+b^2} = \frac{a}{a^2+b^2} - \frac{b}{a^2+b^2} i.$$

Complex conjugation:

The complex number $\overline{a+bi} = a-bi$ is called *complex conjugate* of $a+bi$, and the map $\mathbb{C} \rightarrow \mathbb{C}$, $z \mapsto \overline{z}$ *complex conjugation*. It satisfies $\overline{z+w} = \overline{z} + \overline{w}$, $\overline{zw} = \overline{z}\overline{w}$, i.e., forms a so-called *field automorphism* of \mathbb{C} .

The absolute value of \mathbb{C} :

This is just the Euclidean length

$|a+ib| = |(a, b)| = \sqrt{a^2+b^2} = \sqrt{(a+ib)(a-ib)}$, which also gives that in general $z\overline{z} = |z|^2$ and $1/z = \overline{z}/(z\overline{z}) = \overline{z}/|z|^2$

Example (Complex functions cont'd)

Just like the real absolute value, the Euclidean length is multiplicative also for $n = 2$, i.e., $|zw| = |z| |w|$ holds for all $z, w \in \mathbb{C}$ (easy exercise), justifying the name “absolute value”.

Euler's Identity:

This is the stunning formula

$$e^{i\phi} = \cos \phi + i \sin \phi, \quad \phi \in \mathbb{R},$$

which requires knowledge of the complex exponential function $\mathbb{C} \rightarrow \mathbb{C}$, $z \mapsto e^z = \exp(z) = \sum_{n=0}^{\infty} z^n / n!$; cf. Homework 5, Exercise H28 c). It shows that the polar coordinate representation

$$(x, y) = r(\cos \phi, \sin \phi) = r e^{i\phi}, \quad r \geq 0, \phi \in [0, 2\pi)$$

has a succinct “complex formulation” as $z = r e^{i\phi} = |z| e^{i\phi}$, and for $\phi = \pi$ specializes to the even more stunning $e^{i\pi} = -1$.

The geometry of complex multiplication:

Since $zw = (r e^{i\phi})(s e^{i\psi}) = rs e^{i(\phi+\psi)}$ (using the functional equation for $z \mapsto e^z$; cf. Worksheet 6), the multiplication map $\mathbb{C} \rightarrow \mathbb{C}$, $w \mapsto zw$ consists of a rotation with angle ϕ centered at the origin, followed (or preceded) by a scaling with scale factor $r = |z|$.

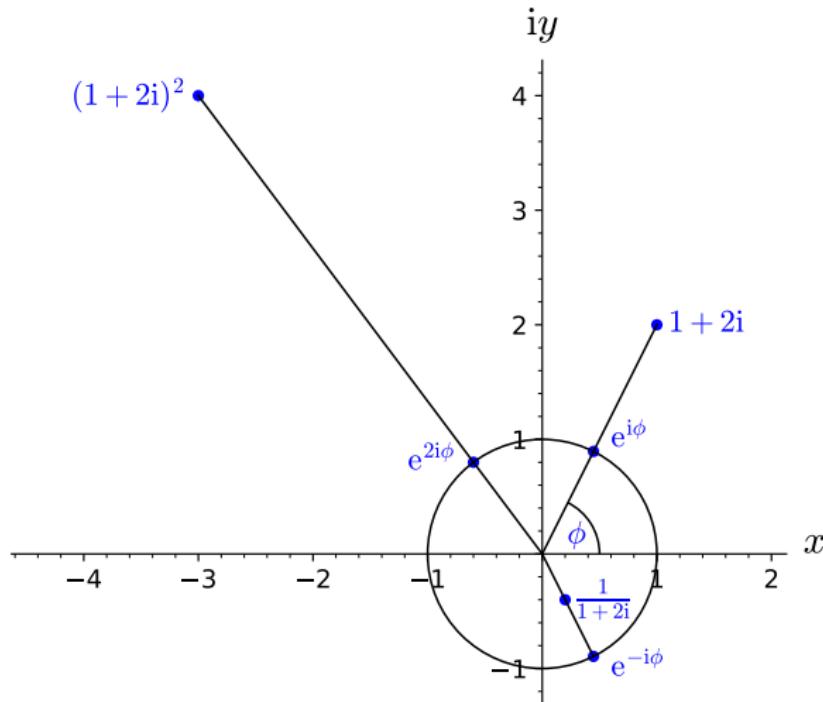


Figure: The complex numbers $1 + 2i$, $(1 + 2i)^2 = -3 + 4i$, $\frac{1}{1+2i} = \frac{1}{5}(1 - 2i)$, graphed together with their arguments $\phi \approx 1.07$ (satisfying $\tan \phi = 2$), 2ϕ , $-\phi$, respectively, their normalizations to unit length (the angle $\phi \in [0, 2\pi]$ in $z = |z| e^{i\phi}$ is called the *argument* of z , notation $\phi = \arg z$)

Example (Complex functions cont'd)

Roots of unity: In \mathbb{C} the equation $z^n = 1$ has the n solutions

$$z = e^{2\pi i k/n} = \cos\left(\frac{2\pi k}{n}\right) + i \sin\left(\frac{2\pi k}{n}\right), \quad k = 0, 1, \dots, n-1.$$

This follows from the functional equation for $z \mapsto e^z$, which implies $(e^{i\phi})^n = e^{in\phi}$, and $e^{2\pi i k} = 1$. The solutions are powers of $z_n = e^{2\pi i/n}$ and form the vertices of a regular n -gon centered at 0. They are called *n-th roots of unity*. Accordingly, we have over \mathbb{C} the polynomial factorization $X^n - 1 = \prod_{k=0}^{n-1} (X - e^{2\pi i k/n})$.

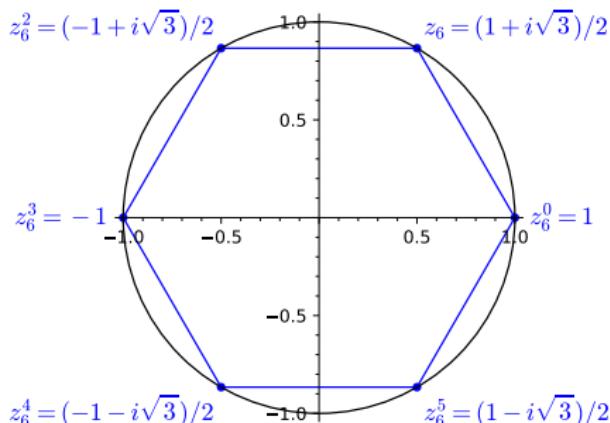


Figure: The 6-th roots of unity in \mathbb{C}

Example (Complex functions cont'd)

The Fundamental Theorem of Algebra (FTA):

FTA says that every non-constant polynomial $p(X)$ with coefficients in \mathbb{C} has a zero in \mathbb{C} . Using induction and the fact that $p(c) = 0$ implies $p(X) = (X - c)q(X)$ for some polynomial $q(X)$, we obtain that $p(X)$ splits into linear factors over \mathbb{C} , i.e.,

$$p(X) = \alpha(X - c_1)(X - c_2) \cdots (X - c_d)$$

with $d \in \mathbb{N}$ (the degree of $p(X)$), $\alpha \in \mathbb{C} \setminus \{0\}$, and $c_1, \dots, c_d \in \mathbb{C}$.

For (monic) polynomials $p(X)$ with real coefficients, FTA yields a factorization into linear polynomials $X - a$, $a \in \mathbb{R}$, and quadratic polynomials $(X - c)(X - \bar{c}) = X^2 + BX + C$ with $B, C \in \mathbb{R}$ and $B^2 - 4C < 0$. The key observation behind this is that non-real zeros of a real polynomial must occur in complex conjugate pairs c, \bar{c} because of $p(\bar{c}) = \overline{p(c)}$.

No easy proof of FTA is known. A rather “elementary” (but still highly nontrivial) proof due to ARGAND assumes that FTA is false and concludes from this that the real-valued function $z \mapsto |p(z)|$ attains a positive minimum at some point $z_0 \in \mathbb{C}$. After some calculations using algebraic properties of \mathbb{C} this leads to a contradiction.

Exercise

- 1 According to the previous remarks, the polynomial $X^4 + 1$, which has no real root and hence no real linear factors, must factor in $\mathbb{R}[X]$ into two (monic) quadratic polynomials. Find this factorization.
- 2 There is a simple relation between the two factors. Can you explain it?

Example (Complex functions cont'd)

Now we list a few examples of complex functions and their real counterparts

- 1 The complex squaring map $h(z) = z^2$, $z \in \mathbb{C}$, has real representation $h(x, y) = (x^2 - y^2, 2xy)$, $(x, y) \in \mathbb{R}^2$, as we have seen.
- 2 Complex conjugation $\mathbb{C} \rightarrow \mathbb{C}$, $z \mapsto \bar{z}$ has real representation $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, $(x, y) \mapsto (x, -y)$. Geometrically, it affords the reflection at the x -axis (“real” axis).
- 3 Multiplicative inversion $\mathbb{C} \setminus \{0\} \rightarrow \mathbb{C} \setminus \{0\}$, $z \mapsto 1/z$ has real representation $\mathbb{R}^2 \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}^2 \setminus \{\mathbf{0}\}$, $(x, y) \mapsto \left(\frac{x}{x^2+y^2}, -\frac{y}{x^2+y^2} \right)$.
- 4 $\mathbb{C} \setminus \{0\} \rightarrow \mathbb{C} \setminus \{0\}$, $z \mapsto 1/\bar{z} = z / |z|^2$ has real representation $\mathbb{R}^2 \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}^2 \setminus \{\mathbf{0}\}$, $(x, y) \mapsto \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2} \right)$. This map fixes the unit circle point-wise, the lines through the origin set-wise, and maps points outside the unit disk inside the unit disk and vice versa. It is sometimes called *reflection at the unit circle*.

Example (Complex functions cont'd)

- 5** $\mathbb{C} \setminus \{0\} \rightarrow \mathbb{C} \setminus \{0\}$, $z \mapsto z + 1/z$ has real representation $\mathbb{R}^2 \setminus \{\mathbf{0}\} \rightarrow \mathbb{R}^2 \setminus \{\mathbf{0}\}$, $(x, y) \mapsto \left(x + \frac{x}{x^2+y^2}, y - \frac{y}{x^2+y^2}\right)$. This map is called ZHUKOVSKII transformation and plays an important role in airfoil design.
- 6** The polar coordinate map $f(r, \phi) = (r \cos \phi, r \sin \phi)$, $(r, \phi) \in \mathbb{R}^+ \times \mathbb{R}$, has “semi-complex” representation $f(r, \phi) = (r \cos \phi) + i(r \sin \phi) = r e^{i\phi} = e^{\ln(r)+i\phi}$.

Further examples of complex functions are polynomials

$$a(x+iy) = a(x, y) = \sum_{(i_1, i_2) \in I} a_{i_1, i_2} x^{i_1} y^{i_2} \text{ with coefficients } a_{i_1, i_2} \in \mathbb{C}.$$

An observation in this regard, which is sometimes useful, is that such polynomials can also be expressed as polynomials in z and \bar{z} via $x = \frac{1}{2}(z + \bar{z})$, $y = \frac{1}{2i}(z - \bar{z})$.

For example, we have for $z = x + iy$ the identities

$$x^2 - y^2 = \operatorname{Re}(z^2) = \frac{1}{2} \left(z^2 + \bar{z}^2 \right),$$

$$x^3 - 3xy^2 = \operatorname{Re}(z^3) = \frac{1}{2} \left(z^3 + \bar{z}^3 \right),$$

$$x^4 - 6x^2y^2 + y^4 = \operatorname{Re}(z^4) = \frac{1}{2} \left(z^4 + \bar{z}^4 \right), \quad \text{etc.}$$

Further Rules

When computing with continuous functions, we have used the following additional rules:

Quotients

Suppose $f: D_1 \rightarrow \mathbb{R}$ and $g: D_2 \rightarrow \mathbb{R}$, $D_1, D_2 \subseteq \mathbb{R}^n$, are continuous at $\mathbf{x}_0 \in D_1 \cap D_2$ and $g(\mathbf{x}_0) \neq 0$. Then $\mathbf{x} \mapsto f(\mathbf{x})/g(\mathbf{x})$ is also continuous at \mathbf{x}_0 .

Note that the maximal domain of $\mathbf{x} \mapsto f(\mathbf{x})/g(\mathbf{x})$ is $\{\mathbf{x} \in D_1 \cap D_2; g(\mathbf{x}) \neq 0\}$. The proof of this rule is almost the same as that for quotients of univariate functions.

Composition

Suppose $f: D_1 \rightarrow \mathbb{R}^m$, $D_1 \subseteq \mathbb{R}^n$, is continuous in $\mathbf{x}_0 \in D_1$ and $g: D_2 \rightarrow \mathbb{R}^p$, $f(D_1) \subseteq D_2 \subseteq \mathbb{R}^m$ is continuous in $\mathbf{y}_0 = f(\mathbf{x}_0)$. Then $g \circ f: D_1 \rightarrow \mathbb{R}^p$ is continuous in \mathbf{x}_0 .

“Ball proof”.

Given $\epsilon > 0$ choose $\delta' > 0$ such that $g(B_{\delta'}(\mathbf{y}_0) \cap D_2) \subseteq B_\epsilon(\mathbf{z}_0)$, $\mathbf{z}_0 = f(\mathbf{y}_0)$, and $\delta > 0$ such that $f(B_\delta(\mathbf{x}_0) \cap D_1) \subseteq B_{\delta'}(\mathbf{y}_0)$.

$$\Rightarrow g(f(B_\delta(\mathbf{x}_0) \cap D_1)) \subseteq B_\epsilon(\mathbf{z}_0)$$


Why continuity of $f(x_1, \dots, x_n)$ is useful

The following properties of continuous n -variable functions $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, will be proved later:

- ① If D is *closed* (i.e., D contains its limit points) and *bounded* (i.e., D is contained in a ball), then f attains a maximum and a minimum.
- ② If D is *connected* (i.e., D is not the union of non-empty subsets A, B with $\overline{A} \cap B = A \cap \overline{B} = \emptyset$) and there exist $\mathbf{x}_1, \mathbf{x}_2 \in D$ such that $f(\mathbf{x}_1) = a < b = f(\mathbf{x}_2)$ then f attains every value in $[a, b]$.

Limit Computations

Example

Consider the two functions $f, g: \mathbb{R}^2 \setminus \{(0, 0)\} \rightarrow \mathbb{R}$ defined by

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

Determine whether $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$ and $\lim_{(x,y) \rightarrow (0,0)} g(x, y)$ exist.

f and g satisfy

$$f(x, 0) = f(0, y) = 0, \quad g(x, 0) = g(0, y) = 0,$$

i.e., both functions vanish on the coordinate axes.

⇒ If the limits exist, they must be equal to zero.

Example (cont'd)

On the line $x = y$ we have

$$f(x, y) = f(x, x) = \frac{x^2}{x^2 + x^2} = \frac{1}{2} \neq 0$$

$\implies \lim_{(x,y) \rightarrow (0,0)} f(x, y)$ doesn't exist.

For g we can use the estimate

$$|g(x, y)| = \frac{|y|}{2} \cdot \frac{|2xy|}{x^2 + y^2} \leq \frac{|y|}{2}$$

to conclude that

$$0 \leq \lim_{(x,y) \rightarrow (0,0)} |g(x, y)| \leq \lim_{(x,y) \rightarrow (0,0)} |y| / 2 = 0,$$

i.e., $\lim_{(x,y) \rightarrow (0,0)} |g(x, y)| = 0$ and $\lim_{(x,y) \rightarrow (0,0)} g(x, y) = 0$.

Note

Since $\lim_{(x,y) \rightarrow (0,0)} g(x, y) = 0$, we can extend g to a continuous function on \mathbb{R}^2 by defining $g(0, 0) = 0$.

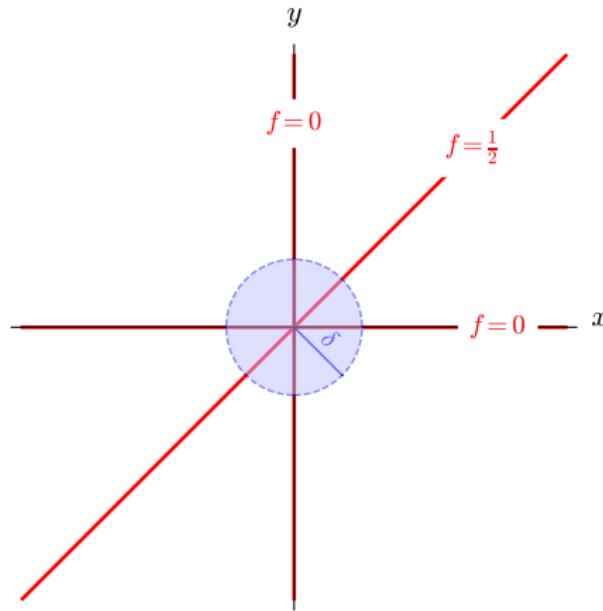


Figure: Non-existence of the limit $\lim_{(x,y) \rightarrow (0,0)} f(x,y)$

Every ball $B_\delta(0,0)$ contains points (x, y) with $f(x, y) = 0$ and points (x, y) with $f(x, y) = \frac{1}{2}$.

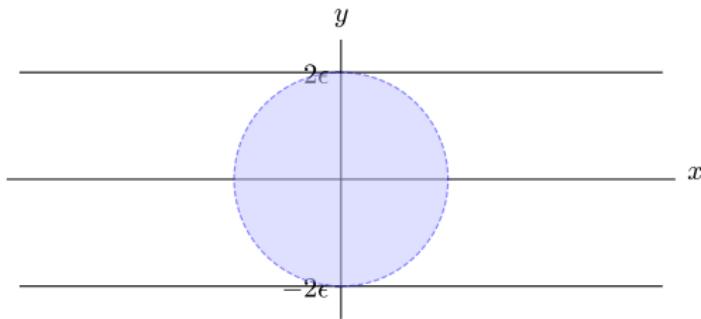
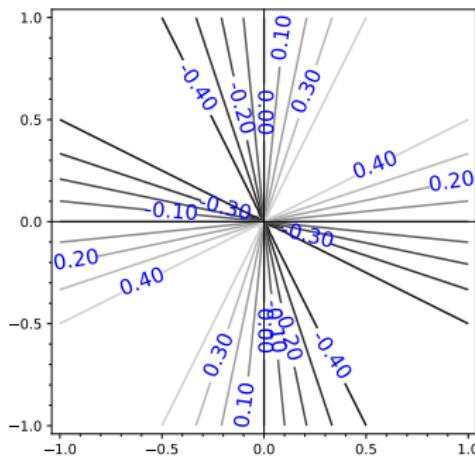


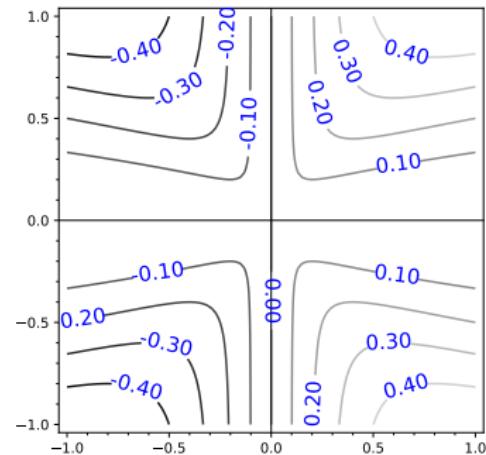
Figure: Existence of the limit $\lim_{(x,y) \rightarrow (0,0)} g(x,y)$

$|y| < 2\epsilon$ implies $|g(x,y)| < \epsilon$, so that we can take $\delta := 2\epsilon$ as response.

Contour Plots



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



$$(b) g(x, y) = \frac{xy^2}{x^2+y^2}$$

f in polar coordinates

$$f(r \cos \phi, r \sin \phi) = \frac{(r \cos \phi)(r \sin \phi)}{r^2} = \cos \phi \sin \phi = \frac{1}{2} \sin(2\phi)$$

$\Rightarrow -\frac{1}{2} \leq f(x, y) \leq \frac{1}{2}$ and the k -contour $N_f(k)$ consists of 0, 1 or 2 (possibly broken) lines through the origin.

Problem

Let $D = \{(x, y) \in \mathbb{R}^2; x > 0, y > 0\}$ and $g: D \rightarrow \mathbb{R}$ be defined by $g(x, y) = \frac{1}{x+y}$.

- 1 For which points of $(x_0, y_0) \in \mathbb{R}^2$ does it make sense to ask for $\lim_{(x,y) \rightarrow (x_0,y_0)} g(x, y)$? Decide whether the corresponding limits exist.
- 2 Does $\lim_{|(x,y)| \rightarrow \infty} g(x, y)$ exist?

Solution

(1) The limits make only sense in accumulation points of D . These are all points in D and the points on the non-negative x - and y -axis.

Since g is a rational function, it is continuous in D and we have

$$\lim_{(x,y) \rightarrow (x_0,y_0)} g(x, y) = \lim_{(x,y) \rightarrow (x_0,y_0)} \frac{1}{x+y} = \frac{1}{x_0+y_0} = g(x_0, y_0)$$

for $(x_0, y_0) \in D$.

Solution (cont'd)

For points $(x_0, 0), (0, y_0)$ on the positive coordinate axes we have similarly

$$\lim_{(x,y) \rightarrow (x_0, 0)} \frac{1}{x+y} = \frac{1}{x_0}, \quad \lim_{(x,y) \rightarrow (0, y_0)} \frac{1}{x+y} = \frac{1}{y_0}.$$

This resembles the fact that g has a continuous extension to $\{(x, y) \in \mathbb{R}^2; y \neq -x\}$ (the maximal domain of $\frac{1}{x+y}$), which includes these points.

Now let $(x_0, y_0) = (0, 0)$.

We can satisfy any inequality $\frac{1}{x+y} > R$ by choosing x and y sufficiently small, viz., $x < \frac{1}{2R}$ and $y < \frac{1}{2R}$. The points (x, y) in the disk $B_{\frac{1}{2R}}(0, 0)$ satisfy both inequalities and hence also $\frac{1}{x+y} > R$.
 $\implies \delta = \frac{1}{2R} > 0$ serves as a response for any given $R > 0$, showing that

$$\lim_{(x,y) \rightarrow (0,0)} \frac{1}{x+y} = +\infty.$$

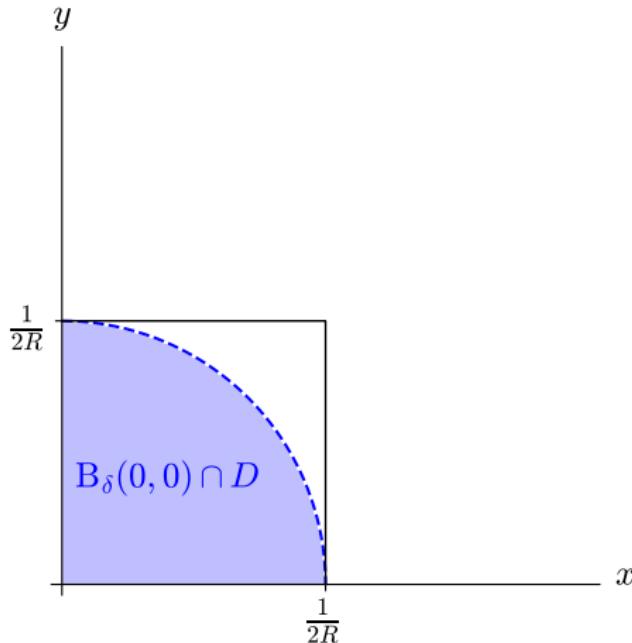


Figure: Illustration of $\lim_{\substack{|(x,y)| \rightarrow (0,0) \\ x>0, y>0}} \frac{1}{x+y} = +\infty$

Solution (cont'd)

(2) For the limit $\lim_{|(x,y)| \rightarrow \infty} g(x, y) = \lim_{\substack{|(x,y)| \rightarrow \infty \\ x>0, y>0}} \frac{1}{x+y}$ the points

$(x, y) \in D \setminus B_R(0, 0)$ with R large have to be considered. Since any such set is non-empty, it makes sense to consider this limit.

It is fairly obvious that $\lim_{\substack{|(x,y)| \rightarrow \infty \\ x>0, y>0}} \frac{1}{x+y} = 0$.

For a rigorous proof note that for $(x, y) \in D$ the inequality

$0 < \frac{1}{x+y} < \epsilon$ is equivalent to $x + y > 1/\epsilon$ and is certainly satisfied if at least one of x, y is $\geq 1/\epsilon$. If we choose the disk $B_R(0, 0)$ in such a way that it includes the square

$\{(x, y) \in \mathbb{R}^2; 0 < x, y < 1/\epsilon\}$, e.g. $R = \sqrt{2}/\epsilon$, then all points in D outside the disk have this property and hence satisfy $\frac{1}{x+y} < \epsilon$. We have thus exhibited a response R for any given $\epsilon > 0$.

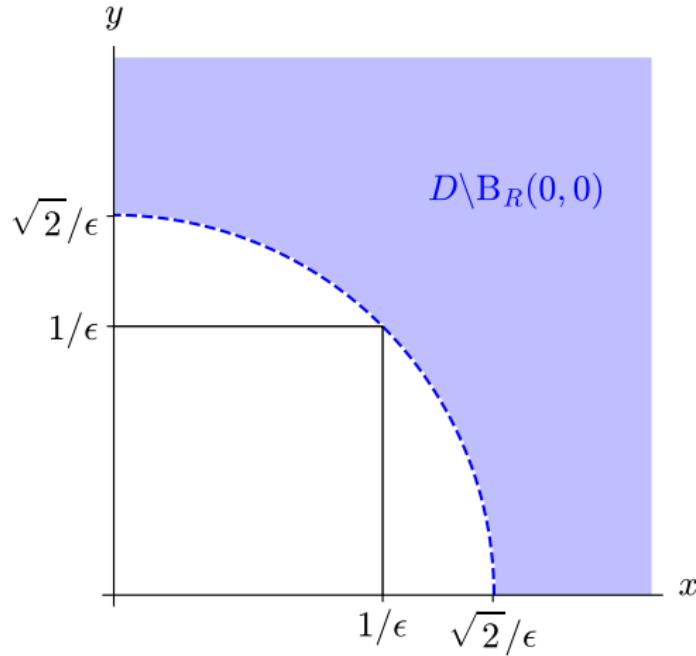


Figure: Illustration of $\lim_{\substack{|(x,y)| \rightarrow \infty \\ x>0, y>0}} \frac{1}{x+y} = 0$

Notes

- A more concise derivation of the limits for $(x, y) \rightarrow (0, 0)$ and $|(x, y)| \rightarrow \infty$ uses the estimate

$$\sqrt{x^2 + y^2} \leq |x| + |y| \leq \sqrt{2} \sqrt{x^2 + y^2}, \text{ or}$$

$$\frac{1}{\sqrt{2}r} \leq \frac{1}{|x| + |y|} \leq \frac{1}{r}, \quad r = \sqrt{x^2 + y^2}.$$

From this it is immediate that $\frac{1}{\sqrt{2}r} > R$ (i.e., $r < \frac{1}{\sqrt{2}R}$) implies $\frac{1}{|x| + |y|} > R$, and $\frac{1}{r} < \epsilon$ (i.e., $r > \frac{1}{\epsilon}$) implies $\frac{1}{|x| + |y|} < \epsilon$. Thus $\delta = \frac{1}{\sqrt{2}R}$ and $R = \frac{1}{\epsilon}$ can serve as responses for R and ϵ , respectively.

- In the first solution we have essentially worked with squares $C_r(0, 0) = \{(x, y) \in \mathbb{R}^2; |x| < r, |y| < r\}$ instead of the disks $B_r(0, 0)$. This is possible, since every disk contains a square (possibly with smaller r) and vice versa. The same applies of course to disks and squares with arbitrary center, and also to their higher-dimensional analogues (balls and n -dimensional cubes).

Problem

Repeat the preceding exercise for $h: D \rightarrow \mathbb{R}$ defined by $h(x, y) = \frac{x-y}{x+y}$. (The domain D remains the same.)

Solution

For points $(x_0, y_0) \in D$ and on the positive coordinate axes we have, as before,

$$\lim_{(x,y) \rightarrow (x_0, y_0)} h(x, y) = \lim_{(x,y) \rightarrow (x_0, y_0)} \frac{x - y}{x + y} = \frac{x_0 - y_0}{x_0 + y_0}.$$

The limits for $(x, y) \rightarrow (0, 0)$ and $|(x, y)| \rightarrow \infty$ don't exist in this case, as the behaviour of $h(x, y)$ on the lines $y = x$ and $y = 2x$, shows:

$$h(x, x) = 0, \quad h(x, 2x) = \frac{x - 2x}{x + 2x} = -\frac{1}{3}.$$

Balls (i.e., disks) $B_r(0, 0)$ around the origin contain points of both lines, implying that both $\lim_{x \downarrow 0} h(x, x)$ and $\lim_{x \downarrow 0} h(x, 2x)$ should be equal to the putative limit $\lim_{(x,y) \rightarrow (0,0)} h(x, y)$. But these limits are 0 and $-1/3$, respectively. Contradiction!

For $|(x, y)| \rightarrow \infty$ the same reasoning applies, since the sets $\mathbb{R}^2 \setminus B_R(0, 0)$ contain points of both lines as well.

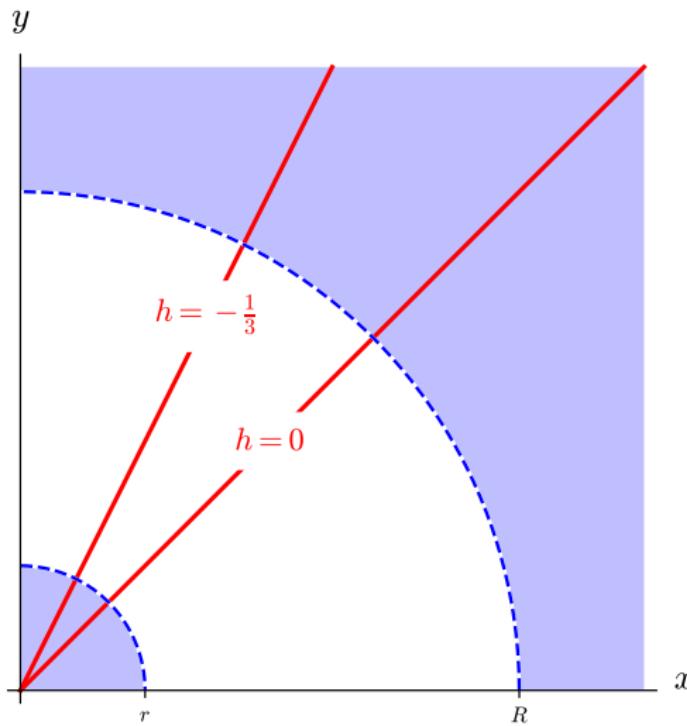


Figure: Non-existence of the limits of $h(x, y)$ for $(x, y) \rightarrow (0, 0)$ and $|(x, y)| \rightarrow \infty$

All sets $B_r(0, 0) \cap D$, $r > 0$, and $D \setminus B_R(0, 0)$, $R > 0$, contain points (x, y) with $h(x, y) = 0$ and points (x, y) with $h(x, y) = -1/3$.

The Last Example

Problem

Decide whether the limits $\lim_{\substack{(x,y) \rightarrow (0,0) \\ x>0, y>0}} \frac{1}{xy}$ and $\lim_{\substack{|(x,y)| \rightarrow \infty \\ x>0, y>0}} \frac{1}{xy}$ exist?

Solution

This is the same problem for the function $f: D \rightarrow \mathbb{R}$, $(x, y) \mapsto \frac{1}{xy}$ with D as before. We consider only the case $|(x, y)| \rightarrow \infty$ and leave the case $(x, y) \rightarrow (0, 0)$ as an exercise.

Here we have on any line $y = mx + t$, $m > 0$,

$$f(x, mx + t) = \frac{1}{x(mx + t)} = \frac{1}{mx^2 + tx} \rightarrow 0 \quad \text{for } x \rightarrow +\infty,$$

but this does not imply that the limit $\lim_{\substack{|(x,y)| \rightarrow \infty \\ x>0, y>0}} \frac{1}{xy}$ is zero.

In fact the limit does not exist, since the contours of f are the curves $y = 1/(kx)$, $k > 0$ (for $k < 0$ the contours are empty), and every set $D \setminus B_R(0, 0)$ meets these curves (all of them!).

We could also argue that $\lim_{x \rightarrow +\infty} f(x, \frac{1}{x}) = 1 \neq \lim_{x \rightarrow +\infty} f(x, \frac{2}{x}) = \frac{1}{2}$.

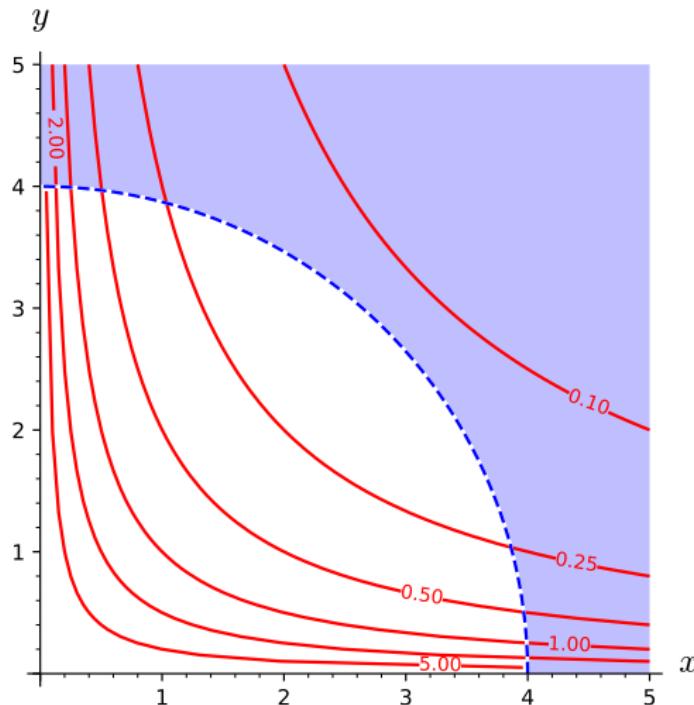


Figure: Non-existence of the limit $\lim_{\substack{|(x,y)| \rightarrow \infty \\ x>0, y>0}} \frac{1}{xy}$

Every set $D \setminus B_R(0, 0)$ contains points of every contour of f . (The picture shows this for $R = 4$ and a few selected contours.)

The examples $(x, y) \mapsto \frac{1}{x+y}$ and $(x, y) \mapsto \frac{1}{xy}$ also show that for a function $f: D \rightarrow \mathbb{R}$ (with D the same as in the previous examples)

$$\lim_{x \rightarrow +\infty} f(x, y_0) = 0 \quad \text{and} \quad \lim_{y \rightarrow +\infty} f(x_0, y) = 0$$

may hold for all $x_0, y_0 > 0$ despite the fact that $\lim_{|(x,y)| \rightarrow \infty} f(x, y)$ doesn't exist.

For the function $h(x, y) = \frac{x-y}{x+y}$ we can use the behavior on the coordinate lines to disprove in another way the existence of the limit $\lim_{|(x,y)| \rightarrow \infty} h(x, y)$: Since

$$\lim_{x \rightarrow +\infty} \frac{x - y_0}{x + y_0} = 1, \quad \lim_{y \rightarrow +\infty} \frac{x_0 - y}{x_0 + y} = -1$$

are different, the limit $\lim_{|(x,y)| \rightarrow \infty} \frac{x-y}{x+y}$ cannot exist. (If it would, these two limits should be equal to it.)

Math 241
Calculus III

Thomas
Honold

Introduction

Differentiable
maps

Partial
Derivatives

Further
Concepts

Directional
Derivatives

The Gradient

Tangent Spaces

True Meaning of
Differentials

The Chain Rule

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Introduction

2 Differentiable maps

3 Partial Derivatives

4 Further Concepts

Directional Derivatives

The Gradient

Tangent Spaces

True Meaning of Differentials

The Chain Rule

Introduction

Differentiable
maps

Partial
Derivatives

Further
Concepts

Directional
Derivatives

The Gradient
Tangent Spaces

True Meaning of
Differentials

The Chain Rule

Introduction

Differentiable
maps

Partial
Derivatives

Further
Concepts

Directional
Derivatives

The Gradient

Tangent Spaces

True Meaning of
Differentials

The Chain Rule

Today's Lecture: Differentiation of Multivariable Functions

Introduction

The following examples, of which you should know the first one, illustrate the main purpose of differentiation:

Local approximation by a linear map

Example ($f(x) = x^3$)

Considering $x_0 \in \mathbb{R}$ as fixed, real numbers close to x_0 have the form $x = x_0 + h$ with $|h|$ small.

$$\begin{aligned}f(x) &= f(x_0 + h) = (x_0 + h)^3 = x_0^3 + 3x_0^2h + 3x_0h^2 + h^3 \\&= f(x_0) + \text{something linear in } h + R(h) \\&\approx f(x_0) + \text{something linear in } h\end{aligned}$$

with approximation error $R(h) = 3x_0h^2 + h^3$.

The error satisfies $R(h)/h = 3x_0h + h^2 \rightarrow 0$ for $h \rightarrow 0$.

This is exactly what we need to show that

$$\frac{f(x_0 + h) - f(x_0)}{h} = 3x_0^2 + \frac{R(h)}{h} \rightarrow 3x_0^2 \quad \text{for } h \rightarrow 0, \quad \text{i.e.,} \quad f'(x_0) = 3x_0^2.$$

Using little-o notation, $\lim_{h \rightarrow 0} R(h)/h = 0$ is expressed as $R(h) = o(h)$, and hence the approximation in the previous example as

$$f(x_0 + h) = f(x_0) + \text{something linear in } h + o(h).$$

Now comes the first multivariable example.

Example ($f(x, y) = x^3 - 3xy^2$)

Here the displacement is of the form $\mathbf{h} = (h_1, h_2)$ and we get (dropping the index '0' in the fixed point (x, y) considered)

$$\begin{aligned} f(x + h_1, y + h_2) &= (x + h_1)^3 - 3(x + h_1)(y + h_2)^2 \\ &= x^3 + 3x^2h_1 + 3xh_1^2 + h_1^3 - 3xy^2 - 6xyh_2 - 3xh_2^2 - 3h_1y^2 - 6h_1yh_2 - 3h_1h_2^2 \\ &= x^3 - 3xy^2 + 3(x^2 - y^2)h_1 - 6xyh_2 + 3xh_1^2 - 6yh_1h_2 - 3xh_2^2 + h_1^3 - 3h_1h_2^2 \\ &= f(x, y) + \text{something linear in } \mathbf{h} + R(\mathbf{h}) \end{aligned}$$

Since every monomial appearing in $R(\mathbf{h})$ has degree ≥ 2 and

$$\frac{|h_i|}{|\mathbf{h}|} = \frac{|h_i|}{\sqrt{h_1^2 + h_2^2}} \leq 1 \quad \text{for } i = 1, 2,$$

we have $R(\mathbf{h})/|\mathbf{h}| \rightarrow 0$ for $\mathbf{h} \rightarrow \mathbf{0} \in \mathbb{R}^2$ (i.e., $h_1 \rightarrow 0$ and $h_2 \rightarrow 0$ in \mathbb{R}).

[Introduction](#)[Differentiable
maps](#)[Partial
Derivatives](#)[Further
Concepts](#)[Directional
Derivatives](#)[The Gradient](#)[Tangent Spaces](#)[True Meaning of
Differentials](#)[The Chain Rule](#)

Using little-o notation, we can express $\lim_{\mathbf{h} \rightarrow \mathbf{0}} R(\mathbf{h}) / |\mathbf{h}| = 0$ as $R(\mathbf{h}) = o(\mathbf{h})$, and hence the approximation in the previous example as

$$f((x, y) + \mathbf{h}) = f(x, y) + \text{something linear in } \mathbf{h} + o(\mathbf{h}).$$

Example ($V(x, y, z) = xyz$)

The function $V(x, y, z)$ returns the volume of a cuboid with side lengths x, y, z . We have

$$\begin{aligned} V(x + h_1, y + h_2, z + h_3) - V(x, y, z) &= (x + h_1)(y + h_2)(z + h_3) - xyz \\ &= yzh_1 + xzh_2 + xyh_3 + zh_1h_2 + yh_1h_3 + xh_2h_3 + h_1h_2h_3 \\ &\approx yzh_1 + xzh_2 + xyh_3 \end{aligned}$$

with an error of order $o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$, and thus substantially smaller than the maximum of $|h_1|, |h_2|, |h_3|$.

This says that a small change/error in the input of V , represented by $\mathbf{h} = (h_1, h_2, h_3)$, “propagates” to a change/error of approximately $yzh_1 + xzh_2 + xyh_3$ in the output of V , i.e., in the computed volume.

Example (squaring map)

The squaring map (real representation) was defined as $s(x, y) = (x^2 - y^2, 2xy)$ for $(x, y) \in \mathbb{R}^2$.

Using column vectors, its linear approximation in (x, y) is

$$\begin{aligned}s(x + h_1, y + h_2) &= \begin{pmatrix} (x + h_1)^2 - (y + h_2)^2 \\ 2(x + h_1)(y + h_2) \end{pmatrix} \\&= \begin{pmatrix} x^2 - y^2 + 2xh_1 - 2yh_2 + h_1^2 - h_2^2 \\ 2xy + 2yh_1 + 2xh_2 + 2h_1h_2 \end{pmatrix} \\&= \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix} + \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} h_1^2 - h_2^2 \\ 2h_1h_2 \end{pmatrix} \\&= s(x, y) + \text{something linear in } \mathbf{h} + R(\mathbf{h})\end{aligned}$$

Here $R(\mathbf{h})$ is vector-valued, but from

$$|R(\mathbf{h})| \leq \sqrt{2} \max\{|h_1^2 - h_2^2|, |2h_1h_2|\}$$

we still get $|R(\mathbf{h})| / |\mathbf{h}| \rightarrow 0$ for $\mathbf{h} \rightarrow \mathbf{0}$ in the same way as before, i.e., $R(\mathbf{h}) = o(\mathbf{h})$.

In the complex world, using $z = (x, y) = x + yi$,
 $h = (h_1, h_2) = h_1 + h_2i$, the approximation just obtained reads

$$(z + h)^2 = z^2 + 2zh + h^2 = z^2 + 2zh + o(h),$$

so that the approximating linear map is multiplication by
 $2z = (z^2)'$. This is no coincidence.

Example ($f(x, y) = e^{xy}$)

This example has been included, because it is genuinely non-polynomial. Here we can argue as follows:

$$\begin{aligned} e^{(x+h_1)(y+h_2)} - e^{xy} &= e^{xy + yh_1 + xh_2 + h_1 h_2} - e^{xy} = e^{xy} (e^{yh_1 + xh_2 + h_1 h_2} - 1) \\ &= e^{xy} (yh_1 + xh_2 + \text{terms of degree } \geq 2 \text{ in } \mathbf{h}) \\ &= (ye^{xy})h_1 + (xe^{xy})h_2 + o(\mathbf{h}). \end{aligned}$$

Without going into details, this follows by expanding

$$e^{yh_1 + xh_2 + h_1 h_2} = \sum_{k=0}^{\infty} \frac{(yh_1 + xh_2 + h_1 h_2)^k}{k!}$$

into a double series and suitably rearranging terms.

Notes on the preceding examples

- According to the subsequent definition of differentiable maps, all five functions are differentiable everywhere (the case of $V(x, y, z)$ requires further justification!), and the differential of the function in a particular point is the linear map which sends \mathbf{h} (a vector in \mathbb{R} , \mathbb{R}^2 , resp., \mathbb{R}^3) to the blue expression stated in the approximation formula.
- Be sure to understand that a given function f can be associated with many linear maps in this way, one for each point \mathbf{x} (denoted by x_0 , (x, y) , or (x, y, z) in the examples) of its domain. The differential df of f is the map (non-linear in general) that sends \mathbf{x} to the linear map used at \mathbf{x} , viz. $df(\mathbf{x})$.
- For the explicit computation of the linear maps involved we need a formula for obtaining their “coefficients”, i.e., the entries of their representing matrices. In the examples the underlying pattern can be already seen, e.g., think how the coefficients of $(h_1, h_2) \mapsto (ye^{xy})h_1 + (xe^{xy})h_2$ arise from e^{xy} . In this regard also note that the coefficient of h_1 , say, can be obtained by setting $h_2 = 0$ and considering the resulting one-dimensional approximation problem.

Definition

Differentiable Maps

Suppose $D \subseteq \mathbb{R}^n$ and $\mathbf{x}_0 \in D$. The point \mathbf{x}_0 is said to be an *inner point* of D if D contains a ball of positive radius around \mathbf{x}_0 , i.e., there exists $r > 0$ such that $|\mathbf{x} - \mathbf{x}_0| < r$ implies $\mathbf{x} \in D$. The set of inner points of D is denoted by D° .

The remaining points of D are boundary points (but the boundary ∂D may contain points in $\mathbb{R}^n \setminus D$).

Now comes the most important definition of Calculus III.

Definition

Suppose $f: D \rightarrow \mathbb{R}^m$ is a map with domain $D \subseteq \mathbb{R}^n$ and \mathbf{x}_0 is an inner point of D . The map f is said to be *differentiable* at \mathbf{x}_0 if there exists a linear map $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + L(\mathbf{h}) + o(\mathbf{h}) \quad \text{for } \mathbf{h} \rightarrow \mathbf{0}. \quad (\text{TD})$$

If this is the case then the linear map L , which is uniquely determined, is called the *differential* of f at \mathbf{x}_0 and denoted by $df(\mathbf{x}_0)$.

Notes

- If you are uncomfortable with the little-o notation, take the following equivalent formulation of (TD):

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) - L(\mathbf{h})}{|\mathbf{h}|} = \mathbf{0} \in \mathbb{R}^m$$

Both formulations say in particular that for $\mathbf{h} \rightarrow \mathbf{0}$ the error term of the linear approximation $f(\mathbf{x}_0 + \mathbf{h}) \approx f(\mathbf{x}_0) + L(\mathbf{h})$ is substantially smaller than \mathbf{h} in length.

- The linear map L in (TD) is indeed uniquely determined: If L_1 and L_2 satisfy (TD), we must have $L_1(\mathbf{h}) - L_2(\mathbf{h}) = o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$. Now let $\mathbf{h} = h\mathbf{v}$ with $\mathbf{v} \in \mathbb{R}^n$ a fixed nonzero vector. Since $h\mathbf{v} \rightarrow \mathbf{0}$ for $h \rightarrow 0$, the quotient

$$\frac{|L_1(h\mathbf{v}) - L_2(h\mathbf{v})|}{|h\mathbf{v}|} = \frac{|h(L_1(\mathbf{v}) - L_2(\mathbf{v}))|}{|h\mathbf{v}|} = \frac{|L_1(\mathbf{v}) - L_2(\mathbf{v})|}{|\mathbf{v}|}$$

tends to 0 for $h \rightarrow 0$, which can't be unless $L_1(\mathbf{v}) = L_2(\mathbf{v})$.

Question: Where have we used that \mathbf{x}_0 is an inner point of D ? *Answer:* To have $\mathbf{x}_0 + h\mathbf{v} \in D$ for small $|h|$.

Notes cont'd

- Linear maps $L: \mathbb{R} \rightarrow \mathbb{R}$ have the form $L(h) = ah$ for some $a \in \mathbb{R}$, since $L(h) = L(h1) = hL(1) = L(1)h$, i.e., $a = L(1)$. Hence in the case $m = n = 1$ (TD) reduces to the familiar

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0) - ah}{h} = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} - a = 0,$$

which just says $f'(x_0) = a$. In other words, differentiable functions $f: I \rightarrow \mathbb{R}$, $I \subseteq \mathbb{R}$, in the old sense remain differentiable in the new sense and have differential $x_0 \mapsto df(x_0): \mathbb{R} \rightarrow \mathbb{R}$, $h \mapsto f'(x_0)h$.

For curves $f: I \rightarrow \mathbb{R}^n$ the same is true (mutatis mutandis).

- As we have seen, linear maps $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$ have the form $L(\mathbf{x}) = \mathbf{Ax}$ for some (uniquely determined) matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. Hence the condition in (TD) can be rephrased as: There exists a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ such that

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \mathbf{Ah} + o(\mathbf{h}) \quad \text{for } \mathbf{h} \rightarrow \mathbf{0}.$$

The matrix \mathbf{A} , which is uniquely determined by the previous note, is called *Jacobi(an) matrix* of f at \mathbf{x}_0 and denoted by $\mathbf{J}_f(\mathbf{x}_0)$.

Notes cont'd

- A vector-valued function $f = (f_1, \dots, f_m)$ is differentiable at \mathbf{x}_0 iff each coordinate function f_i is differentiable at \mathbf{x}_0 . This is due to the fact that limits of vector-valued functions can be computed coordinate-wise.
- Finally a note on the various sets \overline{D} , ∂D , D' , D° defined for any set $D \subseteq \mathbb{R}^n$. In what follows, \uplus denotes the *disjoint union* of sets, i.e., $M = S \uplus T$ means $M = S \cup T$ and $S \cap T = \emptyset$.

For any D we have $D^\circ \subseteq D \subseteq \overline{D}$, $D^\circ \subseteq D' \subseteq \overline{D}$, $\overline{D} = D \cup D' = D \cup \partial D$, and the decomposition

$$\begin{aligned}\mathbb{R}^n &= D^\circ \uplus \partial D \uplus (\mathbb{R}^n \setminus D)^\circ \\ &= \overline{D} \uplus (\mathbb{R}^n \setminus D)^\circ && (\overline{D} = D^\circ \uplus \partial D) \\ &= D^\circ \uplus \overline{\mathbb{R}^n \setminus D}. && \text{(by symmetry)}\end{aligned}$$

The boundary $\partial D = \partial(\mathbb{R}^n \setminus D)$ consists of those points \mathbf{x} for which every ball around \mathbf{x} contains points of D as well as points of $\mathbb{R}^n \setminus D$. Points in $D \cap \partial D$ must be accumulation points of $\mathbb{R}^n \setminus D$ but need not be accumulation points of D .

The Five Examples reconsidered

- ① $f(x) = x^3$ is differentiable in \mathbb{R} with differential

$$df(x): \mathbb{R} \rightarrow \mathbb{R}, h \mapsto 3x^2h.$$

- ② $f(x, y) = x^3 - 3xy^2$ is differentiable in \mathbb{R}^2 with differential

$$df(x, y): \mathbb{R}^2 \rightarrow \mathbb{R}, (h_1, h_2) \mapsto 3(x^2 - y^2)h_1 - 6xyh_2.$$

- ③ $V(x, y, z) = xyz$ is differentiable in \mathbb{R}^3 with differential

$$dV(x, y, z): \mathbb{R}^3 \rightarrow \mathbb{R}, (h_1, h_2, h_3) \mapsto yzh_1 + xzh_2 + xyh_3.$$

- ④ $s(x, y) = (x^2 - y^2, 2xy)$ is differentiable in \mathbb{R}^2 with differential

$$ds(x, y): \mathbb{R}^2 \rightarrow \mathbb{R}^2, \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \mapsto \underbrace{\begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}}_{\mathbf{J}_s(x, y)} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

- ⑤ $f(x, y) = e^{xy}$ is differentiable in \mathbb{R}^2 with differential

$$df(x, y): \mathbb{R}^2 \rightarrow \mathbb{R}, (h_1, h_2) \mapsto ye^{xy}h_1 + xe^{xy}h_2.$$

Introduction

Differentiable
maps

Partial
Derivatives

Further
Concepts

Directional
Derivatives

The Gradient

Tangent Spaces

True Meaning of
Differentials

The Chain Rule

Further Examples

Example (linear maps)

Consider a linear map $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{x} \mapsto \mathbf{Ax}$. Here we have

$$f(\mathbf{x}_0 + \mathbf{h}) = \mathbf{A}(\mathbf{x}_0 + \mathbf{h}) = \mathbf{Ax}_0 + \mathbf{Ah},$$

and there is no remainder term. This implies that f is differentiable at \mathbf{x}_0 with differential $df(\mathbf{x}_0): \mathbf{h} \mapsto \mathbf{Ah}$.

In other words, a linear map f is differentiable everywhere and the differential $df(\mathbf{x})$ coincides with f at any point $\mathbf{x} \in \mathbb{R}^n$.

For affine maps $f(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$ the same is true (except that the differential doesn't coincide with f if $\mathbf{b} \neq \mathbf{0}$), because the constant vector \mathbf{b} does not matter for differentiation.

Example (quadratic forms)

A *quadratic form* on \mathbb{R}^n is a map $q: \mathbb{R}^n \rightarrow \mathbb{R}$ of the form $q(\mathbf{x}) = \sum_{1 \leq i \leq j \leq n} q_{ij}x_i x_j$ (i.e., a homogeneous polynomial of degree 2).

Setting $a_{ii} = q_{ii}$ and $a_{ij} = a_{ji} = q_{ij}/2$ for $i < j$ and viewing $\mathbf{x} \in \mathbb{R}^n$ as a column vector, we have

$$q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} \quad \text{with } \mathbf{A} \in \mathbb{R}^{n \times n}, \mathbf{A} = \mathbf{A}^T.$$

This representation is best suited for differentiating:

$$\begin{aligned} q(\mathbf{x} + \mathbf{h}) &= (\mathbf{x} + \mathbf{h})^T \mathbf{A} (\mathbf{x} + \mathbf{h}) \\ &= \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{h}^T \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{A} \mathbf{h} + \mathbf{h}^T \mathbf{A} \mathbf{h} \\ &= q(\mathbf{x}) + 2\mathbf{x}^T \mathbf{A} \mathbf{h} + \mathbf{h}^T \mathbf{A} \mathbf{h} \quad (\text{since } \mathbf{A} = \mathbf{A}^T) \end{aligned}$$

$\mathbf{h}^T \mathbf{A} \mathbf{h} = q(\mathbf{h})$ is a sum of terms $q_{ij} h_i h_j$, and we have

Example (cont'd)

$$\frac{|q_{ij}h_ih_j|}{|\mathbf{h}|} \leq |q_{ij}| |h_j| \leq |q_{ij}| |\mathbf{h}|.$$

This shows $\mathbf{h}^T \mathbf{A} \mathbf{h} = o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$ and hence that q is differentiable everywhere with differential

$$dq(\mathbf{x}): \mathbb{R}^n \rightarrow \mathbb{R}, \mathbf{h} \mapsto 2\mathbf{x}^T \mathbf{A} \mathbf{h} = 2(\mathbf{A}\mathbf{x})^T \mathbf{h}.$$

In other words, the differential of q at $\mathbf{x} \in \mathbb{R}^n$ is “taking the dot product with the column vector $2(\mathbf{A}\mathbf{x})$ ”, which represents a linear map.

In contrast with the linear case, however, the differential $dq(\mathbf{x})$ of a quadratic form depends on the particular point \mathbf{x} .

As a concrete example, in the two-variable case

$$q(x, y) = ax^2 + 2bxy + cy^2 = \begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

we have $dq(x, y)(h_1, h_2) = 2(ax + by)h_1 + 2(bx + cy)h_2$.

Example

The length function $\mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{x} \mapsto |\mathbf{x}|$ is differentiable at any point $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ but not differentiable at the origin.

We restrict ourselves to the case $n = 2$. (The proof in the general case can be easily inferred from this.)

First we show that $\mathbf{x} \mapsto |\mathbf{x}|$ is not differentiable at $\mathbf{0} = (0, 0)$.

For the special choice $\mathbf{h} = (h, 0) = h\mathbf{e}_1$ we have $L(\mathbf{h}) = hL(\mathbf{e}_1)$ but

$$|\mathbf{0} + \mathbf{h}| - |\mathbf{0}| = |\mathbf{h}| = |(h, 0)| = |h| = \pm h,$$

which cannot be approximated within $o(h)$ by a single linear map.
(We should define $L(\mathbf{e}_1) = 1$ for $h > 0$, but $L(\mathbf{e}_1) = -1$ for $h < 0$.)

Now consider $\mathbf{x} = (x_1, x_2) \neq (0, 0)$. Here we must estimate

$$\begin{aligned} |\mathbf{x} + \mathbf{h}| - |\mathbf{x}| &= \sqrt{(x_1 + h_1)^2 + (x_2 + h_2)^2} - \sqrt{x_1^2 + x_2^2} \\ &= \sqrt{x_1^2 + x_2^2 + 2(x_1 h_1 + x_2 h_2) + h_1^2 + h_2^2} - \sqrt{x_1^2 + x_2^2} \\ &= \frac{2x_1 h_1 + 2x_2 h_2 + h_1^2 + h_2^2}{\sqrt{x_1^2 + x_2^2 + 2(x_1 h_1 + x_2 h_2) + h_1^2 + h_2^2} + \sqrt{x_1^2 + x_2^2}} \end{aligned}$$

Example (cont'd)

This has the form

$$|\mathbf{x} + \mathbf{h}| - |\mathbf{x}| = \frac{2x_1 h_1 + 2x_2 h_2 + h_1^2 + h_2^2}{g(h_1, h_2)},$$

where $g(h_1, h_2)$ is continuous at $(0, 0)$ and $g(0, 0) = 2\sqrt{x_1^2 + x_2^2} \neq 0$.

We now show that the linear map $L: \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$\begin{aligned} L(\mathbf{h}) &= L(h_1, h_2) = \frac{2x_1 h_1 + 2x_2 h_2}{g(0, 0)} \\ &= \frac{x_1}{\sqrt{x_1^2 + x_2^2}} \cdot h_1 + \frac{x_2}{\sqrt{x_1^2 + x_2^2}} \cdot h_2 \end{aligned}$$

has the required approximation property:

$$\begin{aligned} |\mathbf{x} + \mathbf{h}| - |\mathbf{x}| - L(\mathbf{h}) &= (2x_1 h_1 + 2x_2 h_2) \left(\frac{1}{g(h_1, h_2)} - \frac{1}{g(0, 0)} \right) + \frac{h_1^2 + h_2^2}{g(h_1, h_2)} \\ &= O(|\mathbf{h}|) o(1) + O(|\mathbf{h}|^2) = o(|\mathbf{h}|) = o(\mathbf{h}), \end{aligned}$$

as claimed.

Partial derivatives

How to get the entries of the Jacobi matrix

In the preceding example we have obtained the Jacobi matrix of the length function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $\mathbf{x} \mapsto \sqrt{x_1^2 + x_2^2}$:

$$\mathbf{J}_f(\mathbf{x}) = \begin{pmatrix} \frac{x_1}{\sqrt{x_1^2 + x_2^2}}, & \frac{x_2}{\sqrt{x_1^2 + x_2^2}} \end{pmatrix}.$$

Question

How to obtain the entries of $\mathbf{J}_f(\mathbf{x})$, and hence the differential $df(\mathbf{x}): \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{h} \mapsto \mathbf{J}_f(\mathbf{x})\mathbf{h}$, in general?

The answer uses only Calculus I and can be found by inspecting our earlier examples. It involves the so-called partial derivatives of f .

Introduction

Differentiable
mapsPartial
DerivativesFurther
ConceptsDirectional
Derivatives

The Gradient

Tangent Spaces

True Meaning of
Differentials

The Chain Rule

Example (squaring map continued)

We have seen that $s(x, y) = (x^2 - y^2, 2xy)$ is differentiable in \mathbb{R}^2 with differential

$$ds(x, y)(h_1, h_2) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}.$$

The entries of the Jacobi matrix can be obtained without the (rather complicated) expansion step by setting $h_1 = 0$, respectively, $h_2 = 0$ in the approximation formula

$$s(x + h_1, y + h_2) = s(x, y) + \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + o((h_1, h_2)).$$

For example, setting $h_2 = 0$ gives

$$s(x + h_1, y) = s(x, y) + h_1 \begin{pmatrix} 2x \\ 2y \end{pmatrix} + o(h_1).$$

$$\implies \begin{pmatrix} 2x \\ 2y \end{pmatrix} = \lim_{h_1 \rightarrow 0} \frac{s(x + h_1, y) - s(x, y)}{h_1} = \frac{d}{dx}(x \mapsto s(x, y)).$$

Definition

Let $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, be a real-valued function. The *partial derivative* of f with respect to the variable x_j , $1 \leq j \leq n$, is the function that assigns to $\mathbf{x} = (x_1, \dots, x_n) \in D$ the derivative of $t \mapsto (x_1, \dots, x_{j-1}, t, x_{j+1}, \dots, x_n)$ at $t = x_j$. The partial derivatives of f are denoted by f_{x_j} or $\frac{\partial f}{\partial x_j}$.

Notes

- According to the definition of derivatives in Calculus I we have

$$f_{x_j}(\mathbf{x}) = \frac{\partial f}{\partial x_j}(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h}.$$

- The partial derivatives of f are obtained by viewing f as a function of one variable x_j (keeping all other variables fixed) and applying the usual rules for computing derivatives learned in Calculus I to this function (resp., to its coordinate functions).

Notes cont'd

- The (maximal) domain D_j of f_{x_j} consists of all $\mathbf{x} \in D$ for which the limit $\lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h}$ exists.
- If $\mathbf{x} \in D$ is such that all partial derivatives $f_{x_j}(\mathbf{x})$, $1 \leq j \leq n$, exist (i.e., $\mathbf{x} \in \bigcap_{j=1}^n D_j$), we say that f is *partially differentiable* at \mathbf{x} (and partially differentiable per se if this is true for all $\mathbf{x} \in D$).
- We will see in a moment that differentiability implies partial differentiability (at a point $\mathbf{x} \in D$) but not conversely. To make this difference clear, differentiability is also referred to as "*total* differentiability".
- Partial derivatives $\frac{\partial f}{\partial x_j}$ for vectorial functions $f = (f_1, \dots, f_m)$ are defined in the same way. The rules for computing limits of vectorial functions imply that $\frac{\partial f}{\partial x_j} = \left(\frac{\partial f_1}{\partial x_j}, \dots, \frac{\partial f_m}{\partial x_j} \right)$. Thus, anticipating Part (2) of the theorem on the next slide, we can say that the entries of $\mathbf{J}_f(\mathbf{x})$ are the scalar partial derivatives $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$, and the columns of $\mathbf{J}_f(\mathbf{x})$ are the vectorial partial derivatives $\frac{\partial f}{\partial x_j}(\mathbf{x})$. (Recall that we should consider a vectorial function f as a column vector $(f_1, \dots, f_m)^T$.)

Theorem

Let $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, be a function with coordinate functions f_1, \dots, f_m and $\mathbf{x} \in D^\circ$.

- ① If f is differentiable at \mathbf{x} then f is continuous at \mathbf{x} .
- ② If f is differentiable at \mathbf{x} then the partial derivatives $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ exist for $1 \leq i \leq m$, $1 \leq j \leq n$, and

$$\mathbf{J}_f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \frac{\partial f_m}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix}.$$

- ③ Conversely, if all partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist near \mathbf{x} (i.e., f is partially differentiable in some ball around \mathbf{x}) and are continuous at \mathbf{x} , then f is differentiable at \mathbf{x} .

Proof.

(1) With $L = df(\mathbf{x})$ we have

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + L(\mathbf{h}) + o(\mathbf{h}) = f(\mathbf{x}) + L(\mathbf{h}) + o(1)$$

for $\mathbf{h} \rightarrow \mathbf{0}$, and it remains to show that

$$\lim_{\mathbf{h} \rightarrow \mathbf{0} \in \mathbb{R}^n} L(\mathbf{h}) = \mathbf{0} \in \mathbb{R}^m.$$

This amounts to L being continuous at $\mathbf{0} \in \mathbb{R}^n$ and is easily verified from the matrix representation $L(\mathbf{h}) = \mathbf{A}\mathbf{h}$, $\mathbf{A} = \mathbf{J}_f(\mathbf{x})$.

(2) Specializing the approximation property to $\mathbf{h} = h\mathbf{e}_j$, $h \in \mathbb{R}$, gives

$$f(\mathbf{x} + h\mathbf{e}_j) = f(\mathbf{x}) + L(h\mathbf{e}_j) + o(h\mathbf{e}_j) = f(\mathbf{x}) + hL(\mathbf{e}_j) + o(h).$$

Subtracting $f(\mathbf{x})$ and dividing by h gives further

$$\frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h} = L(\mathbf{e}_j) + o(1), \quad \text{i.e.} \quad \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_j) - f(\mathbf{x})}{h} = L(\mathbf{e}_j).$$

Passing to the coordinate functions f_i then shows that $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ exists for all i, j and forms the (i, j) entry of $\mathbf{J}_f(\mathbf{x})$. □

Proof cont'd.

(3) We assume $n = 2$ and $m = 1$ for simplicity (but the proof in the general case can be easily inferred from this case).

For sufficiently small $\mathbf{h} = (h_1, h_2)$ we have

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) &= f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) \\ &= f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2 + h_2) + f(x_1, x_2 + h_2) - f(x_1, x_2). \end{aligned}$$

Applying the Mean Value Theorem from Calculus I to the functions $g_1(s) = f(s, x_2 + h_2)$ and $g_2(t) = f(x_1, t)$ shows the existence of $\xi_1 \in (x_1, x_1 + h_1)$, $\xi_2 \in (x_2, x_2 + h_2)$ such that

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) &= g'_1(\xi_1)h_1 + g'_2(\xi_2)h_2 \\ &= \frac{\partial f}{\partial x_1}(\xi_1, x_2 + h_2)h_1 + \frac{\partial f}{\partial x_2}(x_1, \xi_2)h_2. \end{aligned}$$

Finally, the continuity of $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$ at \mathbf{x} gives for $\mathbf{h} \rightarrow \mathbf{0}$ the estimate

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) &= \left(\frac{\partial f}{\partial x_1}(x_1, x_2) + o(1) \right) h_1 + \left(\frac{\partial f}{\partial x_2}(x_1, x_2) + o(1) \right) h_2 \\ &= \frac{\partial f}{\partial x_1}(x_1, x_2)h_1 + \frac{\partial f}{\partial x_2}(x_1, x_2)h_2 + \underbrace{o(1)h_1 + o(1)h_2}_{=o(\mathbf{h})}. \end{aligned}$$

Introduction

Differentiable
mapsPartial
DerivativesFurther
ConceptsDirectional
Derivatives

The Gradient

Tangent Spaces

True Meaning of
Differentials

The Chain Rule

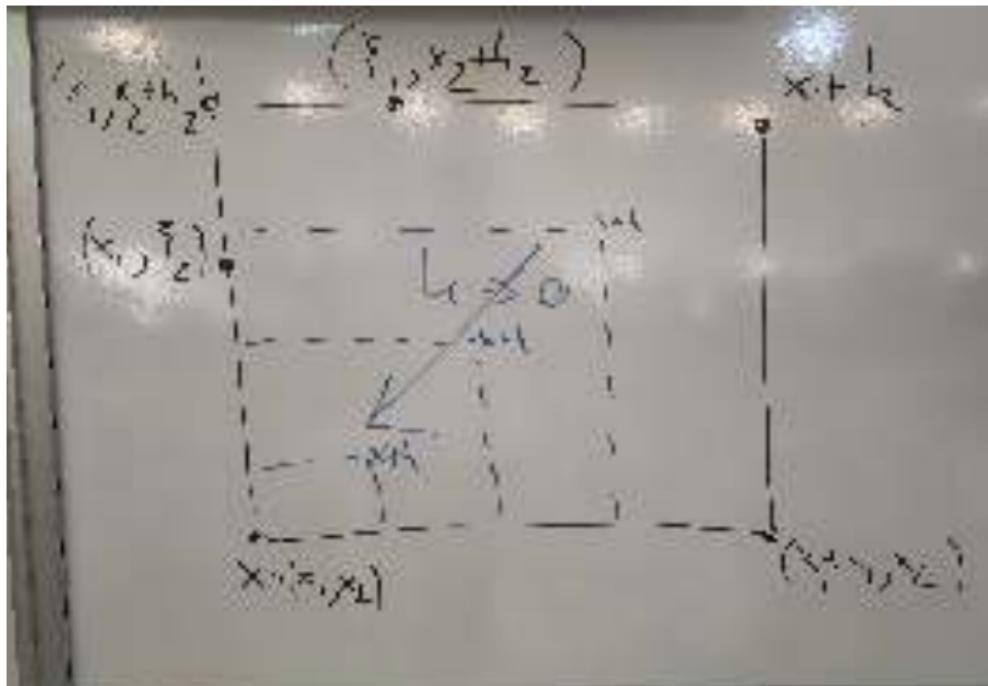


Figure: Illustration of the proof of Part 3 of the theorem:
For $\mathbf{h} \rightarrow \mathbf{0}$ the rectangle shrinks to its lower left vertex \mathbf{x} , and
hence also the intermediate points $(\xi_1, x_2 + h_2)$, (x_1, ξ_2) converge
to \mathbf{x} .

Afternote

After class I realized that students have some problems with the use of Big-O/little-o notation in the wider setting of vectorial and multivariable functions, including mixed-dimension cases. Here is the definition in more detail:

Suppose $D \subseteq \mathbb{R}^n$, $f: D \rightarrow \mathbb{R}^{m_1}$, $g: D \rightarrow \mathbb{R}^{m_2}$, and $\mathbf{x}_0 \in D^\circ$.

- 1 We say $f(\mathbf{x}) = O(g(\mathbf{x}))$ for $\mathbf{x} \rightarrow \mathbf{x}_0$ if there exist constants $C, \delta > 0$ such that $|f(\mathbf{x})| \leq C |g(\mathbf{x})|$ for all $\mathbf{x} \in B_\delta(\mathbf{x}_0) \setminus \{\mathbf{x}_0\}$
- 2 We say $f(\mathbf{x}) = o(g(\mathbf{x}))$ for $\mathbf{x} \rightarrow \mathbf{x}_0$ if for every $\epsilon > 0$ there exists $\delta = \delta(\epsilon) > 0$ such that $|f(\mathbf{x})| \leq \epsilon |g(\mathbf{x})|$ for all $\mathbf{x} \in B_\delta(\mathbf{x}_0) \setminus \{\mathbf{x}_0\}$.

Here $|f(\mathbf{x})|$, $|g(\mathbf{x})|$ denote the Euclidean lengths of $f(\mathbf{x})$, $g(\mathbf{x})$, and it is tacitly assumed that δ is chosen in such a way that $B_\delta(\mathbf{x}_0) \subseteq D$.

“ $f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + L(\mathbf{h}) + o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$ ” is a special case of (2): Here $\mathbf{x}_0 = \mathbf{0}$, the vectorial variable is \mathbf{h} instead of \mathbf{x} , the function $\mathbf{h} \mapsto f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - L(\mathbf{h})$ plays the role of f , and $g(\mathbf{h}) = \mathbf{h}$.

“ $f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + L(\mathbf{h}) + o(1)$ ” is also a special case of (2): Use $g(\mathbf{h}) = 1$.

“ $o(\mathbf{h}) = o(1)$ ” means “if $f(\mathbf{h}) = o(\mathbf{h})$ then $f(\mathbf{h}) = o(1)$ ”. Note that ...

Afternote cont'd

... the equality sign doesn't obey the usual rules here but is used informally just like "is" is often in common English: Any function that is $o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$ is also $o(1)$, but not conversely.

The main purpose of using Big-O/little-o notation in the lecture is to make limit calculations more concise. Compare the final part of the proof of Part (3) of the theorem to the following:

$$\begin{aligned}
 f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) &= \frac{\partial f}{\partial x_1}(\xi_1, x_2 + h_2)h_1 + \frac{\partial f}{\partial x_2}(x_1, \xi_2)h_2 \\
 &= \frac{\partial f}{\partial x_1}(x_1, x_2)h_1 + \frac{\partial f}{\partial x_2}(x_1, x_2)h_2 + \underbrace{\left(\frac{\partial f}{\partial x_1}(\xi_1, x_2 + h_2) - \frac{\partial f}{\partial x_1}(x_1, x_2) \right) h_1}_{\rightarrow 0} \\
 &\quad + \underbrace{\left(\frac{\partial f}{\partial x_2}(x_1, \xi_2) - \frac{\partial f}{\partial x_2}(x_1, x_2) \right) h_2}_{\rightarrow 0} \\
 &= \frac{\partial f}{\partial x_1}(x_1, x_2)h_1 + \frac{\partial f}{\partial x_2}(x_1, x_2)h_2 + o(\mathbf{h}) \quad \text{for } \mathbf{h} = (h_1, h_2) \rightarrow (0, 0).
 \end{aligned}$$

The two $o(1)$'s have been replaced. Now imagine we want to get rid of the $o(\mathbf{h})$ as well!

Example

We show that the length function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{x} \mapsto |\mathbf{x}|$ is differentiable at every point $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ with

$$df(\mathbf{x})(\mathbf{h}) = \frac{\mathbf{x} \cdot \mathbf{h}}{|\mathbf{x}|}.$$

For the proof we use Part 3 of the theorem.

$$\begin{aligned} f_{x_j}(x_1, \dots, x_n) &= \frac{\partial}{\partial x_j} \sqrt{x_1^2 + \dots + x_j^2 + \dots + x_n^2} \\ &= \frac{2x_j}{2\sqrt{x_1^2 + \dots + x_n^2}} = \frac{x_j}{|\mathbf{x}|}. \end{aligned}$$

\implies The partial derivatives of f exist and are continuous on $\mathbb{R}^n \setminus \{\mathbf{0}\}$.

$\implies f$ is differentiable on $\mathbb{R}^n \setminus \{\mathbf{0}\}$ with differential

$$df(\mathbf{x})(\mathbf{h}) = \left(\frac{x_1}{|\mathbf{x}|}, \dots, \frac{x_n}{|\mathbf{x}|} \right) \mathbf{h} = \frac{\mathbf{x} \cdot \mathbf{h}}{|\mathbf{x}|}, \quad \mathbf{h} \in \mathbb{R}^n.$$

Example

We compute the differential of the polar coordinate map

$$f(r, \phi) = \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix}, \quad (r, \phi) \in \mathbb{R}^+ \times \mathbb{R}.$$

$$\mathbf{J}_f(r, \phi) = \begin{pmatrix} \frac{\partial(r \cos \phi)}{\partial r} & \frac{\partial(r \cos \phi)}{\partial \phi} \\ \frac{\partial(r \sin \phi)}{\partial r} & \frac{\partial(r \sin \phi)}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix}$$

Since the entries of $\mathbf{J}_f(r, \phi)$ are continuous functions of (r, ϕ) , the polar coordinate map f is differentiable in $\mathbb{R}^+ \times \mathbb{R}$ with differential

$$df(r, \phi)(\mathbf{h}) = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} h_1 \cos \phi - h_2 r \sin \phi \\ h_1 \sin \phi + h_2 r \cos \phi \end{pmatrix}$$

for $\mathbf{h} = (h_1, h_2)^T \in \mathbb{R}^2$.

Example (squaring map in \mathbb{C})

Consider again the squaring map $f: \mathbb{C} \rightarrow \mathbb{C}$, $z \mapsto z^2$, i.e.,

$$f(z) = f(x + y\mathbf{i}) = (x + y\mathbf{i})^2 = x^2 - y^2 + 2xy\mathbf{i} = \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix}.$$

We have

$$\mathbf{J}_f(x, y) = \begin{pmatrix} \frac{\partial(x^2 - y^2)}{\partial x} & \frac{\partial(x^2 - y^2)}{\partial y} \\ \frac{\partial(2xy)}{\partial x} & \frac{\partial(2xy)}{\partial y} \end{pmatrix} = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix},$$

$$df(x, y)(\mathbf{h}) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} 2xh_1 - 2yh_2 \\ 2yh_1 + 2xh_2 \end{pmatrix}.$$

Switching back to \mathbb{C} ,

$$df(z)(h) = df(x + y\mathbf{i})(h_1 + h_2\mathbf{i}) = 2(xh_1 - yh_2) + 2(yh_1 + xh_2)\mathbf{i} = 2zh,$$

so that—as we have mentioned before—the (real) differential $df(z)$ is multiplication by the complex derivative $f'(z) = 2z$.

This holds more generally for complex functions whose complex derivative exists.

Example

Consider again the functions $f, g: \mathbb{R}^2 \setminus \{(0, 0)\} \rightarrow \mathbb{R}$ defined by

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

We extend f, g to \mathbb{R}^2 by defining $f(0, 0) = g(0, 0) = 0$. Then, as we have seen earlier, g is continuous in $(0, 0)$ but f is not.

Reasoning as in the previous examples, one can easily show that f and g are differentiable in $\mathbb{R}^2 \setminus \{(0, 0)\}$.

It turns out, however, that f and g are only partially but not totally differentiable at $(0, 0)$. We show this for f and leave the case of g as a worksheet exercise.

By Part 1 of the theorem, since f is not continuous at $(0, 0)$, it cannot be differentiable at $(0, 0)$.

Since $f(x, 0) = f(0, y) = f(0, 0) = 0$ for all $x, y \in \mathbb{R}$, we have

$$f_x(0, 0) = \lim_{h \rightarrow 0} \frac{f(h, 0) - f(0, 0)}{h} = 0, \quad f_y(0, 0) = \lim_{h \rightarrow 0} \frac{f(0, h) - f(0, 0)}{h} = 0,$$

i.e., the partial derivatives of f exist also at $(0, 0)$.

Example (cont'd)

Why can Part 3 of the theorem not be applied here?

$$f_x(x, y) = \frac{y(x^2 + y^2) - xy(2x)}{(x^2 + y^2)^2} = \frac{y^3 - x^2y}{(x^2 + y^2)^2}$$

Substituting $y = mx$, $m \in \mathbb{R}$ fixed, gives

$$f_x(x, mx) = \frac{(mx)^3 - mx^3}{(x^2 + m^2x^2)^2} = \frac{m^3 - m}{(1 + m^2)^2 x} \quad \text{for } x \in \mathbb{R} \setminus \{0\}.$$

$$\implies \lim_{x \rightarrow 0^\pm} f_x(x, mx) = \pm\infty \text{ or } \mp\infty \quad \text{if } m \neq 0, \pm 1$$

and $\lim_{(x,y) \rightarrow (0,0)} f_x(x, y)$ does not exist (not even in the improper sense).

$\implies f_x$ is discontinuous at $(0, 0)$.

Similarly, f_y is discontinuous at $(0, 0)$.

Directional Derivatives

Definition

Suppose $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is a function, $\mathbf{x} \in D^\circ$ and $\mathbf{u} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. The *derivative* of f at \mathbf{x} in the direction \mathbf{u} is defined as

$$f_{\mathbf{u}}(\mathbf{x}) = \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x})}{t}.$$

Other notations in use for $f_{\mathbf{u}}(\mathbf{x})$ are $D_{\mathbf{u}}f(\mathbf{x})$ and $\frac{\partial f}{\partial \mathbf{u}}(\mathbf{x})$.

Notes

- Partial derivatives form a special case of directional derivatives: $\frac{\partial f}{\partial x_j}(\mathbf{x}) = f_{\mathbf{e}_j}(\mathbf{x})$.
- In general, $f_{\mathbf{u}}(\mathbf{x})$ is equal to the derivative at $t = 0$ of the function $t \mapsto f(\mathbf{x} + t\mathbf{u})$, which describes the behaviour of f on the line $\mathbf{x} + \mathbb{R}\mathbf{u}$. However, since $f_{\lambda\mathbf{u}}(\mathbf{x}) = \lambda f_{\mathbf{u}}(\mathbf{x})$ for $\lambda \in \mathbb{R}$, different choices of the direction vector for this line usually result in different values of the directional derivative.
- For functions $f: D \rightarrow \mathbb{R}$ (i.e., $m = 1$), the quantity $f_{\mathbf{u}}(\mathbf{x})$ measures the slope of G_f at \mathbf{x} in the direction \mathbf{u} ; equality holds in the case $|\mathbf{u}| = 1$.

Notes cont'd

- If f is differentiable at \mathbf{x} , then all directional derivatives $f_{\mathbf{u}}(\mathbf{x})$, $\mathbf{u} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, exist and are obtained as $f_{\mathbf{u}}(\mathbf{x}) = df(\mathbf{x})(\mathbf{u})$. This follows from

$$f(\mathbf{x} + t\mathbf{u}) - f(\mathbf{x}) = L(t\mathbf{u}) + o(t\mathbf{u}) = tL(\mathbf{u}) + o(t)$$

for $t \rightarrow 0$, where $L = df(\mathbf{x})$.

- Returning to the case $m = 1$, the slope of G_f at \mathbf{x} is maximized if the direction vector \mathbf{u} (assumed to have unit length) is taken as a positive multiple of $\mathbf{J}_f(\mathbf{x})$ (and minimized for negative multiples). This follows from
$$f_{\mathbf{u}}(\mathbf{x}) = \mathbf{J}_f(\mathbf{x})\mathbf{u} = \mathbf{J}_f(\mathbf{x})^T \cdot \mathbf{u} = |\mathbf{J}_f(\mathbf{x})^T| |\mathbf{u}| \cos \theta = |\mathbf{J}_f(\mathbf{x})^T| \cos \theta.$$
- The preceding theorem has a coordinate-independent generalization, which uses directional derivatives.
For example, Part 3 generalizes to:

Suppose there exists a basis $\mathbf{u}_1, \dots, \mathbf{u}_n$ of \mathbb{R}^n such that the directional derivatives $f_{\mathbf{u}_j}(\mathbf{x})$, $1 \leq j \leq n$, exist and are continuous at \mathbf{x} . Then f is differentiable at \mathbf{x} , and

$$df(\mathbf{x}) \left(\sum_{j=1}^n h_j \mathbf{u}_j \right) = \sum_{j=1}^n f_{\mathbf{u}_j}(\mathbf{x}) h_j \quad \text{for } (h_1, \dots, h_n) \in \mathbb{R}^n.$$

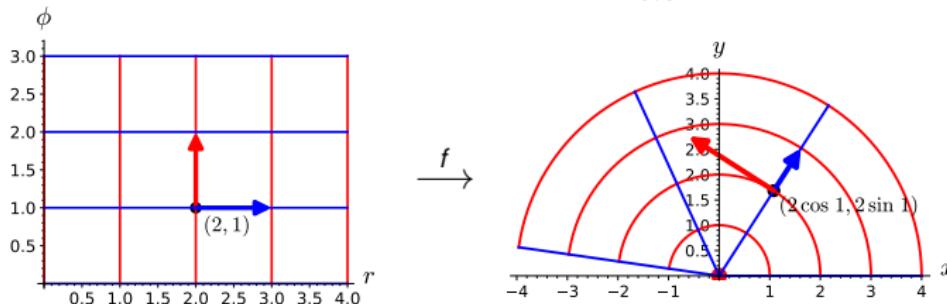
Remark (columns of the Jacobi matrix)

For a differentiable map $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, the vectorial partial derivatives $\frac{\partial f}{\partial x_j}(\mathbf{x}) = \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{e}_j) - f(\mathbf{x})}{t}$ provide tangent vectors to the curves $t \mapsto f(\mathbf{x} + t\mathbf{e}_j)$ (images of the coordinate lines under f) in \mathbf{x} .

For example, the polar coordinate map $f(r, \phi) = \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix}$, which

has $\mathbf{J}_f(r, \phi) = \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix}$, maps the two coordinate lines

through $(2, 1)$ to curves through $f(2, 1) = (2 \cos 1, 2 \sin 1)$ with tangent vectors $\begin{pmatrix} \cos 1 \\ \sin 1 \end{pmatrix}$, respectively, $\begin{pmatrix} -2 \sin 1 \\ 2 \cos 1 \end{pmatrix}$.



In general the tangent vector of an image curve is obtained by applying the differential at the corresponding point to the tangent vector of the original curve (in our case $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ resp. $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$); cf. next lecture.

The Gradient

We consider a real-valued function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$. Suppose f is differentiable at \mathbf{x} .

Observation

For $\mathbf{h} \in \mathbb{R}^n$ (represented as a column vector) we have

$$\begin{aligned} df(\mathbf{x})(\mathbf{h}) &= \mathbf{J}_f(\mathbf{x})\mathbf{h} = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\mathbf{x}) h_j \\ &= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix} \cdot \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}, \end{aligned}$$

i.e., the differential $df(\mathbf{x})$ is “taking the dot product with the column vector $\mathbf{J}_f(\mathbf{x})^T$ of partial derivatives”.

Definition

The column vector $\mathbf{J}_f(\mathbf{x})^T = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right)^T \in \mathbb{R}^n$ is called *gradient* of f at \mathbf{x} and denoted by $\nabla f(\mathbf{x})$ (“nabla” $f(\mathbf{x})$) or $\text{grad } f(\mathbf{x})$.

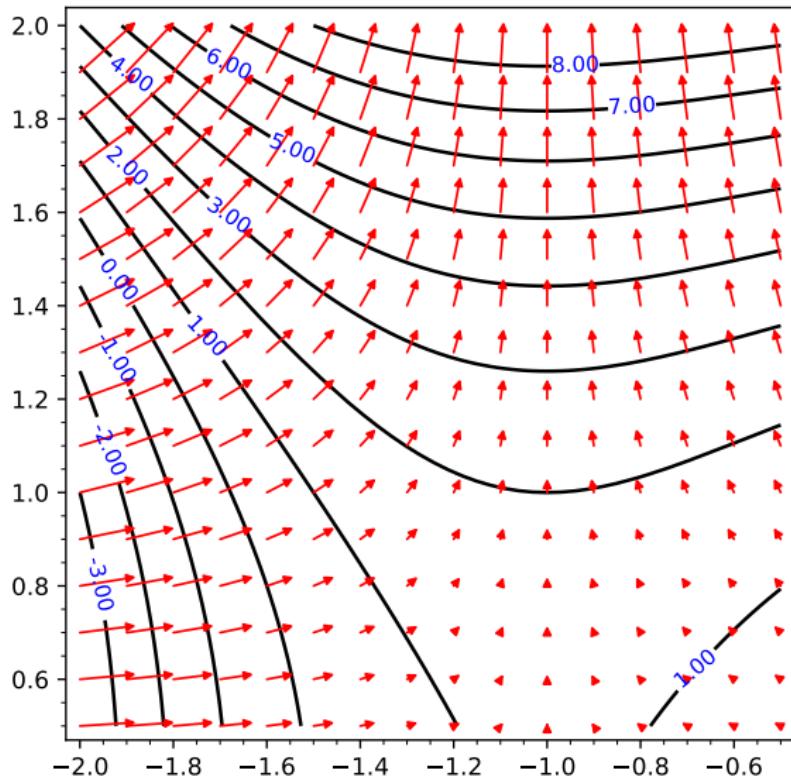


Figure: Contours of $f(x, y) = 2x^3 + 3x^2 + y^3$ and gradients $\nabla f(x, y) = (6x^2 + 6x, 3y^2)^T$ scaled by 0.01

Notes

- In terms of the gradient, the approximation property of the differential takes the form $f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{h} + o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$. Here \mathbf{x} and \mathbf{h} are viewed as column vectors.
- $\nabla f(\mathbf{x})$ contains of course the same information as $\mathbf{J}_f(\mathbf{x})$, but it “lives” in the ambient space of D and interacts with the points in D through vector arithmetic; cf. the picture.
- The gradient $\nabla f(\mathbf{x})$ points into the direction of the steepest ascent of the graph G_f at \mathbf{x} .
More precisely, the slope m of the one-variable function $t \mapsto f(\mathbf{x} + t\mathbf{u})$, $|\mathbf{u}| = 1$, at $t = 0$ is maximized for $\mathbf{u} = \frac{\nabla f(\mathbf{x})}{|\nabla f(\mathbf{x})|}$, as follows from $m = f_{\mathbf{u}}(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{u} = |\nabla f(\mathbf{x})| |\mathbf{u}| \cos \theta$. The maximal slope is $|\nabla f(\mathbf{x})|$.
- The gradient $\nabla f(\mathbf{x})$ is perpendicular to the contour of f through \mathbf{x} (provided that the contour admits a parametrization that is smooth at \mathbf{x}); see the corollary to the Chain Rule.

Tangent Spaces

Suppose $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is differentiable at \mathbf{x}_0 . Setting $\mathbf{x} = \mathbf{x}_0 + \mathbf{h}$, the approximation property can be written as

$$f(\mathbf{x}) = f(\mathbf{x}_0) + L(\mathbf{x} - \mathbf{x}_0) + o(\mathbf{x} - \mathbf{x}_0) \quad \text{for } \mathbf{x} \rightarrow \mathbf{x}_0.$$

This says that the affine map $A: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{x} \mapsto f(\mathbf{x}_0) + L(\mathbf{x} - \mathbf{x}_0)$, $L = df(\mathbf{x}_0)$, approximates f very well near \mathbf{x}_0 . The graph G_A appears to touch G_f in $(\mathbf{x}_0, f(\mathbf{x}_0))$.

Definition

The graph G_A is called *(affine) tangent space* of G_f at \mathbf{x}_0 .

Notes

- In parametric form the tangent space is given as

$$\begin{aligned} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} &= \begin{pmatrix} \mathbf{x} \\ \mathbf{y}_0 + \mathbf{A}(\mathbf{x} - \mathbf{x}_0) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{y}_0 - \mathbf{A}\mathbf{x}_0 \end{pmatrix} + \begin{pmatrix} \mathbf{I}_n \\ \mathbf{A} \end{pmatrix} \mathbf{x} \\ &= \begin{pmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{pmatrix} + \begin{pmatrix} \mathbf{I}_n \\ \mathbf{A} \end{pmatrix} \mathbf{h}, \end{aligned} \quad (\text{Subst. } \mathbf{x} = \mathbf{x}_0 + \mathbf{h})$$

where $\mathbf{y}_0 = f(\mathbf{x}_0)$ and $\mathbf{A} = \mathbf{J}_f(\mathbf{x}_0)$. It follows that $\dim(G_A) = n$.

Notes cont'd

- An equational form for the tangent space is $\mathbf{y} - \mathbf{y}_0 = \mathbf{A}(\mathbf{x} - \mathbf{x}_0)$ or, as a proper linear system of equations,

$$(-\mathbf{A} \quad \mathbf{I}_n) \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{y}_0 - \mathbf{Ax}_0.$$

In the case $m = 1$ the tangent space is a hyperplane of \mathbb{R}^{n+1} (*tangent hyperplane*). It is then defined by the single equation $y - y_0 = \mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0)$, or

$$x_{n+1} = f(\mathbf{x}^{(0)}) + \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\mathbf{x}^{(0)})(x_j - x_j^{(0)}),$$

where we have written y as x_{n+1} and, in order to avoid double subscripts, \mathbf{x}_0 as $\mathbf{x}^{(0)} = (x_1^{(0)}, \dots, x_n^{(0)})$.

If f is a function of two variables, we can use x, y, z -notation instead. The *tangent plane* to the graph of f in (x_0, y_0, z_0) , $z_0 = f(x_0, y_0)$ is then given by

$$z = z_0 + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

Example

We determine the tangent plane of the parabolic cylinder P in \mathbb{R}^3 with equation $z = x^2 + y^2$ in the point $(1, 1, 2) \in P$.

P is the graph of $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $(x, y) \mapsto x^2 + y^2$.

We compute $f_x(x, y) = 2x$, $f_y(x, y) = 2y$ and hence $\mathbf{J}_f(x, y) = (2x, 2y)$.

⇒ An equation for the tangent plane of P in $(1, 1, 2)$ is

$$\begin{aligned} z &= f(1, 1) + f_x(1, 1)(x - 1) + f_y(1, 1)(y - 1) \\ &= 2 + 2(x - 1) + 2(y - 1) = 2x + 2y - 2. \end{aligned}$$

The corresponding parametric form is

$$\begin{aligned} \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} x \\ y \\ 2x + 2y - 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -2 \end{pmatrix} + x \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} + h_1 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + h_2 \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}. \end{aligned}$$

Generalization

Our earlier definition of the tangent line to a (smooth) parametric curve doesn't fit the present definition of "tangent space", since curves usually are not specified as graphs of maps $D \rightarrow \mathbb{R}^{n-1}$, $D \subseteq \mathbb{R}$.

Definition (Parametric surface)

A map $g: D \rightarrow \mathbb{R}^n$, $D \subseteq \mathbb{R}^k$, is called a parametric surface in \mathbb{R}^n .

The parametric surface is said to be *smooth* and of *dimension k* if g is differentiable and $\mathbf{J}_g(\mathbf{x})$ has full column rank k for all $\mathbf{x} \in D$.

Just like a parametric curve, a parametric surface has a geometric object associated with it, viz. the range $g(D)$. This is called a non-parametric surface.

Parametric surfaces will be discussed later in more detail, when we do surface integration.

For a parametric surface we can define the tangent space in a way similar to the definition of the tangent line to a parametric curve. If g is differentiable in \mathbf{x}_0 , the distance between $g(\mathbf{x})$ and the linear approximation $\mathbf{x} \mapsto g(\mathbf{x}_0) + \mathbf{J}_g(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$ near \mathbf{x}_0 is much smaller than $|\mathbf{x} - \mathbf{x}_0|$, so that the range of the linear approximation appears to touch the surface in $g(\mathbf{x}_0)$.

Definition

Suppose $g: D \rightarrow \mathbb{R}^n$, $D \subseteq \mathbb{R}^k$, is differentiable in $\mathbf{x}_0 \in D$ and $\text{rk } \mathbf{J}_g(\mathbf{x}_0) = k$. Then the range

$$T = \{g(\mathbf{x}_0) + \mathbf{J}_g(\mathbf{x}_0)\mathbf{h}; \mathbf{h} \in \mathbb{R}^k\}$$

of the linear approximation of g in \mathbf{x}_0 is called *(affine) tangent space* of g in the point $g(\mathbf{x}_0)$, or of the non-parametric surface $g(D)$ associated with g .

By assumption, T is a k -dimensional affine subspace of \mathbb{R}^n , whose associated linear subspace ("direction space") is the column space of $\mathbf{J}_g(\mathbf{x}_0)$.

The last part of the definition actually requires justification (which is omitted): Show that if a non-parametric surface S is represented as $S = g_1(D_1) = g_2(D_2)$ with parametrizations g_1, g_2 then the tangent spaces T_1, T_2 of g_1, g_2 at $\mathbf{y} = g_1(\mathbf{x}_1) = g_2(\mathbf{x}_2)$ according to the previous definition are equal.

Example

Consider the (non-parametric) surface

$$S = \{(u + v, u^2 + v^2, u^3 + v^3); u, v \in \mathbb{R}\}.$$

Here we have

$$g(u, v) = \begin{pmatrix} u + v \\ u^2 + v^2 \\ u^3 + v^3 \end{pmatrix}, \quad \mathbf{J}_g(u, v) = \begin{pmatrix} 1 & 1 \\ 2u & 2v \\ 3u^2 & 3v^2 \end{pmatrix}.$$

Transforming $\mathbf{J}_g(u, v)$ into column-echelon form, viz.

$$\begin{pmatrix} 1 & 0 \\ 2u & 2(u-v) \\ 3u^2 & 3(u-v)(u+v) \end{pmatrix}, \text{ shows that } \mathbf{J}_g(u, v) \text{ has rank 2 if } u \neq v.$$

The points on S with $u = v$, i.e., $g(u, u) = (2u, 2u^2, 2u^3)$ form a twisted cubic (enlarged by the factor 2), and S (which one may call “twisted cubic surface”) appears to have this curve C as a 1-dimensional boundary; cf. subsequent picture. Removing this curve from S results in a smooth 2-dimensional parametric surface, which can be parametrized bijectively by restricting the domain of g to $\{(u, v) \in \mathbb{R}^2; u > v\}$, say.

[Introduction](#)

Differentiable
maps

Partial
Derivatives

Further
Concepts

Directional
Derivatives

The Gradient

Tangent Spaces

True Meaning of
Differentials

The Chain Rule

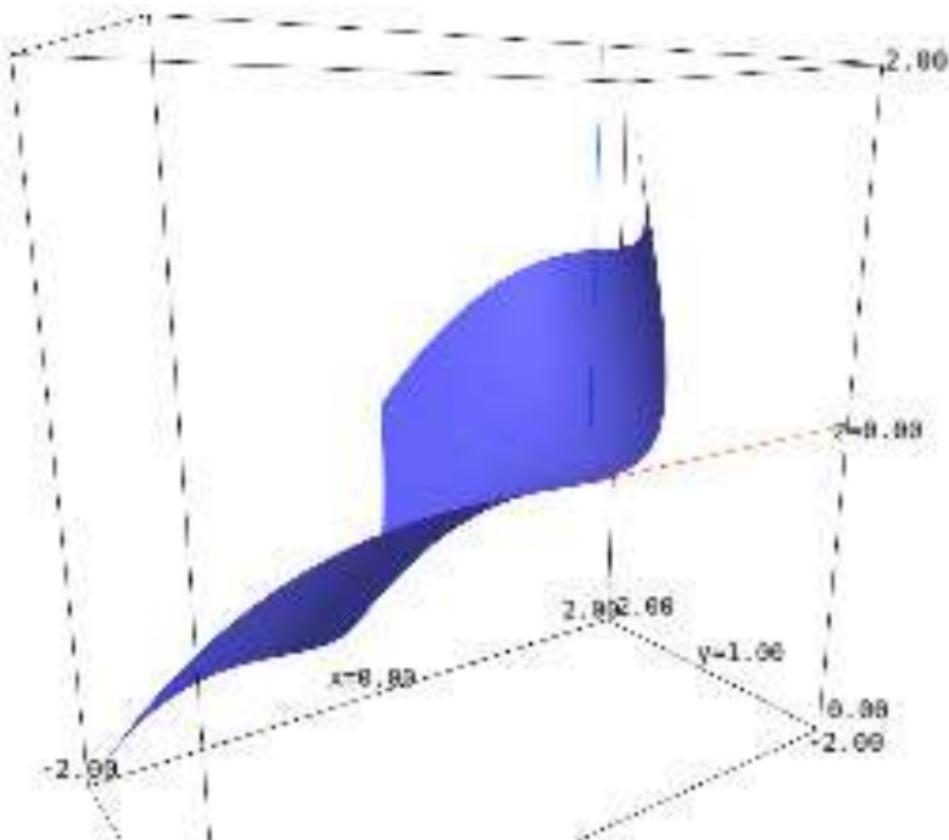


Figure: The twisted-cubic surface with parameter domain restricted to $[-1, 1]^2$

Example (cont'd)

As an example for computing tangent planes we consider the point $g(1, 0) = (1, 1, 1)$. Since $\mathbf{J}_g(1, 0) = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 0 \end{pmatrix}$, the tangent plane to S in $(1, 1, 1)$ has parametric form

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \mathbb{R} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \mathbb{R} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

and equation $3y - 2z = 1$.

Exercise

- ① Show that the restriction of $g(u, v) = (u + v, u^2 + v^2, u^3 + v^3)$ maps $\{(u, v) \in \mathbb{R}^2; u > v\}$ bijectively onto $S \setminus C$.
- ② Show that S can be represented as graph of a function $z = h(x, y)$, compute the tangent plane to G_h in $(x, y, z) \in S$ according to our earlier definition, and verify that both definitions yield the same tangent planes.

Hint: Eliminate u, v from $x = u + v, y = u^2 + v^2, z = u^3 + v^3$.

What are dx , dx_j , $d\mathbf{x}$, dz ?

Have you ever wondered what “ dx ” in the notation for integrals, e.g., in $\int_0^1 x^2 dx$ means?

Answer: dx and its friends are differentials.

- dx denotes the differential of $\mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto x$ (the identity map $\text{id}_{\mathbb{R}}$ of \mathbb{R}), which is $\mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto \text{id}_{\mathbb{R}}$. Thus $dx(h) = dx(x_0)(h) = h$ for all $x_0 \in \mathbb{R}$ and $h \in \mathbb{R}$. Moreover, $df(x) = f'(x)dx$ in the sense that $df(x)(h) = f'(x)h = f'(x)dx(h)$ for all x at which f is differentiable and all $h \in \mathbb{R}$.
- dx_j denotes the differential of the coordinate projection $\mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{x} \mapsto x_j$, which is given by $dx_j(\mathbf{h}) = dx_j(\mathbf{x}_0)(\mathbf{h}) = h_j$ for all $\mathbf{x}_0 \in \mathbb{R}^n$ and $\mathbf{h} \in \mathbb{R}^n$. Moreover, $df = \sum_{j=1}^n \frac{\partial f}{\partial x_j} dx_j$ in the sense that if f is differentiable at \mathbf{x} then

$$df(\mathbf{x})(\mathbf{h}) = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\mathbf{x}) dx_j(\mathbf{h}) \quad \text{for } \mathbf{h} \in \mathbb{R}^n.$$

Introduction

Differentiable
mapsPartial
DerivativesFurther
ConceptsDirectional
DerivativesThe Gradient
Tangent SpacesTrue Meaning of
Differentials

The Chain Rule

Answer (cont'd)

- $d\mathbf{x}$ is the differential of $\mathbb{R}^n \rightarrow \mathbb{R}^n$, $\mathbf{x} \mapsto \mathbf{x}$ (the identity on \mathbb{R}^n), which is given by $d\mathbf{x}(\mathbf{h}) = d\mathbf{x}(\mathbf{x}_0)(\mathbf{h}) = \mathbf{h}$ for all $\mathbf{x}_0 \in \mathbb{R}^n$, $\mathbf{h} \in \mathbb{R}^n$.
- dz is the differential of $\mathbb{C} \rightarrow \mathbb{C}$, $z \mapsto z$ (the identity on \mathbb{C}) and thus equal to $d\mathbf{x}$ for $n = 2$, provided that \mathbb{C} is identified with \mathbb{R}^2 . We have $df(z) = f'(z)dz$ for those z in the domain of f for which the complex derivative $f'(z) = \lim_{h \rightarrow 0} \frac{f(z+h)-f(z)}{h}$ exists.

Caution

In [Ste21] the differential df of a two-variable scalar function $z = f(x, y)$ is sometimes denoted by dz as well. In the lecture we won't use dz in this sense. Instead, when writing $x = x_1$, $y = y_2$, $z = x_3$, we denote the corresponding differentials by dx , dy , and dz . So in the lecture, dz has two different meanings: (i) $dz = dx_3$; (ii) the complex differential $dz = dx + i dy$.

The differentials $d\mathbf{x}$, dz , which belong to vector-valued functions, will be rarely used in the sequel. The coordinate differentials dx_j , or dx , dy , dz , however, are so convenient to use (and will be used frequently).

Example

Consider $f(x, y) = x^3 - 3xy^2$ defined on $D = \mathbb{R}^2$.

$$f_x(x, y) = 3x^2 - 3y^2, \quad f_y(x, y) = -6xy,$$

$$\mathbf{J}_f(x, y) = \begin{pmatrix} 3x^2 - 3y^2 & -6xy \end{pmatrix}, \quad \nabla f(x, y) = \begin{pmatrix} 3x^2 - 3y^2 \\ -6xy \end{pmatrix}$$

$$df = f_x dx + f_y dy = (3x^2 - 3y^2) dx - 6xy dy$$

The purpose of the differential is to approximate

$$f(x+h_1, y+h_2) - f(x, y) \approx df(x, y)(h_1, h_2) = (3x^2 - 3y^2)h_1 - 6xyh_2$$

for small h_1, h_2 .

Mnemonic: In order to apply $df(x, y)$ to $\mathbf{h} = (h_1, h_2)$, substitute the coordinates h_1, h_2 for dx, dy in the expression
 $df = (3x^2 - 3y^2) dx - 6xy dy$.

The Multivariable Chain Rule

generalizing $(g \circ f)'(x) = g'(y)f'(x)$

Theorem

Suppose $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is differentiable in $\mathbf{x}_0 \in D$, $g: E \rightarrow \mathbb{R}^p$, $E \subseteq \mathbb{R}^m$, is differentiable in $\mathbf{y}_0 \in E$, and f satisfies $f(D) \subseteq E$, $f(\mathbf{x}_0) = \mathbf{y}_0$. Then $g \circ f: D \rightarrow \mathbb{R}^p$ is differentiable in \mathbf{x}_0 , and its differential satisfies

$$d(g \circ f)(\mathbf{x}_0) = dg(\mathbf{y}_0) \circ df(\mathbf{x}_0).$$

In other words, the differential of a composition is the composition of the differentials (evaluated at the respective points).

Proof.

Writing $L = df(\mathbf{x}_0)$, $M = dg(\mathbf{y}_0)$, we have $L: \mathbb{R}^n \rightarrow \mathbb{R}^m$, $M: \mathbb{R}^m \rightarrow \mathbb{R}^p$, $M \circ L: \mathbb{R}^n \rightarrow \mathbb{R}^p$ and must show that $d(g \circ f)(\mathbf{x}_0) = M \circ L$ (hence at least the dimensions match).

The differentiability conditions say that for $\mathbf{h} \rightarrow \mathbf{0}$, $\mathbf{k} \rightarrow \mathbf{0}$,

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + L(\mathbf{h}) + \phi(\mathbf{h}), \quad \phi(\mathbf{h}) = o(\mathbf{h}),$$

$$g(\mathbf{y}_0 + \mathbf{k}) = g(\mathbf{y}_0) + M(\mathbf{k}) + \psi(\mathbf{k}), \quad \psi(\mathbf{k}) = o(\mathbf{k}).$$

Proof cont'd.

Hence

$$\begin{aligned}g(f(\mathbf{x}_0 + \mathbf{h})) &= g(f(\mathbf{x}_0) + L(\mathbf{h}) + \phi(\mathbf{h})) \\&= g(f(\mathbf{x}_0)) + M(L(\mathbf{h}) + \phi(\mathbf{h})) + \psi(L(\mathbf{h}) + \phi(\mathbf{h})) \\&= g(f(\mathbf{x}_0)) + M(L(\mathbf{h})) + M(\phi(\mathbf{h})) + \psi(L(\mathbf{h}) + \phi(\mathbf{h}))\end{aligned}$$

for $\mathbf{h} \rightarrow \mathbf{0}$, and it remains to show that (i) $M(\phi(\mathbf{h})) = o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$ and (ii) $\psi(L(\mathbf{h}) + \phi(\mathbf{h})) = o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$.

(i) This part is easy and can be done as follows:

$$\frac{M(\phi(\mathbf{h}))}{|\mathbf{h}|} = M\left(\frac{\phi(\mathbf{h})}{|\mathbf{h}|}\right) \rightarrow M(\mathbf{0}) = \mathbf{0} \quad \text{for } \mathbf{h} \rightarrow \mathbf{0},$$

using the linearity of M , $\lim_{\mathbf{h} \rightarrow \mathbf{0}} \phi(\mathbf{h})/|\mathbf{h}| = \mathbf{0}$, and the continuity of M in $\mathbf{0}$.

Proof cont'd.

(ii) This part is more difficult. We have

$$\frac{\psi(L(\mathbf{h}) + \phi(\mathbf{h}))}{|\mathbf{h}|} = \frac{\psi(L(\mathbf{h}) + \phi(\mathbf{h}))}{|L(\mathbf{h}) + \phi(\mathbf{h})|} \cdot \frac{|L(\mathbf{h}) + \phi(\mathbf{h})|}{|\mathbf{h}|}.$$

The 1st factor tends to $\mathbf{0} \in \mathbb{R}^p$ for $\mathbf{h} \rightarrow \mathbf{0}$, since $L(\mathbf{h}) + \phi(\mathbf{h}) \rightarrow \mathbf{0}$ and $\lim_{\mathbf{k} \rightarrow \mathbf{0}} \psi(\mathbf{k}) / |\mathbf{k}| = \mathbf{0}$.

The 2nd factor remains bounded for $\mathbf{h} \rightarrow \mathbf{0}$, since $|\phi(\mathbf{h})| / |\mathbf{h}| \rightarrow 0$ and

$$\begin{aligned} \frac{|L(\mathbf{h})|}{|\mathbf{h}|} &= \frac{|L(h_1 \mathbf{e}_1 + \cdots + h_n \mathbf{e}_n)|}{|\mathbf{h}|} = \frac{|h_1 L(\mathbf{e}_1) + \cdots + h_n L(\mathbf{e}_n)|}{|\mathbf{h}|} \\ &\leq |L(\mathbf{e}_1)| + \cdots + |L(\mathbf{e}_n)|. \end{aligned}$$

This shows $\psi(L(\mathbf{h}) + \phi(\mathbf{h})) = o(\mathbf{h})$ for $\mathbf{h} \rightarrow \mathbf{0}$ and completes the proof of the chain rule. □

Corollary (chain rule in matrix form)

Under the assumptions of the theorem we have

$$\mathbf{J}_{g \circ f}(\mathbf{x}_0) = \mathbf{J}_g(\mathbf{y}_0) \mathbf{J}_f(\mathbf{x}_0).$$

Proof.

Use $L(\mathbf{h}) = \mathbf{J}_f(\mathbf{x}_0)\mathbf{h}$, $M(\mathbf{k}) = \mathbf{J}_g(\mathbf{y}_0)\mathbf{k}$, and the fact that the composition of two linear maps is represented by the product of the corresponding matrices. □

Remark

If g is scalar-valued ($p = 1$), we have, suppressing arguments,

$$\mathbf{J}_g = \left(\frac{\partial g}{\partial y_1}, \dots, \frac{\partial g}{\partial y_m} \right) \text{ and similarly, writing } h = g \circ f,$$

$$\mathbf{J}_h = \left(\frac{\partial h}{\partial x_1}, \dots, \frac{\partial h}{\partial x_n} \right). \text{ Since } \mathbf{J}_f = \left(\frac{\partial f_i}{\partial x_j} \right), \text{ the corollary gives}$$

$$\frac{\partial h}{\partial x_j} = (\mathbf{J}_g \mathbf{J}_f)_j = \sum_{i=1}^m \frac{\partial g}{\partial y_i} \frac{\partial f_i}{\partial x_j} = \sum_{i=1}^m \frac{\partial u}{\partial y_i} \frac{\partial y_i}{\partial x_j}.$$

In terms of the variables $y_i = f_i(x_1, \dots, x_n)$, $u = g(y_1, \dots, y_m) = h(x_1, \dots, x_n)$ this can be written as $\frac{\partial u}{\partial x_j} = \sum_{i=1}^m \frac{\partial u}{\partial y_i} \frac{\partial y_i}{\partial x_j}$, recovering the “general version” of the chain rule in [Ste21], Ch. 14.5, p. 988.

Example ([Ste21], Ch. 14.5, Example 4, p. 988)

Here $w = g(x, y, z, t)$ is composed with $f(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \\ z(u, v) \\ t(u, v) \end{pmatrix}$, and the chain rule

$$\begin{aligned}\mathbf{J}_{g \circ f}(u, v) &= \mathbf{J}_g(f(u, v)) \mathbf{J}_f(u, v) \\ &= \mathbf{J}_g(x, y, z, t) \mathbf{J}_f(u, v)\end{aligned}$$

takes the form

$$\left(\frac{\partial w}{\partial u} \quad \frac{\partial w}{\partial v} \right) = \left(\frac{\partial w}{\partial x} \quad \frac{\partial w}{\partial y} \quad \frac{\partial w}{\partial z} \quad \frac{\partial w}{\partial t} \right) \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{pmatrix}.$$

The full story is not visible in this form, since the partial derivatives $\frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}, \frac{\partial w}{\partial z}, \frac{\partial w}{\partial t}$ need to be composed with $f(u, v)$.

Corollary

Suppose $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^2$, is differentiable at $\mathbf{x} = (x_1, x_2)$ and the contour of f through \mathbf{x} , viz. $N_f(k)$ with $k = f(\mathbf{x})$, admits a smooth parametrization g near \mathbf{x} . Then $\nabla f(\mathbf{x})$ is perpendicular to the tangent line of $N_f(k)$ at \mathbf{x} .

Proof.

By assumption, there exists $g: (a, b) \rightarrow D$ and $t_0 \in (a, b)$ such that $g(t_0) = \mathbf{x}$, $g'(t_0) \neq \mathbf{0} \in \mathbb{R}^2$, and $f(g(t)) = k$ for $t \in (a, b)$.

The Chain Rule gives

$$0 = \frac{d}{dt}f(g(t)) = d(f \circ g)(t)(1) = \nabla f(g(t)) \cdot g'(t).$$

Plugging in $g(t_0) = \mathbf{x}$ and recalling that the tangent line to the curve at $g(t_0)$ is $g(t_0) + \mathbb{R}g'(t_0)$ completes the proof. □

The proof shows that orthogonality holds at every point $g(t)$, $t \in (a, b)$, but this statement is of course equivalent to the corollary.

Notes on the Corollary

- The Implicit Function Theorem (a pretty advanced result, which is beyond the scope of this course) gives that for a C^1 -function f (i.e., the partial derivatives of f exist and are continuous on D) the condition $\nabla f(\mathbf{x}_0) \neq (0, 0)$ is sufficient for $N_f(k)$ admitting a smooth parametrization near \mathbf{x}_0 .
- In the n -variable case $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, level sets $N_f(k)$ locally admit parametrizations by functions g of $n - 1$ variables (provided f is C^1 and $\nabla f(\mathbf{x}_0) \neq \mathbf{0}$) and form $(n - 1)$ -dimensional smooth parametric surfaces in \mathbb{R}^n . Reasoning as in the proof of the corollary then shows: The gradient $\nabla f(\mathbf{x}_0)$ is orthogonal to the columns of $\mathbf{J}_g(\omega_0)$, where $\mathbf{x}_0 = g(\omega_0)$. But the columns of $\mathbf{J}_g(\omega_0)$ generate the $n - 1$ -dimensional direction space of the tangent hyperplane T of g at ω_0 or, equivalently, of $N_f(k)$ in $\mathbf{x}_0 = g(\omega_0)$. Thus $\nabla f(\mathbf{x}_0)$ serves as normal vector for T , which therefore has equation $\nabla f(\mathbf{x}_0) \cdot (\mathbf{x} - \mathbf{x}_0) = 0$; cp. [Ste21], Ch. 14.6, Eq. (19). The chain rule argument also shows that $\nabla f(\mathbf{x}_0)$ is orthogonal to every curve through \mathbf{x}_0 (i.e., orthogonal to its tangent L at \mathbf{x}_0) that is entirely contained in the corresponding level surface $N_f(k)$, $k = f(\mathbf{x}_0)$. Under the assumption $\text{rk } \mathbf{J}_g(\omega_0) = n - 1$ this implies that L is contained in the tangent hyperplane of $N_f(k)$.

Example

The sphere $x^2 + y^2 + z^2 = 9$ contains the point $(1, 2, 2)$. We compute the tangent plane in $(1, 2, 2)$ in two different ways:

- 1 The sphere is the 9-level surface of $f(x, y, z) = x^2 + y^2 + z^2$. Since $\nabla f(x, y, z) = (2x, 2y, 2z) = 2(x, y, z)$, we can take the point (x, y, z) itself as normal vector and obtain that the tangent plane has equation $(x - 1) + 2(y - 2) + 2(z - 2) = 0$.
- 2 The upper half sphere $x^2 + y^2 + z^2 = 9 \wedge z > 0$, which contains $(1, 2, 2)$, is parametrized by

$$g(x, y) = \left(x, y, \sqrt{9 - x^2 - y^2} \right), \quad x^2 + y^2 < 9.$$

$$\Rightarrow \mathbf{J}_g(x, y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{x}{\sqrt{9-x^2-y^2}} & -\frac{y}{\sqrt{9-x^2-y^2}} \end{pmatrix}, \quad \mathbf{J}_g(1, 2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{1}{2} & -1 \end{pmatrix}$$

\Rightarrow A parametric form for the tangent plane is

$$\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} + \mathbb{R} \begin{pmatrix} 1 \\ 0 \\ -\frac{1}{2} \end{pmatrix} + \mathbb{R} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

Suppose $f: D \rightarrow \mathbb{R}^n$, $D \subseteq \mathbb{R}^n$, is a C^1 -function and $\mathbf{x}_0 \in D$ satisfies $\text{rk}(df(\mathbf{x}_0)) = n$ (i.e., the linear map $df(\mathbf{x}_0)$ or, equivalently, the Jacobi matrix $\mathbf{J}_f(\mathbf{x}_0)$ has an inverse).

In this case the so-called *Inverse Function Theorem* shows that f itself has a C^1 -inverse on a suitably restricted domain $D' \subseteq D$, i.e., there exists an open set $D' \subseteq D$ with $\mathbf{x}_0 \in D'$ and a C^1 -function $g: E' \rightarrow D'$, $E' = f(D')$, such that $f \circ g = \text{id}_{E'}$, $g \circ f = \text{id}_{D'}$.

Corollary

Under the assumptions made above we have

$$dg(\mathbf{y}) = df(g(\mathbf{y}))^{-1} \quad \text{for } \mathbf{y} \in E'$$

or, in terms of Jacobi matrices, $\mathbf{J}_g(\mathbf{y}) = \mathbf{J}_f(g(\mathbf{y}))^{-1}$ for $\mathbf{y} \in E'$.

Proof.

By assumption $g(f(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in D'$. Applying the chain rule gives

$$dg(f(\mathbf{x})) \circ df(\mathbf{x}) = d\mathbf{x} = \text{id}_{\mathbb{R}^n} \quad \text{for } \mathbf{x} \in D'.$$

Similarly, using $f(g(\mathbf{y})) = \mathbf{y}$ one shows $df(g(\mathbf{y})) \circ dg(\mathbf{y}) = \text{id}_{\mathbb{R}^n}$ for all $\mathbf{y} \in E'$. This provides ample proof of the desired equality $dg(\mathbf{y}) = df(\mathbf{x})^{-1}$, $\mathbf{y} = f(\mathbf{x})$. □

The Chain Rule—A Concrete Example

Example (Differential of the length function re-examined)

The length function

$$h(\mathbf{x}) = \|\mathbf{x}\| \equiv \sqrt{x_1^2 + \cdots + x_n^2} \equiv \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

is the composition of $f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{x}$ and $g(y) = \sqrt{y}$. Since

$$df(\mathbf{x})(\mathbf{h}) = 2\mathbf{x}^T \mathbf{h}, \quad g'(y) = \frac{1}{2\sqrt{y}},$$

the chain rule gives

$$\mathbf{J}_h(\mathbf{x}) = g'(f(\mathbf{x}))\mathbf{J}_f(\mathbf{x}) = \frac{2\mathbf{x}^T}{2\sqrt{\mathbf{x} \cdot \mathbf{x}}} = \frac{\mathbf{x}^T}{\|\mathbf{x}\|}, \quad \text{or} \quad \nabla h(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}.$$

In other words, the gradient field of h is radial and consists of unit vectors.

⇒ The graph of h has radial slope 1 (except at the origin), i.e., it is (the surface of) a cone.

Math 241
Calculus III

Thomas
Honold

The Mean
Value
Theorem and
its Friends

Error Propagation

Implicit
Functions and
their
Differentiation

Higher
Derivatives

Examples of Partial
Differential Equations

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 The Mean Value Theorem and its Friends Error Propagation

2 Implicit Functions and their Differentiation

3 Higher Derivatives Examples of Partial Differential Equations

Today's Lecture: Mean Value Theorem, Implicit Differentiation, Higher Derivatives

The Mean Value Theorem

First recall the Mean Value Theorem of single-variable calculus:

If $f: [a, b] \rightarrow \mathbb{R}$ is continuous in $[a, b]$ and differentiable in (a, b) then there exists $\xi \in (a, b)$ such that $f(b) - f(a) = f'(\xi)(b - a)$.

Theorem (Mean Value Theorem)

Suppose $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, is differentiable and $\mathbf{a}, \mathbf{b} \in D$ are such that the line segment $[\mathbf{a}, \mathbf{b}] = \{(1-t)\mathbf{a} + t\mathbf{b}; 0 \leq t \leq 1\}$ is contained in D . Then there exists a point $\mathbf{x} \in [\mathbf{a}, \mathbf{b}]$ such that

$$f(\mathbf{b}) - f(\mathbf{a}) = df(\mathbf{x})(\mathbf{b} - \mathbf{a}) = \nabla f(\mathbf{x}) \cdot (\mathbf{b} - \mathbf{a}).$$

Proof.

Consider the one-variable function $\phi: [0, 1] \rightarrow \mathbb{R}$, $t \mapsto f((1-t)\mathbf{a} + t\mathbf{b})$, which satisfies $\phi(0) = f(\mathbf{a})$, $\phi(1) = f(\mathbf{b})$. The chain rule gives

$$\phi'(t) = df(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a}) \quad \text{for } t \in [0, 1].$$

By the Mean Value Theorem of Calculus I, there exists $\xi \in (0, 1)$ such $\phi(1) - \phi(0) = \phi'(\xi)$.

$\Rightarrow \mathbf{x} = \mathbf{a} + \xi(\mathbf{b} - \mathbf{a})$ has the required property. □

The Mean
Value
Theorem and
its Friends

Error Propagation

Implicit
Functions and
their
Differentiation

Higher
Derivatives

Examples of Partial
Differential Equations

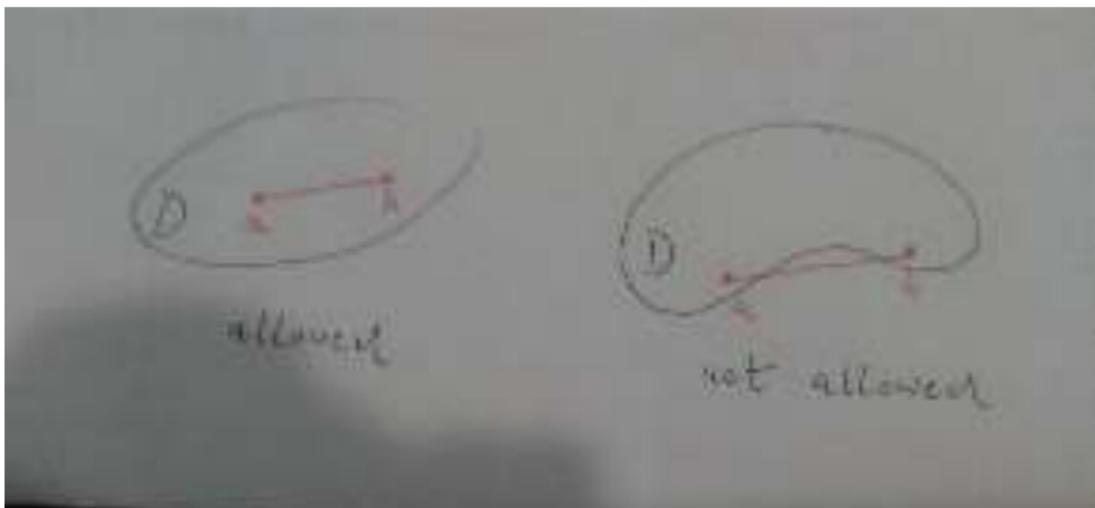


Figure: Illustration of the Mean Value Theorem for $n = 2$

Integral Version

As in Calculus I, there exists an integral representation of the difference $f(\mathbf{b}) - f(\mathbf{a})$, provided that f (and hence ϕ) is continuously differentiable. It is obtained as follows:

$$\begin{aligned} f(\mathbf{b}) - f(\mathbf{a}) &= \int_0^1 \phi'(t) dt = \int_0^1 df(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a}) dt \\ &= \left(\int_0^1 \nabla f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) dt \right) \cdot (\mathbf{b} - \mathbf{a}) \end{aligned}$$

In contrast with the “Lagrange (mid-point) version” this generalizes immediately to the case $m > 1$.

Mean Value Theorem (Integral Version)

Suppose $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is continuously differentiable and $\mathbf{a}, \mathbf{b} \in D$ are such that the line segment

$[\mathbf{a}, \mathbf{b}] = \{(1-t)\mathbf{a} + t\mathbf{b}; 0 \leq t \leq 1\}$ is contained in D . Then

$$f(\mathbf{b}) - f(\mathbf{a}) = \left(\int_0^1 \mathbf{J}_f(\mathbf{a} + t(\mathbf{b} - \mathbf{a})) dt \right) (\mathbf{b} - \mathbf{a}).$$

Example

As an example for the integral version of the Mean Value Theorem we consider the squaring map $s(x, y) = (x^2 - y^2, 2xy)^T$. We have seen that $\mathbf{J}_s(x, y) = \begin{pmatrix} 2x & -2y \\ 2y & 2x \end{pmatrix}$. Hence the theorem gives

$$s(\mathbf{b}) - s(\mathbf{a}) = s(b_1, b_2) - s(a_1, a_2)$$

$$\begin{aligned} &= \int_0^1 \mathbf{J}_s(a_1 + t(b_1 - a_1), a_2 + t(b_2 - a_2)) dt \cdot \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \end{pmatrix} \\ &= \begin{pmatrix} \int_0^1 2a_1 + 2t(b_1 - a_1) dt & \int_0^1 -2a_2 - 2t(b_2 - a_2) dt \\ \int_0^1 2a_2 + 2t(b_2 - a_2) dt & \int_0^1 2a_1 + 2t(b_1 - a_1) dt \end{pmatrix} \cdot \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \end{pmatrix} \\ &= \begin{pmatrix} a_1 + b_1 & -a_2 - b_2 \\ a_2 + b_2 & a_1 + b_1 \end{pmatrix} \cdot \begin{pmatrix} b_1 - a_1 \\ b_2 - a_2 \end{pmatrix}. \end{aligned}$$

You can verify that this matrix-vector product is indeed equal to

$$(b_1^2 - b_2^2, 2b_1b_2)^T - (a_1^2 - a_2^2, 2a_1a_2)^T = (b_1^2 - b_2^2 - a_1^2 + a_2^2, 2b_1b_2 - 2a_1a_2)^T,$$

as claimed.

Afternote

If you, like many other students, had difficulties to understand the meaning of “ $df(\mathbf{x})(\mathbf{b} - \mathbf{a})$ ” in the Mean Value Theorem, here are three equivalent expressions:

$$\begin{aligned} df(\mathbf{x})(\mathbf{b} - \mathbf{a}) &= \frac{\partial f}{\partial x_1}(\mathbf{x})(b_1 - a_1) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x})(b_n - a_n) \\ &= \left(\underbrace{\frac{\partial f}{\partial x_1}(\mathbf{x}) \quad \cdots \quad \frac{\partial f}{\partial x_n}(\mathbf{x})}_{\mathbf{J}_f(\mathbf{x})} \right) \begin{pmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{pmatrix} \cdot \begin{pmatrix} b_1 - a_1 \\ \vdots \\ b_n - a_n \end{pmatrix} \end{aligned}$$

Remember:

The differential of a scalar-valued function f at a particular point \mathbf{x} is a linear map from \mathbb{R}^n to \mathbb{R} , whose effect is “multiplication by the Jacobi matrix $\mathbf{J}_f(\mathbf{x})$ ” (a row vector) or, equivalently, “taking the dot product with the gradient $\nabla f(\mathbf{x})$ ” (a column vector).

Afternote Cont'd

Another point causing some headache was the following:

Defining $\phi(t) = f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$, why is the formula

$$\phi'(t) = df(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a})$$

an instance of the chain rule?

The same reasoning was used in the proof of “contours are orthogonal to the gradient”, and the details are as follows:

With $g(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ we have $\phi(t) = f(g(t))$. The chain rule gives that $d\phi(t) = \phi'(t) dt$ is the composition of $df(g(t))$ and $dg(t) = g'(t) dt = (\mathbf{b} - \mathbf{a}) dt$.

$$\begin{aligned}\implies \phi'(t)h &= d\phi(t)(h) = df(g(t))(dg(t)(h)) = df(g(t))(g'(t)h) \\ &= df(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))((\mathbf{b} - \mathbf{a})h)\end{aligned}$$

for $h \in \mathbb{R}$. Setting $h = 1$ gives the stated result.

Since $d(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$ is linear, we can rewrite the last equation as $\phi'(t)h = h\phi'(t) = hdf(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))(\mathbf{b} - \mathbf{a})$, making the result even more visible.

Another Theorem of Calculus I

Generalized

Recall that a function $f: [a, b] \rightarrow \mathbb{R}$ satisfying $f'(x) = 0$ for $x \in [a, b]$ must be constant, and similarly for parametric curves.

Definition

A subset $D \subseteq \mathbb{R}^n$ is said to be *path-connected* if for any two points $\mathbf{a}, \mathbf{b} \in D$ there exists a continuous curve $g: [0, 1] \rightarrow D$ satisfying $g(0) = \mathbf{a}$, $g(1) = \mathbf{b}$. In other words, one can continuously walk from \mathbf{a} to \mathbf{b} without ever leaving D .

Examples

Convex sets (i.e., sets $D \subseteq \mathbb{R}^n$ which contain with any two points \mathbf{a}, \mathbf{b} also the straight-line segment $[\mathbf{a}, \mathbf{b}]$) are path-connected. A further example is $\mathbb{R}^n \setminus \{\mathbf{0}\}$ for $n \geq 2$, which shows that path-connected sets may contain “holes”.

Theorem

Suppose $D \subseteq \mathbb{R}^n$ is path-connected and $f: D \rightarrow \mathbb{R}^m$ is differentiable with $df(\mathbf{x}) = 0$ for every $\mathbf{x} \in D$. Then f is constant, i.e., there exists $\mathbf{c} \in \mathbb{R}^m$ such that $f(\mathbf{x}) = \mathbf{c}$ for all $\mathbf{x} \in D$.

Proof.

Writing $f = (f_1, \dots, f_m)$, the condition $f(\mathbf{x}) = \mathbf{c}$ is equivalent to $f_i(\mathbf{x}) = c_i$ with $c_i \in \mathbb{R}$. Hence it suffices to consider the case $m = 1$.

Fix $\mathbf{a} \in D$ and let $c = f(\mathbf{a})$. We must show $f(\mathbf{b}) = c$ for all $\mathbf{b} \in D$.

By assumption there exists a continuous curve $g: [0, 1] \rightarrow D$ with $g(0) = \mathbf{a}$, $g(1) = \mathbf{b}$. Since every point $g(t)$ is an inner point of D (because f is assumed to be differentiable at $g(t)$), we can smooth g , if necessary, and assume that g' exists. Then we can apply the theorem of Calculus I to the composition $\phi: [0, 1] \rightarrow \mathbb{R}$, $t \mapsto f(g(t))$, which has derivative $\phi'(t) = df(g(t))(g'(t)) = 0$, and conclude that $f(\mathbf{b}) = \phi(1) = \phi(0) = f(\mathbf{a})$. □

The Mean
Value
Theorem and
its Friends

Error Propagation

Implicit
Functions and
their
Differentiation

Higher
Derivatives

Examples of Partial
Differential Equations

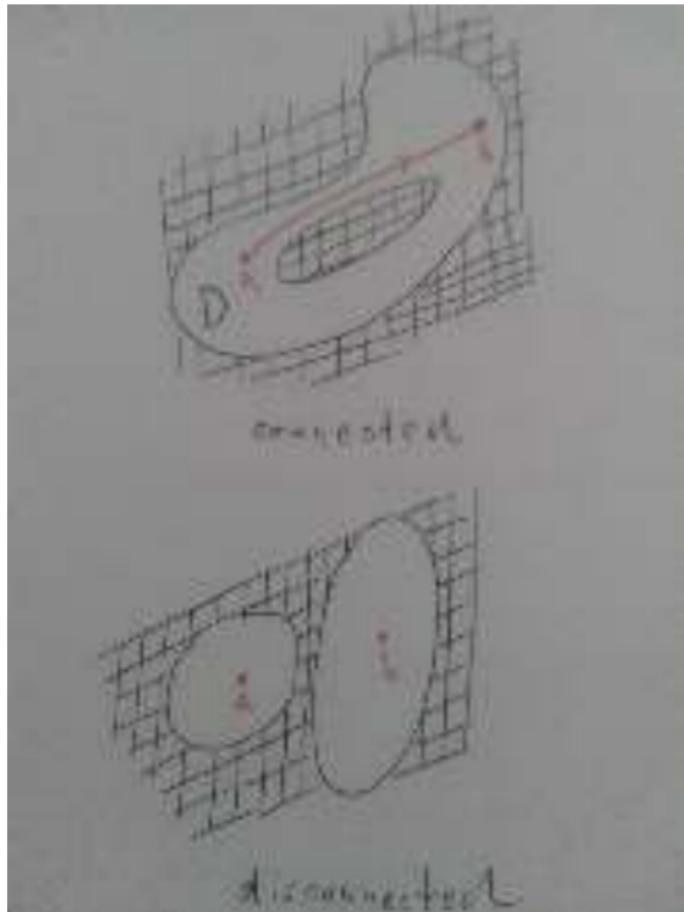


Figure: Illustration of (path-)connectivity for $n = 2$

And Yet another One

Theorem (Intermediate Value Theorem)

Suppose $D \subseteq \mathbb{R}^n$ is open and path-connected and $f: D \rightarrow \mathbb{R}$ is continuous. Then the range $f(D)$ is an interval.

This theorem holds, more generally, for continuous real-valued functions whose domain $D \subseteq \mathbb{R}^n$ is *connected* (but not necessarily open) and says that f must attain every value between any two given values $f(\mathbf{a})$ and $f(\mathbf{b})$.

Proof.

Connect **a** and **b** by a continuous curve $g: [0, 1] \rightarrow D$ and apply the intermediate value theorem for one-variable functions to $[0, 1] \rightarrow \mathbb{R}$, $t \mapsto f(g(t))$, which is continuous since it is a composition of continuous functions. □

Recall that $D \subseteq \mathbb{R}^n$ is said to be *connected* if $D = A \cup B$ with $\overline{A} \cap B = A \cap \overline{B} = \emptyset$ implies $A = \emptyset$ or $B = \emptyset$. The following (rather advanced) exercise clarifies the relation between the properties “connected” and “path-connected” for subsets of \mathbb{R}^n .

Exercise

For $D \subseteq \mathbb{R}^n$ show:

- ① If D is path-connected then D is connected.
- ② If D is connected and open then D is path-connected.
- ③ Give an example of a subset $D \subseteq \mathbb{R}^2$ that is connected but not path-connected.

Hint: Use the graph of $x \mapsto \sin(1/x)$, $x > 0$, and adjoin one or more additional points.

Error Propagation

Especially when using floating-point computations

In real-world applications we are often faced with evaluating a real-valued function $y = f(x_1, \dots, x_n)$ for some input data x_1, \dots, x_n which is not accurately known.

Question

How do small errors in the input affect the output y ?

An answer can be given using differentials. If the true input x_j is replaced by $x_j + \Delta x_j$, the corresponding change in the output y is

$$\begin{aligned}\Delta y &= f(x_1 + \Delta x_1, \dots, x_n + \Delta x_n) - f(x_1, \dots, x_n) \\ &\approx \sum_{j=1}^n \frac{\partial f}{\partial x_j}(x_1, \dots, x_n) \Delta x_j \quad \text{if the } \Delta x_j \text{ are small.}\end{aligned}$$

The Mean Value Theorem provides a rigorous answer:

$$\Delta y = \sum_{j=1}^n \frac{\partial f}{\partial x_j}(\mathbf{x} + \tau \Delta \mathbf{x}) \Delta x_j \quad \text{for some } \tau \in [0, 1],$$

where $\mathbf{x} = (x_1, \dots, x_n)$ and $\Delta \mathbf{x} = (\Delta x_1, \dots, \Delta x_n)$.

Usually we do not know the exact value of Δx_j but only a bound $|\Delta x_j| \leq \delta_j$. In this case the Mean Value Theorem yields the upper bound

$$|\Delta y| \leq \sum_{j=1}^n M_j \delta_j \quad \text{with} \quad M_j = \max_{\mathbf{x}' \in U} \left| \frac{\partial f}{\partial x_j}(\mathbf{x}') \right|,$$

where $U = \{ \mathbf{x}' \in \mathbb{R}^n; |x'_j - x_j| \leq \delta_j \text{ for } 1 \leq j \leq n \}$ is the “uncertainty region” of the input.

Example (taken from [Ste21], p. 981)

The dimensions of a rectangular box are measured as 75 cm, 60 cm, and 40 cm. Each measurement is correct within 0.2 cm. Estimate the largest possible error when the volume of the box is calculated from these measurements.

Solution: We use cm as the unit of measurement.

The differential of the volume $V = V(x, y, z) = xyz$ is

$$dV = dV(x, y, z) = yz dx + xz dy + xy dz,$$

giving the error in the volume computation as

$$\Delta V \approx dV(x, y, z)(\Delta x, \Delta y, \Delta z) = yz \Delta x + xz \Delta y + xy \Delta z.$$

Example (cont'd)

Clearly the partial derivatives take their maximum in the uncertainty region at (75.2, 60.2, 40.2).

$$\begin{aligned}\implies |\Delta V| &\leq (60.2)(40.2)(0.2) + (75.2)(40.2)(0.2) + (60.2)(75.2)(0.2) \\ &= 1994.024\end{aligned}$$

The maximum possible error in the volume computation is therefore 1994.024 cm³.

Note: We have used the fact that the uncertainty region can also be expressed as $U = \{\mathbf{x}' \in \mathbb{R}^n; |x'_j - x_j - \Delta x_j| \leq \delta_j \text{ for } 1 \leq j \leq n\}$.

Relative Errors

For practical purposes it is more useful to determine the propagation of relative errors $\Delta x_j / x_j$, because floating-point computations cause additional errors, whose relative size can be controlled. The preceding considerations give

$$\frac{\Delta y}{y} \approx \sum_{j=1}^n \frac{x_j}{y} \frac{\partial f}{\partial x_j}(x_1, \dots, x_n) \frac{\Delta x_j}{x_j} \quad \text{for small } \Delta_j,$$

and a corresponding exact version.

Example (cont'd)

The relative error of the volume computation is

$$\begin{aligned}\frac{\Delta V}{V} &\approx \frac{\Delta x}{x} + \frac{\Delta y}{y} + \frac{\Delta z}{z} \\ &= \frac{1}{5} \left(\frac{1}{75} + \frac{1}{60} + \frac{1}{40} \right) = 0.011,\end{aligned}$$

or about 1 %.

We have used the non-rigorous error estimate, because it nicely illustrates the following simple error propagation law for products:

The relative error $\frac{\Delta p}{p}$ of a product $p = x_1 x_2 \cdots x_n$ is the sum of the relative errors $\frac{\Delta x_j}{x_j}$ of its factors (in the limit $\Delta \mathbf{x} \rightarrow \mathbf{0}$).

A corresponding rigorous estimate is more complicated, since the true value $V = xyz$ of the box is not known. The Mean Value Theorem gives (ξ, η, ζ) between (x, y, z) and $(x + \Delta x, y + \Delta y, z + \Delta z)$ such that

$$\frac{\Delta V}{V} = \frac{\eta \zeta}{yz} \frac{\Delta x}{x} + \frac{\xi \zeta}{xz} \frac{\Delta y}{y} + \frac{\xi \eta}{xy} \frac{\Delta z}{z},$$

and one can estimate the coefficients as $\frac{\eta \zeta}{yz} \leq \frac{(60.2)(40.2)}{(59.8)(39.8)}, \dots$

Remark

The preceding example is a toy example and it is easy to determine the maximum/minimum absolute error directly:

$$\begin{aligned}\Delta V &\leq (75.2)(60.2)(40.2) - 75 \cdot 60 \cdot 40 \approx 1987, \\ \Delta V &\geq (74.8)(59.8)(39.8) - 75 \cdot 60 \cdot 40 \approx -1973.\end{aligned}$$

Perhaps it is instructive to see how the bound resulting from the Mean Value Theorem overestimates the error (with $h = 0.2$):

$$\begin{aligned}|\Delta V| &\leq (y + h)(z + h)h + (x + h)(z + h)h + (x + h)(y + h)h \\ &= (yz + xz + xy)h + (x + y + z)2h^2 + 3h^3,\end{aligned}$$

whereas direct expansion gives

$$\begin{aligned}|\Delta V| &\leq (x + h)(y + h)(z + h) - xyz \\ &= (yz + xz + xy)h + (x + y + z)h^2 + h^3.\end{aligned}$$

In most cases, however, direct expansion of the absolute error is impossible (too complicated), and the bound resulting from the Mean Value Theorem is a viable alternative.

Floating-point numbers

Definition

A t -digit floating-point number (fpn) is a real number of the form $\pm m \times b^e$, where $b \geq 2$ is a fixed integer (the *base*), e is an integer from some finite range $e_{\min} \leq e \leq e_{\max}$ (*exponent*) and $m > 0$ (*mantissa*) is a number that has a base- b representation using t digits.

Notes

- The representation is usually normalized to $m \in [1, b)$, so that the base- b representation of m has the form $m_1.m_2m_3\dots m_t = \sum_{i=1}^t m_i b^{1-i}$ with $1 \leq m_1 \leq b - 1$ and $0 \leq m_i \leq b - 1$ for $i \geq 2$.
- The number b^{1-t} , which represents the distance from 1 to the smallest fpn > 1 , is called *machine epsilon* and denoted by ε_M .

Example

Double-precision floating-point approximations to π and π^{100} in SageMath :

$$\pi \approx 3.14159265358979, \quad \pi^{100} \approx 5.18784831431959e49$$

They are written as 15-digit decimal fpn on the screen, but the internal representation uses binary 53-digit fpn according to the IEEE double-precision standard.

Fact

The standard arithmetic functions (including square roots, exp, log, sin, cos, etc.) can be implemented in such a way that the result \hat{y} of a computation equals the exact result y rounded to the nearest floating-point number (except when overflow/underflow occurs).

This gives for the relative error $\epsilon = \frac{\hat{y}-y}{y}$ the bound $\epsilon \leq \varepsilon_M$, or

$$\hat{y} = y(1 + \epsilon) \quad \text{with} \quad |\epsilon| \leq \varepsilon_M.$$

Observation

For iterated computations the error in the output of one computation acts as error in the input data of the next computation, and these errors propagate to errors in the final result.

Example

When computing the sum of 3 numbers x_1, x_2, x_3 in floating-point arithmetic, we have

$$\widehat{x_1 + x_2} = (x_1 + x_2)(1 + \epsilon_1), \quad (|\epsilon_1| \leq \varepsilon_M)$$

$$\begin{aligned}\widehat{x_1 + x_2 + x_3} &= ((x_1 + x_2)(1 + \epsilon_1) + x_3)(1 + \epsilon_2) \quad (|\epsilon_2| \leq \varepsilon_M) \\ &= x_1(1 + \epsilon_1)(1 + \epsilon_2) + x_2(1 + \epsilon_1)(1 + \epsilon_2) + x_3(1 + \epsilon_2)\end{aligned}$$

Remarks

- The modern point-of-view emphasizes so-called *backwards analysis*: The approximate result of a machine computation is exact for (slightly) changed input data. For example, the previous computation returned the exact result of adding the 3 numbers $x_1(1 + \epsilon_1)(1 + \epsilon_2)$, $x_2(1 + \epsilon_1)(1 + \epsilon_2)$ and $x_3(1 + \epsilon_2)$. Since these differ from x_1, x_2, x_3 only by a small relative error ($\leq 2\varepsilon_M$), the operation is *well-conditioned*.
- Of the 4 basic arithmetic operations, only subtraction $s(x, y) = x - y$ is ill-conditioned (and only if the numbers involved have approximately the same size).
In order to understand this, we estimate the propagation of the relative error: Since $ds = dx - dy$, we have

$$\Delta s \approx \Delta x - \Delta y,$$

$$\frac{\Delta s}{s} \approx \frac{x}{x-y} \frac{\Delta x}{x} - \frac{y}{x-y} \frac{\Delta y}{y}.$$

It is an instructive exercise to compute $1 - 0.999999$ in 6-digit decimal floating-point arithmetic and compare it with the exact result.

Implicit Differentiation

For a differentiable function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^2$, the gradient $\nabla f(\mathbf{x})$ is perpendicular to the contour line at every point $\mathbf{x} \in D$, as we have seen. Virtually the same proof shows that in the general case $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$ the gradient $\nabla f(\mathbf{x})$ (provided it is nonzero) serves as a normal vector to the contour hypersurface of f at \mathbf{x} . Moreover we can compute the partial derivatives (and hence the differential) of any function g implicitly defined by the contour hypersurface. For this we consider a 3-dimensional example, which is a slight variant of one in our textbook [Ste21].

Example

Find the partial derivatives of the function $z = g(x, y)$ defined implicitly by the 2-dimensional surface S with equation

$$x^3 + y^3 + z^3 + 3xyz = 1 \quad \text{in } \mathbb{R}^3.$$

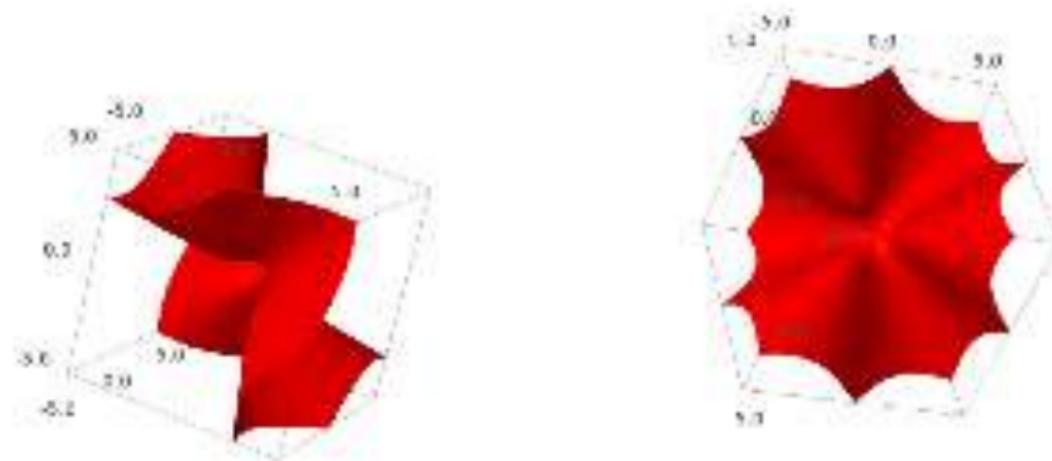


Figure: The surface S from different perspectives

The second picture reveals that S is invariant under a certain rotation of \mathbb{R}^3 by an angle of 120° .

Example (cont'd)

Let $F(x, y, z) = x^3 + y^3 + z^3 + 3xyz - 1$. Then the required function g must satisfy $F(x, y, g(x, y)) = 0$.

Applying the Chain Rule to F and $G(x, y) = (x, y, g(x, y))^T$ (recall that vector-valued functions must be written as column vectors when computing their Jacobi matrices!) gives, writing $z = g(x, y)$ and later suppressing arguments,

$$0 = (F \circ G)(x, y), \quad (\text{for all } (x, y) \text{ in the domain of } g)$$

$$\implies 0 = dF(x, y, g(x, y)) \circ dG(x, y), \quad (\text{zero linear map})$$

$$\implies \mathbf{0} = \mathbf{J}_F(x, y, g(x, y)) \mathbf{J}_G(x, y) \quad (\text{all-zero matrix})$$

$$= (F_x \quad F_y \quad F_z) \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ g_x & g_y \end{pmatrix} = (F_x + F_z g_x \quad F_y + F_z g_y).$$

$$\implies g_x = -\frac{F_x}{F_z} = -\frac{3x^2 + 3yz}{3z^2 + 3xy} = -\frac{x^2 + yz}{z^2 + xy},$$

$$g_y = -\frac{F_y}{F_z} = -\frac{3y^2 + 3xz}{3z^2 + 3xy} = -\frac{y^2 + xz}{z^2 + xy}.$$

Example (cont'd)

How to find the domain of g ?

A qualitative answer is provided by the

Implicit Function Theorem (special case):

Suppose $P_0 = (x_0, y_0, z_0)$ satisfies $F(x_0, y_0, z_0) = 0$ and $F_z(x_0, y_0, z_0) \neq 0$, i.e., P_0 is on the surface and the normal vector to the surface at P_0 is not contained in the x, y -plane. Then there exists a neighborhood of P_0 of the form $U' \times U''$ with

$$U' = \{(x, y) \in \mathbb{R}^2; |x - x_0| < \delta' \wedge |y - y_0| < \delta'\},$$
$$U'' = \{z \in \mathbb{R}; |z - z_0| < \delta''\}, \quad \delta', \delta'' > 0,$$

and a C^1 -function $g: U' \rightarrow U''$ such that the part of the surface contained in $U' \times U''$ is precisely the graph of g , i.e.,

$$F(x, y, z) = 0 \wedge (x, y, z) \in U' \times U'' \iff z = g(x, y)$$

This theorem, which mutatis mutandis applies to any C^1 -function $F: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, and any of its n variables (in place of z), is a mere existence result and doesn't provide any explicit description of U' , U'' , and g .

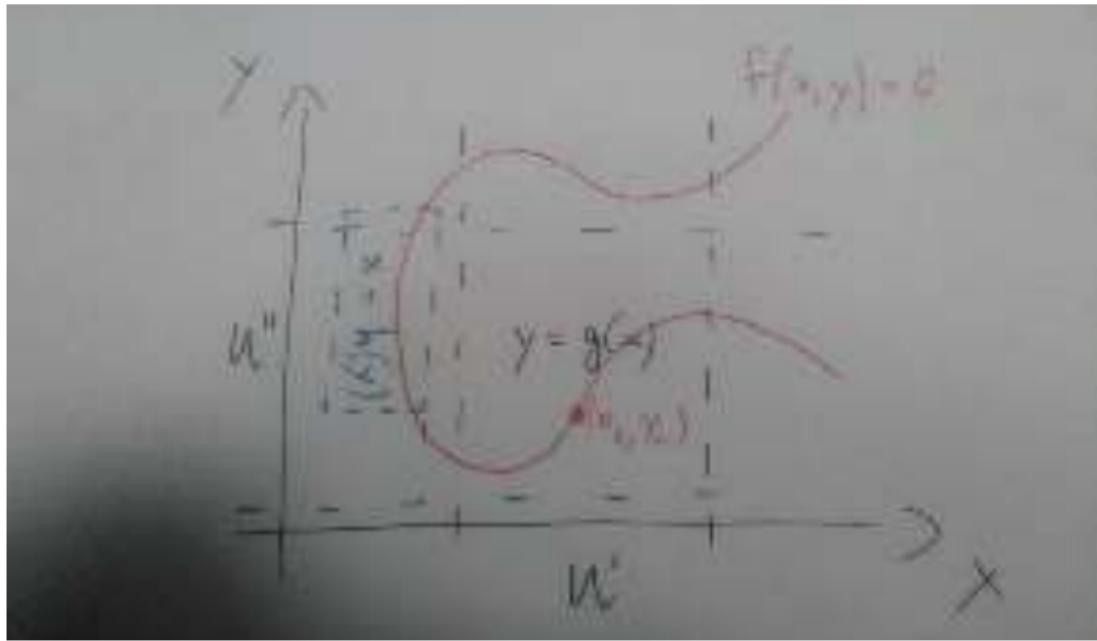


Figure: Illustration of the Implicit Function Theorem for $n = 2$

If the curve C defined by $f(x, y) = 0$ is smooth then locally at every point $(x_0, y_0) \in C$ there exists a representation as graph of a function $y = g(x)$ or $x = h(y)$ (or both).

Example (cont'd)

Now we determine an appropriate neighborhood $U' \times U''$ for the point $P_0 = (0, 0, 1)$, which satisfies the condition

$$F_z(0, 0, 1) = 3 \neq 0.$$

For fixed x, y consider the function

$$h(z) = h_{x,y}(z) = F(x, y, z) = x^3 + y^3 + z^3 + 3xyz - 1.$$

We have

$$h'(z) = F_z(x, y, z) = 3(z^2 + xy) > 0 \quad \text{if } |x| < \frac{1}{2}, |y| < \frac{1}{2}, |z - 1| < \frac{1}{2}.$$

$\Rightarrow h = h_{x,y}$ is strictly increasing on $(\frac{1}{2}, \frac{3}{2})$, provided we restrict (x, y) to $U' = \{(x, y) \in \mathbb{R}^2; |x| < \frac{1}{2}, |y| < \frac{1}{2}\}$. Moreover, for those x, y, z we have

$$h\left(\frac{1}{2}\right) = x^3 + y^3 + \frac{3}{2}xy - \frac{7}{8} \leq \frac{1}{8} + \frac{1}{8} + \frac{3}{2} \cdot \frac{1}{4} - \frac{7}{8} < 0,$$

$$h\left(\frac{3}{2}\right) = x^3 + y^3 + \frac{9}{2}xy + \frac{19}{8} \geq -\frac{1}{8} - \frac{1}{8} - \frac{9}{2} \cdot \frac{1}{4} + \frac{19}{8} > 0,$$

showing that the unique solution of $h(z) = 0$ falls into the interval $(\frac{1}{2}, \frac{3}{2})$. \Rightarrow We can take U' as above and $U'' = (\frac{1}{2}, \frac{3}{2})$.

Example (cont'd)

Note that a similar argument will work for any C^1 -function $F: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$.

It remains to show that g is continuously differentiable. This can be done as follows. Applying the Mean Value Theorem to F , we have for any $(x, y) \in U'$ and s, t sufficiently small an identity

$$\begin{aligned} 0 &= F(x + s, y + t, g(x + s, y + t)) - F(x, y, g(x, y)) \\ &= F_x(\xi, \eta, \zeta)s + F_y(\xi, \eta, \zeta)t + F_z(\xi, \eta, \zeta)(g(x + s, y + t) - g(x, y)), \end{aligned}$$

where (ξ, η, ζ) is a point between $(x, y, g(x, y))$ and $(x + s, y + t, g(x + s, y + t))$. Since F_x , F_y , and $1/F_z$ are continuous on $U' \times U''$, they are bounded, and hence letting $(s, t) \rightarrow (0, 0)$ in the above identity gives $g(x + s, y + t) \rightarrow g(x, y)$, i.e., the continuity of g . Moreover, setting $t = 0$ we get

$$\frac{g(x + s, y) - g(x, y)}{s} = -\frac{F_x(\xi, \eta, \zeta)}{F_z(\xi, \eta, \zeta)} \rightarrow -\frac{F_x(x, y, g(x, y))}{F_z(x, y, g(x, y))} \text{ for } s \rightarrow 0.$$

$\implies \frac{\partial g}{\partial x}$ (and similarly $\frac{\partial g}{\partial y}$) exists and is continuous, because it is composed of continuous functions. $\implies g$ is a C^1 -function.

Example (cont'd)

Why is S "smooth"?

The gradient ∇F doesn't vanish at any point of S , since the only solution of

$$\nabla F(x, y, z) = 0 \iff x^2 + yz = y^2 + xz = z^2 + xy = 0$$

is $(0, 0, 0)$, but $(0, 0, 0) \notin S$.

\implies At every point of S at least one of the representations

$z = g(x, y)$, $y = h(x, z)$ or $x = k(y, z)$ exists.

$\implies S$ has a tangent plane at every point.

How to obtain the tangent plane?

The tangent plane to S in $P_0 = (x_0, y_0, z_0)$ has normal vector $\nabla F(x_0, y_0, z_0)$ and hence the equation

$$F_x(x_0, y_0, z_0)(x - x_0) + F_y(x_0, y_0, z_0)(y - y_0) + F_z(x_0, y_0, z_0)(z - z_0) = 0.$$

For example, $\nabla F(0, 0, 1) = (0, 0, 1)$ gives for the tangent plane at $P_0 = (0, 0, 1)$ the equation $0 \cdot x + 0 \cdot y + 3(z - 1) = 0$ or, simplified, $z - 1 = 0$.

The Mean
Value
Theorem and
its Friends

Error Propagation

Implicit
Functions and
their
Differentiation

Higher
Derivatives

Examples of Partial
Differential Equations

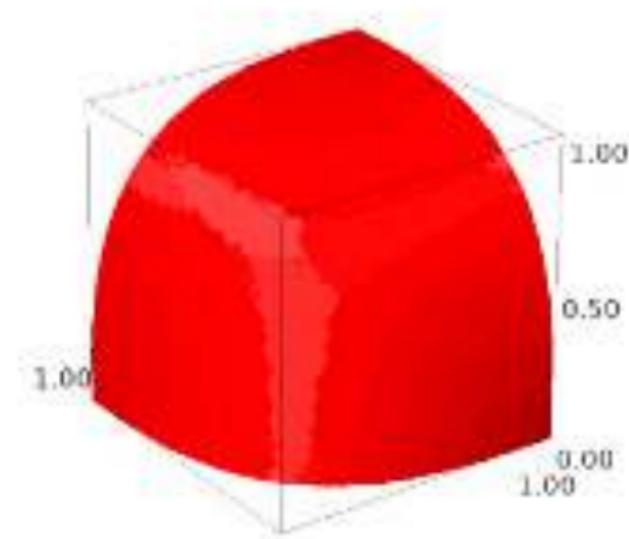


Figure: Intersection of S with the unit cube in \mathbb{R}^3

Higher Derivatives

If $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, is partially differentiable, then $f_{x_j} = \frac{\partial f}{\partial x_j}$ also has domain D and codomain \mathbb{R} . Asking whether the partial derivatives are itself (partially) differentiable, leads to the notion of k -th order partial derivatives for $k = 1, 2, 3, \dots$ (as in Calculus I).

It will be convenient to consider partial differentiation with respect to x_j as an *operator* on the set of functions $f: D \rightarrow \mathbb{R}$, because higher derivatives then correspond to iterating such operators.

Notation

- For 1st-order partial derivatives we use the following additional notation: $D_j f = \frac{\partial}{\partial x_j} f = \frac{\partial f}{\partial x_j} = f_{x_j}$. Thus $D_j = \frac{\partial}{\partial x_j}$ is considered as a map from functions to functions with the same (or smaller) domain.
- For 2nd-order partial derivatives we write

$$D_1 D_1 f = D_1(D_1 f) = \frac{\partial^2 f}{\partial x_1^2} = (f_{x_1})_{x_1} = f_{x_1 x_1},$$

$$D_2 D_1 f = D_2(D_1 f) = \frac{\partial^2 f}{\partial x_2 \partial x_1} = (f_{x_1})_{x_2} = f_{x_1 x_2}, \quad \text{etc.}$$

Example

We compute the higher-order partial derivatives of $f(x, y) = x^3 - 3xy^2$.

$$f_x = 3x^2 - 3y^2, \quad f_y = -6xy$$

$$f_{xx} = 6x, \quad f_{xy} = -6y$$

$$f_{yx} = -6y, \quad f_{yy} = -6x$$

$$f_{xxx} = 6, \quad f_{xxy} = 0$$

$$f_{xyx} = 0, \quad f_{xyy} = -6$$

$$f_{yxx} = 0, \quad f_{yxy} = -6$$

$$f_{yyx} = -6, \quad f_{yyy} = 0$$

One observes that the order in which the operators D_x and D_y are applied does not matter, viz. $f_{xy} = f_{yx}$, $f_{xxy} = f_{xyx} = f_{yxx}$, $f_{xyy} = f_{yxy} = f_{yyx}$. This is no coincidence!

Theorem (Clairaut's Theorem)

Suppose $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^2$, has continuous second-order partial derivatives. Then $f_{xy} = f_{yx}$.

An equivalent statement is the following: If f is partially differentiable in some open disk around (a, b) and the partial derivatives f_{xy} , f_{yx} are continuous on the whole disk, then $f_{xy}(a, b) = f_{yx}(a, b)$.

Corollary

Suppose $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, has continuous k -th order partial derivatives. Then

$$D_{j_1} D_{j_2} \cdots D_{j_k} f = D_{j_{\pi(1)}} D_{j_{\pi(2)}} \cdots D_{j_{\pi(k)}}$$

for any $(j_1, j_2, \dots, j_k) \in \{1, \dots, n\}^k$ and any permutation π of $\{1, 2, \dots, k\}$.

Hence we can write $D_{j_1} D_{j_2} \cdots D_{j_k} = D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n}$, where α_j denotes the number of indexes $s \in \{1, \dots, k\}$ such that $j_s = j$.

Proof of Clairaut's Theorem.

The limit

$$L = \lim_{h \rightarrow 0} \frac{f(a+h, b+h) - f(a, b+h) - f(a+h, b) + f(a, b)}{h^2}$$

is a 2-dimensional analogon of the limit used to define derivatives in Calculus I. We evaluate this limit in two different ways, using the Mean Value Theorem of Calculus I.

- ① Consider $g_1(x) = f(x, b+h) - f(x, b)$:

$$\begin{aligned} L &= \lim_{h \rightarrow 0} \frac{g_1(a+h) - g_1(a)}{h^2} = \lim_{h \rightarrow 0} \frac{g'_1(\xi_1)h}{h^2} \\ &= \lim_{h \rightarrow 0} \frac{f_x(\xi_1, b+h) - f_x(\xi_1, b)}{h} = \lim_{h \rightarrow 0} f_{xy}(\xi_1, \eta_1) = f_{xy}(a, b) \end{aligned}$$

- ② Consider $g_2(y) = f(a+h, y) - f(a, y)$:

$$\begin{aligned} L &= \lim_{h \rightarrow 0} \frac{g_2(b+h) - g_2(b)}{h^2} = \lim_{h \rightarrow 0} \frac{g'_2(\eta_2)}{h} \\ &= \lim_{h \rightarrow 0} \frac{f_y(a+h, \eta_2) - f_y(a, \eta_2)}{h} = \lim_{h \rightarrow 0} f_{yx}(\xi_2, \eta_2) = f_{yx}(a, b) \quad \square \end{aligned}$$

The Mean
Value
Theorem and
its Friends
Error Propagation

Implicit
Functions and
their
Differentiation

Higher
Derivatives

Examples of Partial
Differential Equations

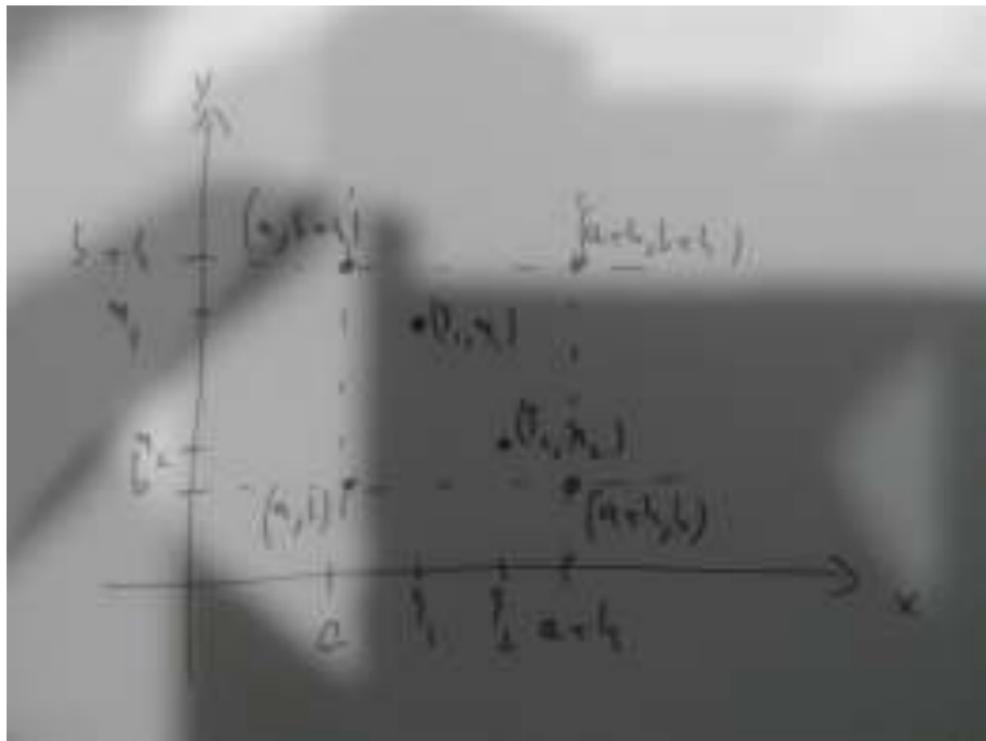


Figure: Illustration of the proof of Clairaut's Theorem

For $h \rightarrow 0$ the auxiliary points (ξ_1, η_1) and (ξ_2, η_2) tend to (a, b) .

Partial Differential Equations

Partial differential equations or *PDE's*, for short, are equations relating the partial derivatives (possibly of higher order and/or order zero) of a function. Often (but not always) they express a law of Physics, like in the following two examples.

Laplace's Equation

The 2-dimensional Laplace equation is

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

This may also be expressed as

$$\Delta u = (D_1^2 + D_2^2)(u) = D_1(D_1 u) + D_2(D_2 u) = 0$$

with $\Delta = D_1^2 + D_2^2$. The differential operator $\Delta = D_1^2 + D_2^2$, or $\Delta = D_1^2 + \cdots + D_n^2$ in general, is called *Laplace operator*. Solutions of the Laplace equation $\Delta u = 0$ are called *harmonic functions*.

Example

Show that $u(x, y) = e^x \cos y$, $(x, y) \in \mathbb{R}^2$, is a solution of the 2-dimensional Laplace equation.

$$u_x = e^x \cos y = u_{xx}$$

$$u_y = -e^x \sin y, \quad u_{yy} = -e^x \cos(y)$$

$$\implies u_{xx} + u_{yy} = 0$$

Note

Since partial derivatives and linear combinations (with constant coefficients!) of solutions of $\Delta u = 0$ are again solutions, all functions of the form

$$u(x, y) = A e^x \cos y + B e^x \sin y, \quad A, B \in \mathbb{R}$$

are solutions of $\Delta u = 0$ (i.e., harmonic).

Example (cont'd)

Question: How can one find such solutions in the first place?

The idea is to try the „Ansatz“

$$u(x, y) = f(x)g(y) \quad \text{for some one-variable functions } f, g.$$
$$\implies u_{xx} + u_{yy} = f''(x)g(y) + f(x)g''(y) \stackrel{!}{=} 0.$$

At points (x, y) with $u(x, y) \neq 0$ this is equivalent to

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

Since the right-hand side is independent of x , we obtain

$f''(x) = k f(x)$, $g''(y) = -k g(y)$ for some constant $k \in \mathbb{R}$ and all x, y in the domain of f resp. g , which are assumed to be intervals. Zeros of f or g make no exception.

For $k = 1$ the solutions of these differential equations are $f(x) = c_1 e^x + c_2 e^{-x}$, $g(y) = c_3 \cos(y) + c_4 \sin(y)$.

$\implies (x, y) \mapsto e^{\pm x} \cos(y), e^{\pm x} \sin(y)$ and all linear combinations thereof solve the Laplace equation. Other values of k yield further solutions.

The Preceding Example Generalized

The functions $e^x \cos y$ and $e^x \sin y$ arise as real and imaginary part of the complex exponential function $z \mapsto e^z$:

$$e^z = e^{x+iy} = e^x e^{iy} = e^x (\cos y + i \sin y) = e^x \cos y + i e^x \sin y$$

They are harmonic because $z \mapsto e^z$ has a complex derivative:

$$\lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{C}}} \frac{e^{z+h} - e^z}{h} = \lim_{h \rightarrow 0} \frac{e^z e^h - e^z}{h} = e^z \lim_{h \rightarrow 0} \frac{e^h - 1}{h} = e^z.$$

Definition

A function $f: D \rightarrow \mathbb{C}$ on an open set $D \subseteq \mathbb{C}$ is said to be *holomorphic* if $f'(z) = \lim_{h \rightarrow 0} \frac{f(z+h)-f(z)}{h}$ exists for every $z \in D$.

Examples

$z \mapsto e^z$; polynomials and rational functions with coefficients in \mathbb{C} ;
power series within their circle of convergence.

Notes

- The theory of holomorphic functions, called *Complex Analysis*, is fundamentally different from the theory of differentiable functions of one real variable. For example, one can show that for a holomorphic function f the derivative f' is again holomorphic, implying the existence of all derivatives of higher order!
- Since $\mathbb{C} \triangleq \mathbb{R}^2$ via $x + iy \mapsto (x, y)$, we can ask about the relation between holomorphic functions and differentiable functions $\mathbb{R}^2 \rightarrow \mathbb{R}^2$. The answer is as follows:
If $f = (u, v) = u + iv$ (i.e., $u = \operatorname{Re} f$, $v = \operatorname{Im} f$) is holomorphic, we have $df(z) = f'(z) dz$. In particular, f and hence its component functions u, v are differentiable, and

$$\begin{aligned}f_x(z) &= df(z)(\mathbf{e}_1) = df(z)(1) = f'(z), \\f_y(z) &= df(z)(\mathbf{e}_2) = df(z)(i) = if'(z).\end{aligned}$$

Since $f_x = (u_x, v_x) = u_x + iv_x$, and similarly for f_y , this shows

$$u_y + iv_y = i(u_x + iv_x) = iu_x - v_x, \quad \text{i.e.,} \quad u_x = v_y \wedge u_y = -v_x.$$

Notes cont'd

- (item cont'd)

⇒ The Jacobi matrix $\mathbf{J}_f = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}$ has the form $\begin{pmatrix} u_x & -v_x \\ v_x & u_x \end{pmatrix}$.

Conversely, if $f = (u, v)$ is differentiable and its partial derivatives satisfy the conditions $u_x = v_y$, $u_y = -v_x$ then

$$\begin{aligned} df(z)(h_1 + ih_2) &= (u_x(z)h_1 + u_y(z)h_2, v_x(z)h_1 + v_y(z)h_2) \\ &= (u_x(z)h_1 - v_x(z)h_2, v_x(z)h_1 + u_x(z)h_2) \\ &= (u_x(z) + iv_x(z))(h_1 + ih_2) \end{aligned}$$

i.e., $df(z)$ is multiplication by $u_x(z) + iv_x(z)$, and

$$\begin{aligned} \frac{f(z+h) - f(z)}{h} &= \frac{df(z)(h) + o(h)}{h} = u_x(z) + iv_x(z) + \frac{o(h)}{h} \\ &\longrightarrow u_x(z) + iv_x(z) \quad \text{for } h \rightarrow 0, \end{aligned}$$

so that f is holomorphic with $f'(z) = u_x(z) + iv_x(z)$.

Notes cont'd

- If $f = (u, v) = u + iv$ is holomorphic then u and v are harmonic. For u , e.g., this follows easily from

$$u_{xx} = (u_x)_x = (v_y)_x = v_{yx},$$

$$u_{yy} = (u_y)_y = (-v_x)_y = -v_{xy},$$

and Clairaut's Theorem.

Here we are assuming that u (or v) has continuous 2nd-order partial derivatives, which is required for the application of Clairaut's Theorem. This can in fact be shown, but it is not at all easy. (Essentially this property is equivalent to the surprising "if f is holomorphic then f' is holomorphic".)

Example

It is an instructive exercise to determine which polynomial functions $p(x, y)$ in two variables of small degree d are harmonic. Since Δ maps monomials of degree n to monomials of degree $n - 2$, it suffices to consider this question for homogeneous polynomials of degree d .

$d \leq 1$ Polynomials of degree ≤ 1 are obviously harmonic.

$d = 2$ You can check that $p(x, y) = ax^2 + bxy + cy^2$ is harmonic iff $c = -a$; i.e., iff $p(x, y)$ is a linear combination of $x^2 - y^2$ and xy .

$d = 3$ We do this case in detail. For

$$p(x, y) = ax^3 + bx^2y + cxy^2 + dy^3 \text{ we have}$$

$$\begin{aligned} p_x &= 3ax^2 + 2bxy + cy^2, & p_y &= bx^2 + 2cxy + 3dy^2, \\ p_{xx} &= 6ax + 2by, & p_{yy} &= 2cx + 6dy. \end{aligned}$$

It follows that $\Delta p = (6a + 2c)x + (2b + 6d)y = 0 \iff c = -3a \wedge b = -3d$; i.e., $p(x, y) = a(x^3 - 3xy^2) + d(y^3 - 3x^2y)$.

Compare this with $z^2 = (x + yi)^2 = x^2 - y^2 + 2xyi$,
 $z^3 = (x + yi)^3 = x^3 - 3xy^2 + (3x^2y - y^3)i$.

Exercise

Show that a homogeneous polynomial $p(x, y)$ of degree 4 is harmonic iff $p(x, y)$ is a linear combination of $x^4 - 6x^2y^2 + y^4$ and $x^3y - y^3x$. Compare this with $\operatorname{Re}(z^4)$ and $\operatorname{Im}(z^4)$, $z = x + yi$.

Exercise

The example and the previous exercise suggest a recipe for obtaining all homogeneous polynomials $p(x, y)$ of degree d that are harmonic. Formulate this recipe, and prove its validity.

Wave equation

The 2-dimensional wave equation is

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}, \quad \text{where } a > 0 \text{ is a constant.}$$

Its solutions $u(x, t)$ describe, for example, the displacement of a vibrating string at time t and distance x from one end of the string.

Example

Show that all functions of the form

$$u(x, t) = \frac{A_0}{2} + \sum_{k=1}^n \left[A_k \cos(k(x - at)) + B_k \sin(k(x - at)) \right], \quad A_i, B_i \in \mathbb{R}$$

are solutions of the 2-dimensional wave equation.

It suffices to show this for $(x, t) \mapsto \cos(k(x - at))$ and $(x, t) \mapsto \sin(k(x - at))$. Let, e.g., $u(x, t) = \cos(k(x - at))$.

$$u_x = -k \sin(kx - kat), \quad u_{xx} = -k^2 \cos(kx - kat),$$

$$u_t = k a \sin(kx - kat), \quad u_{tt} = -k^2 a^2 \cos(kx - kat),$$

$$\implies u_{tt} = a^2 u_{xx}.$$

Example (cont'd)

The functions $(x, t) \mapsto \cos(k(x - at)), \sin(k(x - at))$ provide solutions of the wave equation (with given constant a^2) for all $k \in \mathbb{R}$ and can be found in a way similar to that illustrated for the Laplace equation.

The condition $k \in \mathbb{Z}$ makes these functions periodic in x (with period 2π), which is needed for a solution of the vibrating string problem.

Math 241
Calculus III

Thomas
Honold

Brief
Introduction to
Optimization

Unconstrained
Optimization

Optimization
with Equality
Constraints

Appendix:
Proofs of
Important
Theorems of
1-Variable
Calculus

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Brief Introduction to Optimization

2 Unconstrained Optimization

3 Optimization with Equality Constraints

4 Appendix: Proofs of Important Theorems of 1-Variable Calculus

Brief
Introduction to
Optimization

Unconstrained
Optimization

Optimization
with Equality
Constraints

Appendix:
Proofs of
Important
Theorems of
1-Variable
Calculus

Today's Lecture: Maxima and Minima

Five Optimization Problems

- 1 Solve the linear program

Minimize $3x + 5y$

subject to $2x + y \geq 3$

$$2x + 2y \geq 5$$

$$x + 4y \geq 4$$

$$x, y \geq 0$$

- 2 Determine the maxima and minima of the function $f(x, y) = x^3 - xy + y^3$ on the closed unit square in \mathbb{R}^2 .
- 3 Determine the maxima and minima of the function f in Problem 2 on the closed unit disk in \mathbb{R}^2 .
- 4 Determine the maxima of $g(x, y) = (x + y)e^{-x^2-y^2}$ on \mathbb{R}^2 .
- 5 A rectangular box without lid is to be made from 75 cm^2 of cardboard. Determine the maximum volume of such a box.

Problem 1 belongs to Linear Optimization and will not be considered in this course (except that you will be invited to solve it!). Our present goal is to solve Problems 2–5 (and similar optimization problems).

Some Terminology

Optimization Problem Find a maximum/minimum (or all maxima/minima) of some real-valued function f on some subset S of its domain D . (In our discussion we assume that D and hence S are subsets of \mathbb{R}^n for some positive integer n .)

Feasible solution Elements $\mathbf{x} \in S$ are called *feasible solutions* of the optimization problem. The set S is called the *feasible region*.

Optimal solution For a maximization problem an *optimal solution* is an element $\mathbf{x}^* \in S$ satisfying $f(\mathbf{x}^*) \geq f(\mathbf{x})$ for all $\mathbf{x} \in S$. Such an element \mathbf{x}^* is also referred to as a *global maximum* of f on S . If $f(\mathbf{x}^*) > f(\mathbf{x})$ for all $\mathbf{x} \in S \setminus \{\mathbf{x}^*\}$, the maximum is *strict*.

Extremum Maximum or minimum

Global extremum See explanation under “optimal solution”.

Local extremum An element $\mathbf{x}_0 \in S$ is a *local maximum* of f on S if there exists a neighborhood N of \mathbf{x}_0 such that $f(\mathbf{x}_0) \geq f(\mathbf{x})$ for all $\mathbf{x} \in S \cap N$ (i.e., \mathbf{x}_0 is a global maximum of f on $S \cap N$). In our case the neighborhood can be taken as a ball $B_r(\mathbf{x}_0)$.

Some Terminology (cont'd)

Objective function The function f that is maximized or minimized.

Constraints The conditions defining S (usually given in the form $g(\mathbf{x}) \leq 0$ or $g(\mathbf{x}) = 0$ for some function(s) $g: D \rightarrow \mathbb{R}$).

Unconstrained optimization This refers to the case $S = D$ and usually entails that D is open in \mathbb{R}^n .

Constrained optimization This indicates the presence of at least one constraint, and hence $S \subsetneq D$.

Greedy method An optimization algorithm that at each step selects a locally optimal feasible solution is called a *greedy algorithm* (greedy = “get as much as you can”). For a maximization problem, if f is differentiable and the current feasible solution \mathbf{x} is an inner point of S then a greedy algorithm proceeds in the direction of the gradient $\nabla f(\mathbf{x})$ (*gradient method*).

The Unconstrained Case

Theorem

Suppose $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, has a local extremum in $\mathbf{x}_0 \in D^\circ$ (i.e., \mathbf{x}_0 is an inner point of D). If f is partially differentiable at \mathbf{x}_0 then $\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = 0$ for $1 \leq j \leq n$ or, equivalently, $\nabla f(\mathbf{x}_0) = \mathbf{0} \in \mathbb{R}^n$.

Proof.

Suppose w.l.o.g. that the extremum is a maximum. Choose a ball $B_r(\mathbf{x}_0) \subseteq D$ such that $f(\mathbf{x}_0) \geq f(\mathbf{x})$ for all $\mathbf{x} \in B_r(\mathbf{x}_0)$ (possible, since $\mathbf{x}_0 \in D^\circ$) and consider the functions $g_j: (-r, r) \rightarrow \mathbb{R}$, $t \mapsto f(\mathbf{x}_0 + t\mathbf{e}_j)$.

Clearly g_j has a local extremum at $t = 0$. Hence $g'_j(0) = 0$, as we know from Calculus I.

But $g'_j(0) = \frac{\partial f}{\partial x_j}(\mathbf{x}_0)$, and so $\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = 0$. □

Definition

$\mathbf{x}_0 \in D$ is said to be *critical point* (or *stationary point*) of f if $\nabla f(\mathbf{x}_0) = \mathbf{0}$.

The preceding theorem says that in unconstrained optimization a necessary condition for a local extremum of f at \mathbf{x} is the vanishing of the gradient (or the differential) at \mathbf{x} .

Example

Consider the function $f(x, y) = x^2 + y^2$, $D = \mathbb{R}^2$, which clearly has a global minimum in $(0, 0)$.

We have $\nabla f(x, y) = (2x, 2y)$ and $\nabla f(0, 0) = (0, 0)$, in accordance with the preceding theorem.

Question

How can we decide in general, whether a function at some already known critical point has a maximum or minimum?

As for functions of one variable, the answer is provided by second order Taylor approximation.

Recall that for a C^2 -function $f: I \rightarrow \mathbb{R}$, $I \subseteq \mathbb{R}$, and $t_0 \in I$ we have

$$f(t) = f(t_0) + f'(t_0)(t - t_0) + \frac{f''(t_0)}{2}(t - t_0)^2 + o((t - t_0)^2) \quad \text{for } t \rightarrow t_0.$$

At a critical point t_0 this reduces to

$$f(t) = f(t_0) + \frac{f''(t_0)}{2}(t - t_0)^2 + o((t - t_0)^2),$$

so that the question can be decided by looking at the sign of $f''(t_0)$, provided that $f''(t_0) \neq 0$.

Quadratic Approximation for Multivariable Functions

From the preceding example you may guess that the 2nd-order approximation to a C^2 -function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, in $\mathbf{x}_0 \in D^\circ$ involves the 2nd-order partial derivatives $f_{x_i x_j}(\mathbf{x}_0)$.

In order to save space, we define:

Definition

The matrix $\mathbf{H}_f(\mathbf{x}_0) = (h_{ij}) \in \mathbb{R}^{n \times n}$ whose entries $h_{ij} = f_{x_i x_j}(\mathbf{x}_0)$ are the partial derivatives of f at \mathbf{x}_0 is called *Hesse matrix* of f at \mathbf{x}_0 .

Note that by Clairaut's Theorem $\mathbf{H}_f(\mathbf{x}_0)$ is a symmetric matrix.

Theorem

For any C^2 -function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, and $\mathbf{x}_0 \in D^\circ$ we have

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}_f(\mathbf{x}_0) (\mathbf{x} - \mathbf{x}_0) + o(|\mathbf{x} - \mathbf{x}_0|^2) \quad \text{for } \mathbf{x} \rightarrow \mathbf{x}_0.$$

Setting as usual $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$, this can also be written as

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T \mathbf{h} + \frac{1}{2} \mathbf{h}^T \mathbf{H}_f(\mathbf{x}_0) \mathbf{h} + o(|\mathbf{h}|^2) \text{ for } \mathbf{h} \rightarrow \mathbf{0}.$$

Proof.

We use Taylor approximation of the one-variable function $g(t) = f(\mathbf{x}_0 + t\mathbf{h})$.

Using the Chain Rule, g is a C^2 -function with derivatives

$$g'(t) = \nabla f(\mathbf{x}_0 + t\mathbf{h})^\top \mathbf{h} = \sum_{j=1}^n f_{x_j}(\mathbf{x}_0 + t\mathbf{h}) h_j,$$

$$\begin{aligned} g''(t) &= \sum_{j=1}^n h_j \nabla f_{x_j}(\mathbf{x}_0 + t\mathbf{h})^\top \mathbf{h} = \sum_{i,j=1}^n f_{x_i x_j}(\mathbf{x}_0 + t\mathbf{h}) h_i h_j \\ &= \mathbf{h}^\top \mathbf{H}_f(\mathbf{x}_0 + t\mathbf{h}) \mathbf{h} \end{aligned}$$

The 1st-order approximation with Lagrange remainder term gives

$$\begin{aligned} f(\mathbf{x}_0 + \mathbf{h}) &= g(1) = g(0) + g'(0) + \frac{1}{2}g''(\tau) \quad (\text{with } 0 < \tau < 1) \\ &= g(0) + g'(0) + \frac{1}{2}g''(0) + \frac{1}{2}(g''(\tau) - g''(0)) \\ &= f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top \mathbf{h} + \frac{1}{2}\mathbf{h}^\top \mathbf{H}_f(\mathbf{x}_0) \mathbf{h} + R(\mathbf{h}) \end{aligned}$$

with $R(\mathbf{h}) = \frac{1}{2}\mathbf{h}^\top \mathbf{H}_f(\mathbf{x}_0 + \tau\mathbf{h}) \mathbf{h} - \frac{1}{2}\mathbf{h}^\top \mathbf{H}_f(\mathbf{x}_0) \mathbf{h}$.

$\Rightarrow R(\mathbf{h})$ is a sum of terms $\frac{1}{2}h_i h_j (f_{x_i x_j}(\mathbf{x}_0 + \tau\mathbf{h}) - f_{x_i x_j}(\mathbf{x}_0)) = o(|\mathbf{h}|^2)$. □

The theorem reduces the question about whether a critical point \mathbf{x}_0 of f is a local maximum/minimum to the analysis of the quadratic form (*Hesse form*) $q(\mathbf{h}) = \mathbf{h}^\top \mathbf{A} \mathbf{h} = \sum_{i,j=1}^n a_{ij} h_i h_j$ with $\mathbf{A} = \mathbf{H}_f(\mathbf{x}_0)$, $a_{ij} = f_{x_i x_j}(\mathbf{x}_0)$.

Corollary

Let $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$ be a C^2 -function, $\mathbf{x}_0 \in D^\circ$ a critical point of f and $q(\mathbf{h}) = \mathbf{h}^\top \mathbf{H}_f(\mathbf{x}_0) \mathbf{h}$ the corresponding Hesse quadratic form.

- ① $f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \frac{1}{2} q(\mathbf{h}) + o(|\mathbf{h}|^2)$ for $\mathbf{h} \rightarrow \mathbf{0}$
- ② If $q(\mathbf{h}) > 0$ for all $\mathbf{h} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ then f has a strict local minimum at \mathbf{x}_0 .
- ③ If $q(\mathbf{h}) < 0$ for all $\mathbf{h} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ then f has a strict local maximum at \mathbf{x}_0 .
- ④ If there exist $\mathbf{h}_1, \mathbf{h}_2 \in \mathbb{R}^n$ such that $q(\mathbf{h}_1) > 0$ and $q(\mathbf{h}_2) < 0$ then f has no local extremum at \mathbf{x}_0 .

Definition

A quadratic form q with the property in Part (2), (3), (4) of the corollary is said to be *positive definite*, *negative definite*, and *indefinite*, respectively.

Proof of the corollary.

(1) is clear.

(2) By Part (1) and since $q\left(\frac{\mathbf{h}}{|\mathbf{h}|}\right) = q\left(\frac{1}{|\mathbf{h}|} \mathbf{h}\right) = \frac{1}{|\mathbf{h}|^2} q(\mathbf{h})$, we have

$$\frac{f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)}{|\mathbf{h}|^2} = \frac{1}{2} q\left(\frac{\mathbf{h}}{|\mathbf{h}|}\right) + o(1) \quad \text{for } \mathbf{h} \rightarrow \mathbf{0}.$$

Now we use the nontrivial fact that q attains a minimum m on the unit sphere $S^{n-1} = S_1(\mathbf{0})$ of \mathbb{R}^n . (This will be proved later.)

By assumption, we must have $m > 0$. Further, we can choose a ball $B_\epsilon(\mathbf{0})$ such that the $o(1)$ -term is bounded in absolute value by $m/4$ for all $\mathbf{h} \in B_\epsilon(\mathbf{0})$. For such \mathbf{h} we then have

$$\frac{f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)}{|\mathbf{h}|^2} \geq \frac{m}{2} - \frac{m}{4} = \frac{m}{4} > 0.$$

This shows that f has a strict local minimum in \mathbf{x}_0 .

(3) is proved in the same way as (2).

(4) By Part (1), f has along the line $\mathbf{x}_0 + \mathbb{R}\mathbf{h}_1$ a strict local minimum in \mathbf{x}_0 , and along the line $\mathbf{x}_0 + \mathbb{R}\mathbf{h}_2$ a strict local maximum in \mathbf{x}_0 . This shows that in any neighborhood of \mathbf{x}_0 there exist points $\mathbf{x}', \mathbf{x}'' \in D$ with $f(\mathbf{x}') < f(\mathbf{x}_0) < f(\mathbf{x}'')$. The assertion follows. □

Example (cont'd)

For $f(x, y) = x^2 + y^2$ we computed $f_x(x, y) = 2x$, $f_y(x, y) = 2y$.

$$\implies \mathbf{H}_f(x, y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix},$$

and the Hesse quadratic form is

$$q(h_1, h_2) = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}^\top \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = 2h_1^2 + 2h_2^2.$$

By the corollary, f has a strict local minimum in $(0, 0)$, a fact we already knew of course.

Note

In a way the preceding example is not really illuminating, since f itself is a quadratic form, $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x}$ with $\mathbf{A} = \mathbf{A}^\top$, and hence

$$\begin{aligned} f(\mathbf{x}_0 + \mathbf{h}) &= (\mathbf{x}_0 + \mathbf{h})^\top \mathbf{A} (\mathbf{x}_0 + \mathbf{h}) = \mathbf{x}_0^\top \mathbf{A} \mathbf{x}_0 + 2\mathbf{x}_0^\top \mathbf{A} \mathbf{h} + \mathbf{h}^\top \mathbf{A} \mathbf{h} \\ &= f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^\top \mathbf{h} + \mathbf{h}^\top \mathbf{A} \mathbf{h} \end{aligned}$$

This shows that for a quadratic form at every point the 2nd-order approximation is exact and the Hesse quadratic form is essentially f itself.

The Case $n = 2$

Denoting the critical point by (x_0, y_0) , we have

$$q(h_1, h_2) = (h_1 \ h_2) \begin{pmatrix} A & B \\ B & C \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = A h_1^2 + 2B h_1 h_2 + C h_2^2$$

with $A = f_{xx}(x_0, y_0)$, $B = f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0)$, $C = f_{yy}(x_0, y_0)$.

If $A \neq 0$ we can complete the square:

$$q(h_1, h_2) = A \left(h_1 + \frac{B}{A} h_2 \right)^2 + \frac{AC - B^2}{A} h_2^2 = A' h'_1^2 + B' h'_2^2$$

with $h'_1 = h_1 + (B/A)h_2$, $h'_2 = h_2$ (a linear coordinate change).

Clearly $(h'_1, h'_2) \mapsto A' h'_1^2 + B' h'_2^2$, and hence q , is positive definite/negative definite/indefinite iff the signs of (A', B') are $(+, +)$, $(-, -)$, and one of $(+, -)$, $(-, +)$, respectively. Hence the conditions in the corollary take the form (valid also for $A = 0$):

$$f \text{ has in } (x_0, y_0) \begin{cases} \text{a strict local minimum} & \text{if } A > 0 \wedge AC - B^2 > 0, \\ \text{a strict local maximum} & \text{if } A < 0 \wedge AC - B^2 > 0, \\ \text{no local extremum} & \text{if } AC - B^2 < 0. \end{cases}$$

The Case $n = 2$ Continued

Part (1) of the corollary for $n = 2$ says

$$f(x_0 + h_1, y_0 + h_2) = f(x_0, y_0) + \frac{1}{2} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}^\top \begin{pmatrix} A & B \\ B & C \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + o(h_1^2 + h_2^2)$$

for $(h_1, h_2) \rightarrow (0, 0)$.

Making the change of variables, this becomes

$$f(x_0 + h'_1 - \frac{B}{A} h'_2, y_0 + h'_2) = f(x_0, y_0) + \frac{1}{2} \left(A' h'_1^2 + B' h'_2^2 \right) + o(h'_1^2 + h'_2^2)$$

for $(h'_1, h'_2) \rightarrow (0, 0)$, where $A' = A = f_{xx}(x_0, y_0)$ and

$$B' = \frac{AC - B^2}{A} = \frac{f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}(x_0, y_0)^2}{f_{xx}(x_0, y_0)},$$

or, with $\mathbf{x} = (x_0, y_0)$, $\mathbf{v}_1 = \mathbf{e}_1 = (1, 0)$, $\mathbf{v}_2 = (-B/A, 1)$,

$$f(\mathbf{x} + h'_1 \mathbf{v}_1 + h'_2 \mathbf{v}_2) = f(\mathbf{x}) + \frac{1}{2} \left(A' h'_1^2 + B' h'_2^2 \right) + o(h'_1^2 + h'_2^2).$$

By a further (linear) change of variables, viz. $h'_1 = \sqrt{|A'|/2} h'_1$, $h'_2 = \sqrt{|B'|/2} h'_1$ (assuming $B' \neq 0$), this can even be simplified to

The Case $n = 2$ Continued

$$f(\mathbf{x} + h_1''\mathbf{w}_1 + h_2''\mathbf{w}_2) = f(\mathbf{x}) \pm h_1''^2 \pm h_2''^2 + o(h_1''^2 + h_2''^2)$$

for $(h_1'', h_2'') \rightarrow (0, 0)$.

Since $\mathbf{w}_1, \mathbf{w}_2$ (and likewise $\mathbf{v}_1, \mathbf{v}_2$) form a basis of \mathbb{R}^2 , this says that the graph G_f of f near (x_0, y_0, z_0) , $z_0 = f(x_0, y_0)$, looks like an “affinely distorted” copy of the quadric surface

$z = z_0 \pm (x - x_0)^2 \pm (y - y_0)^2$, and thus f near (x_0, y_0) exhibits the same extremal behavior as the corresponding quadratic form $\pm x^2 \pm y^2$ near $(0, 0)$. The undistorted approximation is

$$f(x, y) \approx f(x_0, y_0) + \frac{1}{2} [A(x - x_0)^2 + 2B(x - x_0)(y - y_0) + C(y - y_0)^2]$$

for $(x, y) \rightarrow (x_0, y_0)$, and accordingly the quadric surface

$z = z_0 + \frac{1}{2} [A(x - x_0)^2 + 2B(x - x_0)(y - y_0) + C(y - y_0)^2]$ is called *osculating quadric* of G_f at (x_0, y_0) .

Note

When, e.g., replacing $o(|\mathbf{h}|^2)$ by $o(|\mathbf{h}'|^2)$, we have tacitly used the fact that for a bijective linear map $L: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ (and likewise in the case $n > 2$) there exist constants $C_1, C_2 > 0$ such that $C_1 |\mathbf{h}| \leq |L(\mathbf{h})| \leq C_2 |\mathbf{h}|$ for all $\mathbf{h} \in \mathbb{R}^2$ (exercise).

Exercise

In the discussion of the case $n = 2$ of the corollary it was stated that the case $A = 0$ provides no exception. Prove this assertion.

Hint: In this case we must have $B \neq 0$, and by symmetry we can assume $C = 0$. Find a linear change of variables that transforms the quadratic form $2Bh_1 h_2$ into $h'_1^2 - h'_2^2$.

Exercise

- 1 Consider a linear map $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, $\mathbf{h} \mapsto \mathbf{A}\mathbf{h} = \mathbf{h}'$. Assuming that $\mathbf{A} \in \mathbb{R}^{2 \times 2}$ is invertible, show that there exist constants $C_1, C_2 > 0$ such that

$$C_1 |\mathbf{h}| \leq |\mathbf{A}\mathbf{h}| \leq C_2 |\mathbf{h}| \quad \text{for all } \mathbf{h} \in \mathbb{R}^2.$$

(This property holds, more generally, for invertible matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$.)

- 2 Using a) show that any function $g(\mathbf{h}) = g(h_1, h_2)$ that is $o(|\mathbf{h}|^2)$ for $\mathbf{h} \rightarrow \mathbf{0}$ is also $o(|\mathbf{h}'|^2)$, and conversely.

Definition

A critical point (x_0, y_0) for which

$$AC - B^2 = f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}(x_0, y_0)^2 < 0$$

(and hence f has no local extremum at this point) is called a *saddle point*.

Reason: Up to scaling, the graph G_f looks approximately like the parabola $z = \pm h'_1{}^2$ on the line $h'_2 = 0$ and like $z = \mp h'_2{}^2$ (with different sign) on the line $h'_1 = 0$; i.e., it looks like a “saddle”.

Related example

Consider the function $u(x, y) = x^3 - 3xy^2$, defined on \mathbb{R}^2 . The graph G_u is known as *Monkey's Saddle*; cf. picture on the next slide.

$$\nabla u(x, y) = (3x^2 - 3y^2, -6xy), \quad \mathbf{H}_u(x, y) = \begin{pmatrix} 6x & -6y \\ -6y & -6x \end{pmatrix}$$

The point $(0, 0)$ is critical, but since $\mathbf{H}_u(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, we cannot apply the preceding theory. It is easy to prove, however, that u has no local extremum in $(0, 0)$ (and hence no local extremum at all).

Brief
Introduction to
Optimization

Unconstrained
Optimization

Optimization
with Equality
Constraints

Appendix:
Proofs of
Important
Theorems of
1-Variable
Calculus

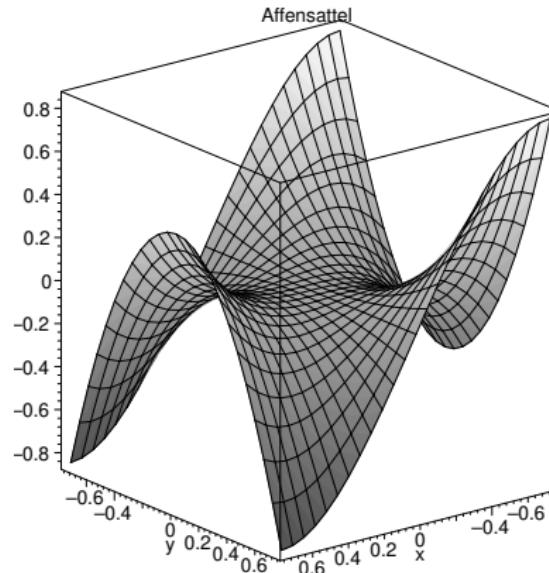


Figure: The graph of $u(x, y) = x^3 - 3xy^2$ near $(0, 0)$

Remark

Suppose \mathbf{x}_0 is a critical point of $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, and $\mathbf{A} = \mathbf{H}_f(\mathbf{x}_0)$ doesn't have full rank n . Then there exists $\mathbf{h} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that $q(\mathbf{h}) = \mathbf{h}^\top \mathbf{A} \mathbf{h} = 0$.

Then along the line $\mathbf{x}_0 + \mathbb{R}\mathbf{h}$ the function f may have a local maximum, minimum, or no extremum at all.

In particular the condition $q(\mathbf{h}) \geq 0$ for all $\mathbf{h} \in \mathbb{R}^n$ (q is *positive semi-definite*) does not imply that f has a possibly non-strict minimum at \mathbf{x}_0 (as one might think in the first place).

Problem 2 (actually a constrained case)

Determine the maxima and minima of the function

$f(x, y) = x^3 - xy + y^3$, which is defined for all $(x, y) \in \mathbb{R}^2$, on the closed unit square

$$S = \{(x, y) \in \mathbb{R}^2; 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

If f has a local maximum at an interior point (x, y) of S , i.e., $0 < |x| < 1 \wedge 0 < |y| < 1$, it must be critical point of f :

$$\nabla f(x, y) = (3x^2 - y, 3y^2 - x) = (0, 0).$$

This system of equations has two solutions, viz. $P_0 = (0, 0)$ and $P_1 = (\frac{1}{3}, \frac{1}{3})$ (to see this, substitute, e.g., $y = 3x^2$ into the second equation to obtain $27x^4 = x$).

Next we look at the corresponding Hesse matrices:

$$\mathbf{H}_f(x, y) = \begin{pmatrix} 6x & -1 \\ -1 & 6y \end{pmatrix}, \quad \mathbf{H}_f(0, 0) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{H}_f(\frac{1}{3}, \frac{1}{3}) = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

$\mathbf{H}_f(0, 0)$ is indefinite $\implies (0, 0)$ is a saddle point

$\mathbf{H}_f(\frac{1}{3}, \frac{1}{3})$ is positive definite $\implies (\frac{1}{3}, \frac{1}{3})$ is a strict local minimum with corresponding objective value $f(\frac{1}{3}, \frac{1}{3}) = -\frac{1}{27}$.

Problem 2 (cont'd)

To determine whether there are extrema on the boundary ∂S of S , we parametrize ∂S :

$$g_1(x) = f(x, 0) = x^3, \quad 0 \leq x \leq 1,$$

$$g_2(x) = f(x, 1) = x^3 - x + 1, \quad 0 \leq x \leq 1,$$

$$g_3(y) = f(0, y) = y^3, \quad 0 \leq y \leq 1,$$

$$g_4(y) = f(1, y) = y^3 - y + 1, \quad 0 \leq y \leq 1.$$

The usual Calculus I machinery shows that $g_2 = g_4$ has two maxima at 0 and 1 with $g_2(0) = g_2(1) = 1$ and a minimum at $\frac{1}{\sqrt{3}} \approx 0.577$ with value $g(\frac{1}{\sqrt{3}}) \approx 0.615$. Together with the obvious form of the graph of $g_1 = g_3$ this shows that f has 3 maxima on ∂S at $P_2 = (1, 0)$, $P_3 = (0, 1)$, $P_4 = (1, 1)$ with corresponding value $f(P_2) = f(P_3) = f(P_4) = 1$, and 3 minima in $P_0 = (0, 0)$, $P_5 = (1, \frac{1}{\sqrt{3}})$, $P_6 = (\frac{1}{\sqrt{3}}, 1)$ with values $f(P_0) = 0$, $f(P_5) = f(P_6) \approx 0.615$.

The maxima in P_2 , P_3 , P_4 and the minimum in P_0 are global relative to ∂S . The minima in P_5 , P_6 are local relative to ∂S .

Brief
Introduction to
Optimization

Unconstrained
Optimization

Optimization
with Equality
Constraints

Appendix:
Proofs of
Important
Theorems of
1-Variable
Calculus

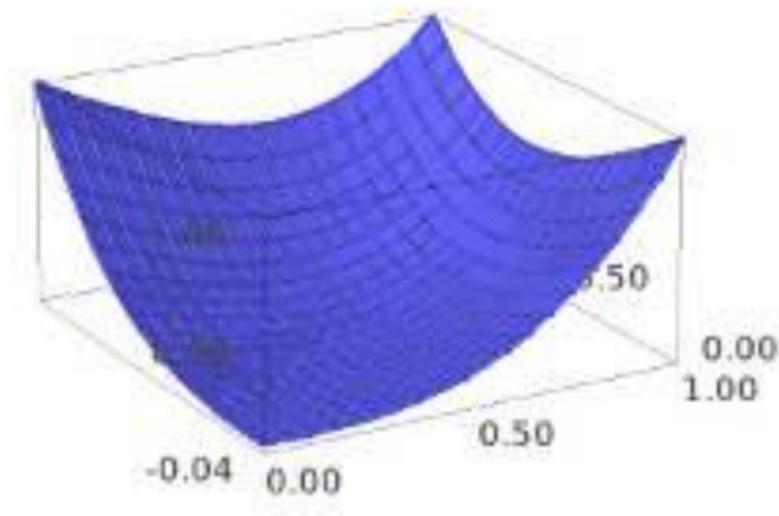


Figure: Graph of $f(x, y) = x^3 + y^3 - xy$, $(x, y) \in S$

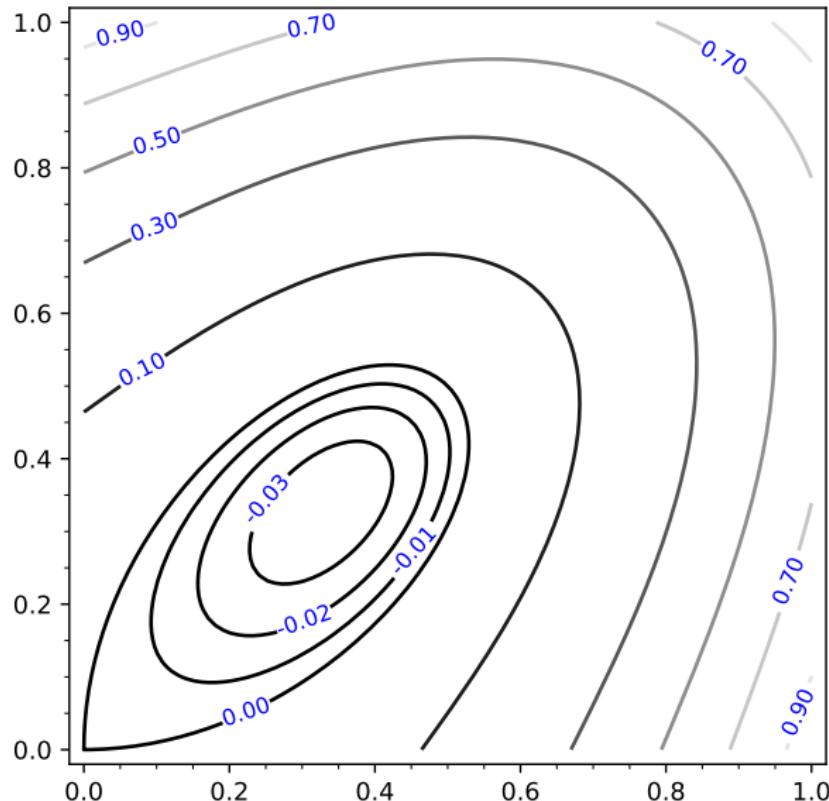


Figure: Contours of $f(x, y) = x^3 + y^3 - xy$, $(x, y) \in S$

Problem 2 (cont'd)

It appears from the pictures that f has on S a global minimum in P_1 and global maxima in P_2, P_3, P_4 .

Further it appears that the boundary points P_0, P_5, P_6 , when viewed relative to S , are no longer local minima.

Question

How can we prove these two facts rigorously?

Answers

- 1 The theorem (to be stated and proved shortly) that a continuous function on a closed and bounded subset of \mathbb{R}^n has a global maximum and a global minimum, applies to f and S . The extrema guaranteed by this theorem must have been found by the preceding analysis, and hence are obtained by comparing the values of f at P_0, \dots, P_6 .
- 2 P_0 is not a local minimum of f on S , since $f(x, x) = 2x^3 - x^2 = x^2(2x - 1) < 0$ for small positive x .
Caution: The saddle point property of P_0 can't be used for the proof, since P_0 is on the boundary ∂S .

Answers cont'd

2 (item cont'd)

To decide this question for P_5 (and, by symmetry, for P_6), we use the gradient:

$$\nabla f(P_5) = (3x^2 - y, 3y^2 - x) \Big|_{x=1, y=\frac{1}{\sqrt{3}}} = \left(3 - \frac{1}{\sqrt{3}}, 0\right)$$

This shows that along the line $y = \frac{1}{\sqrt{3}}$ the function f decreases/increases as we move from P_5 left/right and hence that $f(x, \frac{1}{\sqrt{3}}) < f(1, \frac{1}{\sqrt{3}}) = f(P_5)$ for x slightly smaller than 1.

Remark

The function f satisfies $f(x, y) = f(y, x)$ for all $(x, y) \in \mathbb{R}^2$ and hence its graph $G_f \subset \mathbb{R}^3$ is symmetric with respect to the plane $y = x$. This has been used Answer 2 above, and also explains why the functions parametrizing ∂S satisfy $g_1 = g_3$, $g_2 = g_4$. Moreover, it explains that the two critical points P_0, P_1 of f are on the line $y = x$. (For this consider the two parabolas $3x^2 - y = 0$, $3y^2 - x = 0$, which intersect in P_0 and P_1 .)

The Missing Link

Theorem

Suppose $S \subseteq \mathbb{R}^n$ is non-empty, closed and bounded, and $f: S \rightarrow \mathbb{R}$ is continuous. Then there exist $\mathbf{x}_1, \mathbf{x}_2 \in S$ such that $f(\mathbf{x}_1) \leq f(\mathbf{x}) \leq f(\mathbf{x}_2)$ for all $\mathbf{x} \in S$; i.e., f attains on S a maximum and a minimum.

Proof.

Let $M := \sup\{f(\mathbf{x}); \mathbf{x} \in S\}$ and choose a sequence $(\mathbf{x}^{(k)})$ of points in S with $\lim_{k \rightarrow \infty} f(\mathbf{x}^{(k)}) = M$. (If f is not bounded from above then $M = \lim_{k \rightarrow \infty} f(\mathbf{x}) = +\infty$.)

Since S is bounded, by the n -dimensional Bolzano-Weierstrass Theorem there exists a convergent subsequence $(\mathbf{x}^{(k_j)})$ of $(\mathbf{x}^{(k)})$ with $\lim_{j \rightarrow \infty} \mathbf{x}^{(k_j)} = \mathbf{a}$, say. Further, since S is closed, we must have $\mathbf{a} \in S$.

Clearly, $\lim_{j \rightarrow \infty} f(\mathbf{x}^{(k_j)}) = M$ as well. On the other hand, since f is continuous, $\lim_{j \rightarrow \infty} f(\mathbf{x}^{(k_j)}) = f(\lim_{j \rightarrow \infty} \mathbf{x}^{(k_j)}) = f(\mathbf{a})$.
 $\implies M = f(\mathbf{a})$, and in particular $M \in \mathbb{R}$ (so that f is bounded from above).

Setting $\mathbf{x}_2 = \mathbf{a}$, this proves the second half of the theorem. The first half is proved in the same way using $m = \inf\{f(\mathbf{x}); \mathbf{x} \in S\}$. □

The BOLZANO-WEIERSTRASS Theorem in Dimension n

Every bounded sequence $(\mathbf{x}^{(k)})$ in \mathbb{R}^n contains a convergent subsequence $\mathbf{x}^{(k_1)}, \mathbf{x}^{(k_2)}, \mathbf{x}^{(k_3)}, \dots$.

Recall that a sequence $(\mathbf{x}^{(k)})$ in \mathbb{R}^n is *bounded* if there exists $R > 0$ such that $|\mathbf{x}^{(k)}| < R$ for all $k \in \mathbb{N}$, and “subsequence” requires $k_1 < k_2 < k_3 < \dots$.

Sketch of proof.

Consider a closed cube $C_R(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^n; |x_i| \leq R \text{ for } 1 \leq i \leq n\}$ containing all sequence members and choose k_1 arbitrary. The cube, call it C_1 , can be divided into 2^n subcubes $C_{R/2}(\mathbf{a}_i)$, $1 \leq i \leq 2^n$, in the obvious way. (The coordinates of \mathbf{a}_i are chosen as $\pm R/2$ in all possible ways.) One subcube must contain infinitely many members of the sequence. Call it C_2 and choose $k_2 > k_1$ such that $\mathbf{x}^{(k_2)} \in C_2$. The process can be repeated (i.e., during the next step we divide C_2 into 2^n subcubes $C_{R/4}(\mathbf{b}_i)$, etc.) and yields a sequence of nested cubes $C_1 \supset C_2 \supset C_3 \supset \dots$ with side lengths $R/2^{j-1} \rightarrow 0$ and $k_1 < k_2 < k_3 < \dots$ such that $\mathbf{x}^{(k_j)} \in C_j$ for all $j \in \mathbb{N}$. Now the *completeness* of \mathbb{R} implies that $\bigcap_{j=1}^{\infty} C_j$ contains exactly one point \mathbf{x} . It is then immediate that $\lim_{j \rightarrow \infty} \mathbf{x}^{(k_j)} = \mathbf{x}$. □

The Completeness Property

Recall that the field \mathbb{R} is the unique (up to ordered field isomorphism) complete ordered field, where “complete” refers to the following property: *Every non-empty subset $S \subset \mathbb{R}$, which has an upper bound in \mathbb{R} , has a supremum (least upper bound).*

From this it is easy to conclude that a nested sequence of closed intervals $I_k = [a_k, b_k]$,

$a_1 \leq a_2 \leq \cdots \leq a_k \leq \cdots \leq b_k \leq \cdots \leq b_2 \leq b_1$, with $b_k - a_k \rightarrow 0$ for $k \rightarrow \infty$ contains exactly one point:

$$\bigcap_{k=1}^{\infty} [a_k, b_k] = \{s\} \quad \text{for some } s \in \mathbb{R}.$$

This is proved by setting $s = \sup\{a_k; k \in \mathbb{N}\}$ and showing that s has the required properties; cf. our earlier discussion.

This “nested intervals principle” generalizes easily to \mathbb{R}^n , if we replace the intervals by n -dimensional closed cubes

$$[\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R}^n; a_i \leq x_i \leq b_i \text{ for } 1 \leq i \leq n\}.$$

Application of the theorem

The Hesse quadratic form $q(\mathbf{h})$ is continuous, and the unit sphere $S^{n-1} = S_1(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^n; x_1^2 + \cdots + x_n^2 = 1\}$ is closed and bounded.

$\implies q$ attains a minimum (and a maximum) on S^{n-1} , as previously stated.

Problem 3

Determine the maxima and minima of $f(x, y) = x^3 - xy + y^3$ on the closed unit disk

$$S = B_1(0, 0) = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 1\}.$$

This time we are interested only in global extrema. If f has an extremum in the interior S° then it must be a critical point, and these have been already computed.

$P_0 = (0, 0) \in S^\circ$ is a saddle point and hence not an extremum.
 $P_1 = (\frac{1}{3}, \frac{1}{3})$ is not a global minimum, since $f(P_1) = -\frac{1}{27}$ and $f(-1, 0) = -1 < f(P_1)$.

⇒ The (global) maxima and minima of f lie on the unit circle

$$\partial S = S^1 = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 = 1\}.$$

They can be determined by parametrizing the unit circle as $(x, y) = (\cos t, \sin t)$, $t \in [0, 2\pi]$, and applying the Calculus I machinery to the resulting function

$$g(t) = \cos^3 t - \cos t \sin t + \sin^3 t, \quad t \in [0, 2\pi].$$

If you aren't scared by the plot on the next slide, you are invited to do this.

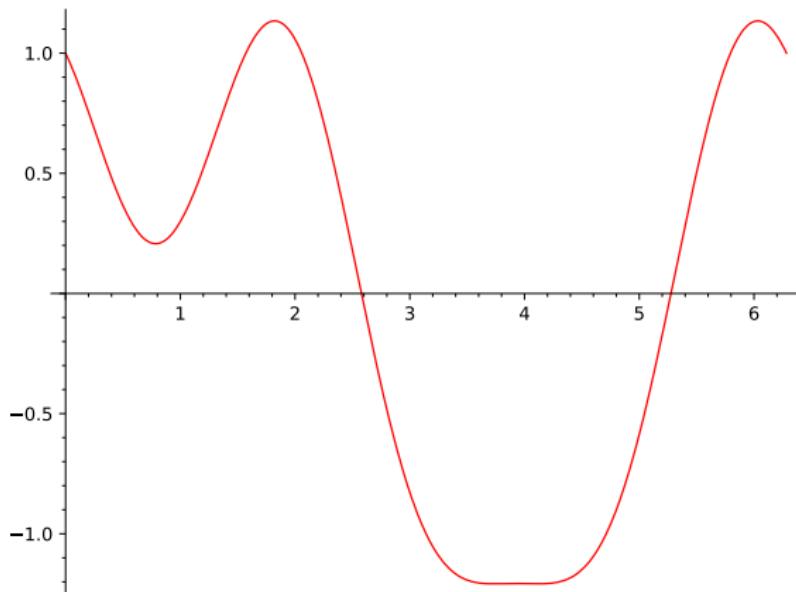


Figure: Graph of $g(t) = \cos^3 t - \cos t \sin t + \sin^3 t$, $t \in [0, 2\pi]$.

Problem 4 (an unconstrained case)

Determine the maxima and minima of $f(x, y) = (x + y)e^{-(x^2 - y^2)}$ on \mathbb{R}^2 .

Since \mathbb{R}^2 has no boundary, a putative maximum must be a critical point, which can be found as usual:

$$f_x = e^{-(x^2+y^2)} + (x + y)e^{-(x^2+y^2)}(-2x) = (1 - 2x^2 - 2xy)e^{-(x^2+y^2)},$$

$$f_y = (1 - 2y^2 - 2xy)e^{-(x^2+y^2)}.$$

The solutions of $f_x = f_y = 0$ are $P_1 = (\frac{1}{2}, \frac{1}{2})$, $P_2 = (-\frac{1}{2}, -\frac{1}{2})$, as is easily seen. The corresponding objective values are

$$f(P_1) = 1/\sqrt{e} \approx 0.607, \quad f(P_2) = -1/\sqrt{e} \approx -0.607.$$

Provided that f attains a global maximum and a global minimum, these must be P_1 , respectively, P_2 .

It remains to show that this is indeed the case.

The Cauchy-Schwarz Inequality gives

$$\begin{aligned}|x + y| &\leq |x| \cdot 1 + |y| \cdot 1 \leq \sqrt{x^2 + y^2} \sqrt{1^2 + 1^2} \\ \implies |f(x, y)| &\leq \sqrt{2}re^{-r^2}, \quad r = \sqrt{x^2 + y^2}.\end{aligned}$$

Problem 4 cont'd

(Or use $|x + y| \leq 2r$ to get the equally sufficient $|f(x, y)| \leq 2r e^{-r^2}$.)

Since $r e^{-r^2} \rightarrow 0$ for $r \rightarrow \infty$, we can conclude that

$$\lim_{|(x,y)| \rightarrow \infty} f(x, y) = 0.$$

This is the key property of f , which together with the continuity and the existence of at least one positive value and one negative value implies the existence of a global maximum and minimum:

Since $f(P_1) > 0$, there exists $R > 0$ such that $f(x, y) < f(P_1)$ for all (x, y) with $x^2 + y^2 > R^2$.

\Rightarrow The maximum of f on the disk $x^2 + y^2 \leq R^2$, which exists (since the disk is closed and bounded and f is continuous), is a global maximum.

For the minimum we argue in a similar way.

Brief
Introduction to
Optimization

Unconstrained
Optimization

Optimization
with Equality
Constraints

Appendix:
Proofs of
Important
Theorems of
1-Variable
Calculus

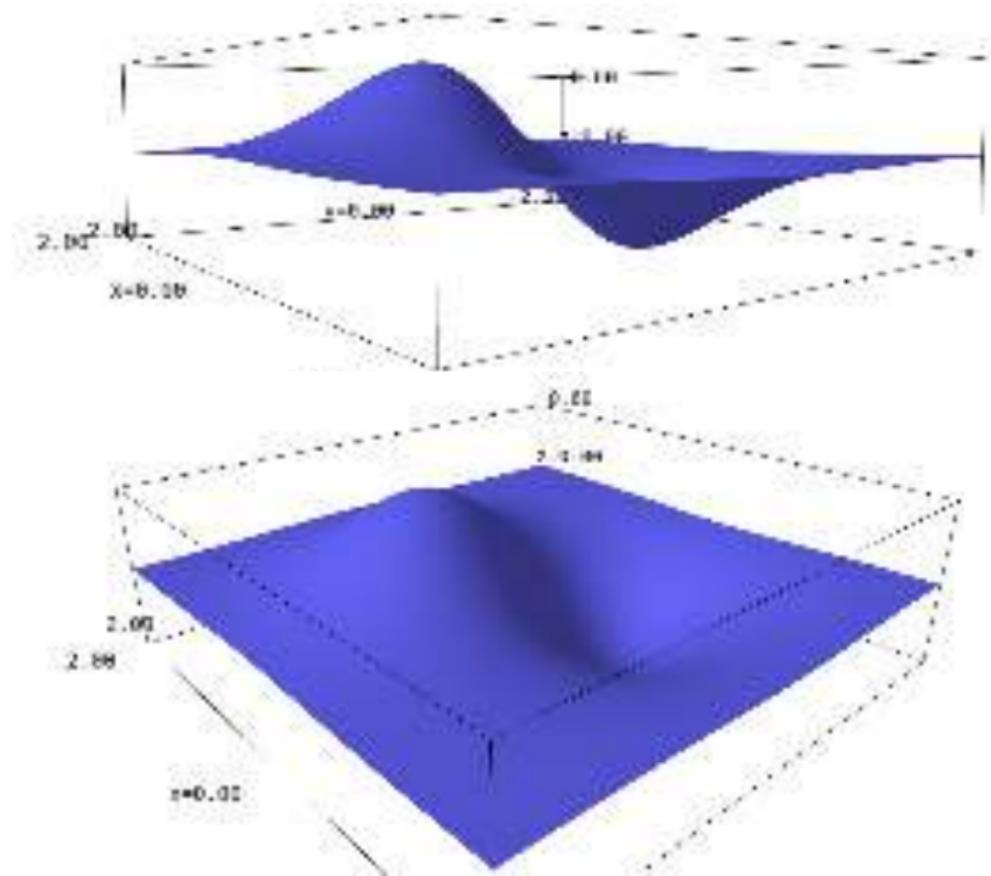


Figure: Graph of $f(x, y) = (x + y)e^{-x^2-y^2}$

Problem 5

A rectangular box without lid is to be made from 75 cm^2 of cardboard. Determine the maximum volume of such a box.

Here the task is to maximize the volume $f(x, y, z) = xyz$ subject to the constraint that the surface of the box, which equals

$$g(x, y, z) = xy + 2xz + 2yz,$$

is equal to 75. So we can take

$$\begin{aligned} D &= (\mathbb{R}_0^+)^3 = \{(x, y, z) \in \mathbb{R}^3; x \geq 0, y \geq 0, z \geq 0\}, \\ S &= \{(x, y, z) \in D; xy + 2xz + 2yz = 75\}. \end{aligned}$$

The constraint is equivalent to $xy + 2z(x + y) = 75$ and can be solved for z , provided that $xy \leq 75$.

$$\Rightarrow f(x, y, z) = xy \cdot \frac{75 - xy}{2(x + y)} = \frac{75xy - x^2y^2}{2x + 2y} \quad \text{on } S.$$

Thus it suffices to maximize the function $f_1(x, y) = \frac{75xy - x^2y^2}{2x + 2y}$ over $S_1 = \{(x, y) \in (\mathbb{R}_0^+)^2; xy \leq 75, (x, y) \neq (0, 0)\}$.

Problem 5 cont'd

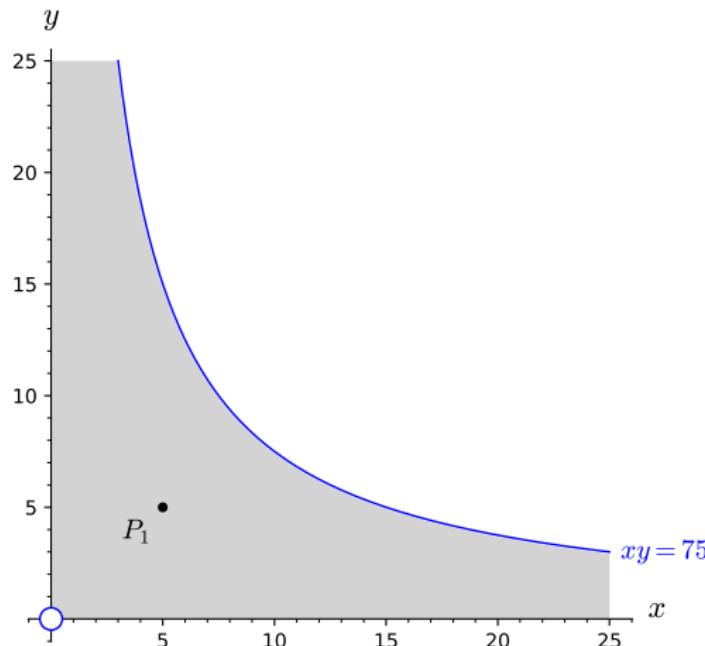


Figure: The domain S_1 together with the critical point P_1

Problem 5 cont'd

A straightforward computation gives

$$\nabla f_1(x, y) = \left(\frac{y^2(75 - x^2 - 2xy)}{2(x+y)^2}, \frac{x^2(75 - y^2 - 2xy)}{2(x+y)^2} \right).$$

The only critical point is $P_1 = (5, 5)$, which corresponds to $P = (5, 5, 2.5) \in S$, with corresponding volume 62.5 cm^3 .

Question

Does this solve Problem 5?

Well, it depends on whether you accept the (obvious?) fact that the volume of the box becomes small if one of x, y, z is large.

For a rigorous proof that f_1 attains its maximum in P_1 we need to show that

- ① f_1 attains a maximum on S_1 ;
- ② the maximum cannot be attained on the boundary ∂S_1 .

Both properties imply that f_1 attains its maximum at an interior point of S_1 , which must be critical and hence equal to P_1 .

Problem 5 cont'd

Property (2) holds, since $f_1(x, y) = 0$ for $(x, y) \in \partial S_1 \cap S_1$.

In order to prove Property (1), we show that the maximum of f_1 on a suitably chosen closed and bounded subset $C \subseteq S_1$, which exists in view of the continuity of f_1 , must be a maximum on S_1 as well. Specifically, we claim that we can take

$$C = \{(x, y) \in S_1; \frac{1}{75} \leq x \leq 75, \frac{1}{75} \leq y \leq 75\}.$$

Proof: We know that on C the function f_1 takes values as large as 62.5. If $x > 75$ then $y < 1$ (since $xy \leq 75$ on S_1) and

$$f_1(x, y) = \frac{x}{x+y} \frac{y(75 - xy)}{2} < 1 \cdot \frac{1}{2} \cdot 75 = 37.5.$$

If $x < \frac{1}{75}$ then

$$f_1(x, y) = \frac{y}{x+y} \frac{x(75 - xy)}{2} < 1 \cdot \frac{1}{2 \cdot 75} \cdot 75 = 0.5.$$

Together with the analogous estimates for $y \notin [\frac{1}{75}, 75]$ this proves our claim.

Another

Question

Can you prove without resort to multivariable calculus that the optimally shaped box has a square as base and height equal to one half of the side length of the base?

The Constrained Case

The cardboard example asks for the maximization of the function $f(x, y, z) = xyz$ under the constraint $xy + 2xz + 2yz = 75$, which we can put into the form $g(x, y, z) = xy + 2xz + 2yz - 75 = 0$.

More precisely, we set

$$D = (\mathbb{R}_0^+)^3 = \{(x, y, z) \in \mathbb{R}^3; x \geq 0 \wedge y \geq 0 \wedge z \geq 0\},$$

define $f, g: D \rightarrow \mathbb{R}$ as above, and the feasible region as

$$S = \{(x, y, z) \in D; g(x, y, z) = 0\}.$$

Then the cardboard problem is to find $\max\{f(x, y, z); (x, y, z) \in S\}$. This is called a maximum of f on S or *relative to S* .

Generalization

Find a maximum or minimum of $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, under a finite number of constraints $g_i(\mathbf{x}) = 0$, $1 \leq i \leq m$, where $g_i: D \rightarrow \mathbb{R}$.

We can encapsulate this into the single vectorial constraint $g(\mathbf{x}) = \mathbf{0}$, if we define g as $g = (g_1, \dots, g_m)$ (with domain D and codomain \mathbb{R}^m).

Problem 3 is also of this type: $\max_{\min} \{f(x, y); g(x, y) = 0\}$
with $f(x, y) = x^3 - xy + y^3$ and $g(x, y) = x^2 + y^2 - 1$.

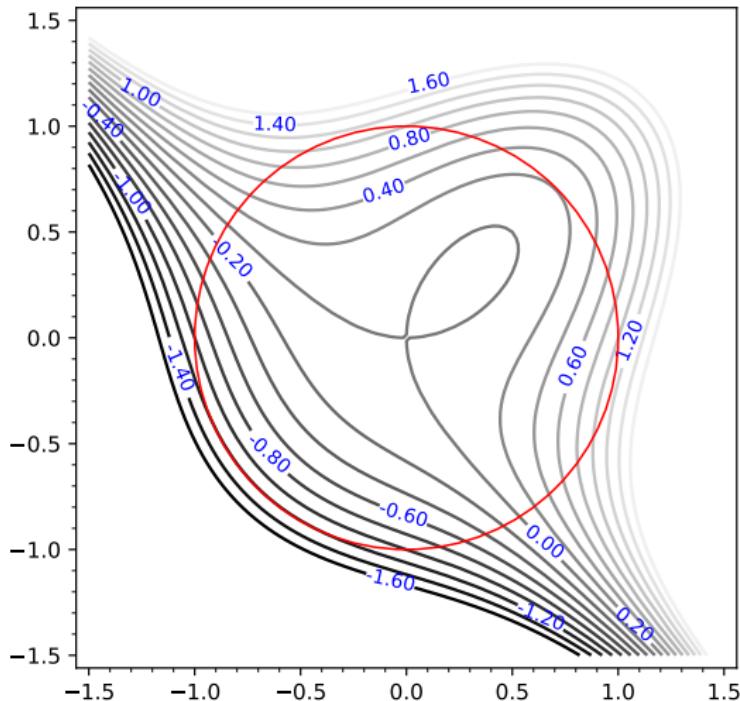


Figure: Contours of $f(x, y) = x^3 - xy + y^3$ near the unit circle

Lagrange Multipliers

Theorem

Let $f: D \rightarrow \mathbb{R}$ and $g = (g_1, \dots, g_m): D \rightarrow \mathbb{R}^m$ be C^1 -functions with common domain $D \subseteq \mathbb{R}^n$, $S = \{\mathbf{x} \in D; g(\mathbf{x}) = \mathbf{0}\}$, and $\mathbf{x}^* \in S$. Suppose that

- ① f has a local extremum relative to S in \mathbf{x}^* ;
- ② $\mathbf{J}_g(\mathbf{x}^*)$ has full row rank m .

Then there exist “multipliers” $\lambda_i \in \mathbb{R}$, $1 \leq i \leq m$, such that

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}^*). \quad (\text{L})$$

In other words, $\nabla f(\mathbf{x}^*)$ is contained in the row space of $\mathbf{J}_g(\mathbf{x}^*)$.

Note

(L) provides n equations for the $n + m$ unknowns x_1^*, \dots, x_n^* , $\lambda_1, \dots, \lambda_m$. Together with the m equations $g_i(\mathbf{x}^*) = 0$ we have $m + n$ equations for $m + n$ unknowns at our disposal.

Further Notes

- In the special case $m = 1$ the function g is scalar and (L) reduces to $\nabla f(\mathbf{x}^*) = \lambda \nabla g(\mathbf{x}^*)$ for some $\lambda \in \mathbb{R}$.

Moreover, in the case $n = 2$ the set S is a contour of g (in fact the 0-contour, but for other contours the theorem also holds, since the k -contour of g is the 0-contour of $\mathbf{x} \mapsto g(\mathbf{x}) - k$), and (L) says that the contours of f and g through \mathbf{x}^* touch in \mathbf{x}^* (i.e., have the same tangent line) if the contour of f is smooth in \mathbf{x}^* . The contour of g must be smooth in \mathbf{x}^* according to Condition (2).

For $n > 2$ the same remark applies mutatis mutandis.

Contours become general level sets, and “touching” means the tangent hyperplanes are equal.

- Be careful to include any points $\mathbf{x} \in S$ with $\nabla g(\mathbf{x}) = \mathbf{0}$ (or $\text{rk } \mathbf{J}_g(\mathbf{x}) < m$ in the general case) in your search for extrema of f relative to S . To such points the theorem does not apply, but nevertheless f may have an extremum relative to S there.
- Condition (2) says that S is smooth at \mathbf{x}^* . This follows from the Implicit Function Theorem, which asserts that under this condition S is locally at \mathbf{x}^* the graph of a differentiable function of some $n - m$ of the variables x_1, \dots, x_n .

For the proof we need two lemmas. The first lemma belongs to Linear Algebra and expresses a fundamental property of the dot product.

Definition

For a (linear) subspace U of \mathbb{R}^n we define

$$U^\perp = \{\mathbf{v} \in \mathbb{R}^n; \mathbf{u} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{u} \in U\}$$

and call U^\perp (which is easily seen to be subspace of \mathbb{R}^n as well) the *orthogonal space* (or *orthogonal complement*) of U .

Lemma

For all subspaces of \mathbb{R}^n the identity $(U^\perp)^\perp = U$ holds.

Proof.

The inclusion $U \subseteq (U^\perp)^\perp$ is trivial.

Suppose $\mathbf{x} \in \mathbb{R}^n$ satisfies $\mathbf{x} \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in U^\perp$. We must show that $\mathbf{x} \in U$.

We may write $\mathbf{x} = \mathbf{u} + \mathbf{v}$ with $\mathbf{u} \in U$, $\mathbf{v} \in U^\perp$; cf. our earlier discussion of orthogonal projections, which yields $\mathbf{u} \in U$ such that $\mathbf{x} - \mathbf{u} \perp U$.

Proof cont'd.

$$0 = \mathbf{x} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{v}.$$

$$\implies \mathbf{v} = \mathbf{0} \text{ and } \mathbf{x} = \mathbf{u} \in U.$$

□

The second Lemma relies on the Implicit Function Theorem and is much more difficult to prove.

Lemma

Suppose $g: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is differentiable at \mathbf{x}_0 and $\mathbf{J}_g(\mathbf{x}_0)$ has full rank m . Let M be the level set of g containing \mathbf{x}_0 , i.e., $M = N_g(g(\mathbf{x}_0)) = \{\mathbf{x} \in D; g(\mathbf{x}) = g(\mathbf{x}_0)\}$. Then every vector \mathbf{v} in the kernel of $\mathbf{J}_g(\mathbf{x}_0)$, i.e. $\mathbf{J}_g(\mathbf{x}_0)\mathbf{v} = \mathbf{0}$, is the tangent vector $\gamma'(0)$ of some curve $\gamma: (-\epsilon, \epsilon) \rightarrow M$ with $\gamma(0) = \mathbf{x}_0$.

Proof.

We may assume that the last m columns of $\mathbf{J}_g(\mathbf{x}_0)$ form an invertible submatrix. Then, by the Implicit Function Theorem, the equation $g(\mathbf{x}) = g(\mathbf{x}_0)$ can be solved locally for the last m variables, i.e., with $\mathbf{x}_0 = (\mathbf{a}, \mathbf{b})$ there exist neighborhoods U' of \mathbf{a} in \mathbb{R}^{n-m} , U'' of $\mathbf{b} \in \mathbb{R}^m$ and a C^1 -function $\phi: U' \rightarrow U''$ such that

$$M \cap (U' \times U'') = \{(\mathbf{x}', \phi(\mathbf{x}')); \mathbf{x}' \in U'\}.$$

Proof cont'd.

Now define for $1 \leq j \leq n - m$ the curves

$$\gamma_j(t) = (\mathbf{a} + t\mathbf{e}_j, \phi(\mathbf{a} + t\mathbf{e}_j)), \quad t \in (-\epsilon, \epsilon),$$

where $\epsilon > 0$ is sufficiently small.

The tangent vectors $\gamma'_j(0)$, which form the columns of the matrix $\begin{pmatrix} \mathbf{I}_{n-m} \\ \mathbf{J}_\phi(\mathbf{a}) \end{pmatrix}$, are linearly independent and hence span an $(n - m)$ -dimensional subspace of \mathbb{R}^n .

Since $\gamma_j(t) \in M$, we have $g(\gamma_j(t)) = g(\mathbf{x}_0)$, a fixed constant, and hence $d(g \circ \gamma_j) = 0$. Applying the chain rule then shows $\mathbf{J}_g(\mathbf{x}_0)\gamma'_j(0) = \mathbf{0}$.

Since the kernel of $\mathbf{J}_g(\mathbf{x}_0)$ has dimension $n - m$ as well, it must be equal to the span of $\gamma'_1(0), \dots, \gamma'_{n-m}(0)$.

Finally, if $\mathbf{J}_g(\mathbf{x}_0)\mathbf{v} = \mathbf{0}$, there exist $\lambda_j \in \mathbb{R}$ such

$$\mathbf{v} = \sum_{j=1}^{n-m} \lambda_j \gamma'_j(0) = \begin{pmatrix} \mathbf{I}_{n-m} \\ \mathbf{J}_\phi(\mathbf{a}) \end{pmatrix} \boldsymbol{\lambda}, \quad \text{where } \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{n-m}).$$

$$\implies \mathbf{v} = \gamma'(0) \text{ for } \gamma(t) = (\mathbf{a} + t\boldsymbol{\lambda}, \phi(\mathbf{a} + t\boldsymbol{\lambda})).$$



Proof of the theorem on Lagrange Multipliers.

(L) is equivalent to $\nabla f(\mathbf{x}^*) \in \langle \nabla g_1(\mathbf{x}^*), \dots, \nabla g_m(\mathbf{x}^*) \rangle$, i.e., to $\mathbb{R}\nabla f(\mathbf{x}^*) \subseteq \langle \nabla g_1(\mathbf{x}^*), \dots, \nabla g_m(\mathbf{x}^*) \rangle$.

Since for any two subspaces of \mathbb{R}^n the inclusion $U \subseteq V$ obviously implies $U^\perp \supseteq V^\perp$, we get from the first lemma the equivalence " $U \subseteq V \iff U^\perp \supseteq V^\perp$ ". Hence the condition above is equivalent to

$$\langle \nabla g_1(\mathbf{x}^*), \dots, \nabla g_m(\mathbf{x}^*) \rangle^\perp \subseteq (\mathbb{R}\nabla f(\mathbf{x}^*))^\perp.$$

The subspace on the left-hand side is precisely the kernel of $\mathbf{J}_g(\mathbf{x}^*)$, which by the second lemma consists of tangent vectors $\gamma'(0)$ of curves $\gamma: (-\epsilon, \epsilon) \rightarrow S$ with $\gamma(0) = \mathbf{x}^*$.

Since f restricted to S has a local extremum in \mathbf{x}^* , the same must be true of the image curve $f \circ \gamma$ with \mathbf{x}^* replaced by $0 \in \mathbb{R}$.

$$\implies (f \circ \gamma)'(0) = \nabla f(\mathbf{x}^*) \cdot \gamma'(0) = 0,$$

as desired. □

Note on the proof

In the proof we have only used that f is differentiable at \mathbf{x}^* , so that the condition on f may be relaxed. But g needs to be C^1 in a neighborhood of \mathbf{x}^* , because this is required for the Implicit Function Theorem.

Aside on Orthogonal Complements

We have stated one important property of orthogonal complements as a lemma and another one used in the proof of the Theorem on Lagrange Multipliers. But there is more to say, and we collect all important properties for easy reference. For any (linear) subspaces U, V of \mathbb{R}^n we have:

- ① $U + U^\perp = \mathbb{R}^n, \quad U \cap U^\perp = \{\mathbf{0}\};$
- ② $\dim(U) + \dim(U^\perp) = n;$
- ③ $U^{\perp\perp} = U;$
- ④ $U \subseteq V \iff U^\perp \supseteq V^\perp;$
- ⑤ $(U + V)^\perp = U^\perp \cap V^\perp, \quad (U \cap V)^\perp = U^\perp + V^\perp.$

The proofs of those which we have not yet done are easy.

As an application of (1) we mention the important fact that if B and C are bases of U and U^\perp , respectively, then $B \cup C$ is a basis of \mathbb{R}^n . Thus (1) and (2) justify the term “orthogonal complement”.

Example

We solve the cardboard problem

$$\max\{xyz; xy + 2xz + 2yz = 75 \wedge x, y, z \geq 0\}$$

a second time using Lagrange multipliers.

Here $f(x, y, z) = xyz$, $\nabla f(x, y, z) = (yz, xz, xy)$, and the equality constraint has the form $g(x, y, z) = xy + 2xz + 2yz - 75 = 0$ and

$$\nabla g(x, y, z) = (y + 2z, x + 2z, 2x + 2y) \neq \mathbf{0}$$

for all $(x, y, z) \in S$. The domain D can be taken w.l.o.g. as $(\mathbb{R}^+)^3$.

\implies The putative maximum must satisfy the system of equations

$$yz - \lambda(y + 2z) = 0$$

$$xz - \lambda(x + 2z) = 0$$

$$xy - \lambda(2x + 2y) = 0$$

$$xy + 2xz + 2yz - 75 = 0$$

Multiplying the 1st and 2nd equation by x resp. y and subtracting gives $2\lambda z(y - x) = 0$ and hence $x = y$ (since $\lambda = 0$ is impossible).

Example (cont'd)

The 3rd equation then yields $x = 4\lambda$, and the 2nd equation further $x = 2z$, so that the optimal solution has the form $(x, x, x/2)$ for some $x > 0$. From there the proof proceeds as before.

For convenience we provide a direct argument that $f(x, y, z) = xyz$ attains a maximum on $S = \{(x, y, z) \in (\mathbb{R}_0^+)^3; xy + 2xz + 2yz = 75\}$: On S the products xy , xz , yz are ≤ 75 . If one variable, say z , is > 150 then x, y are $< \frac{1}{2}$ and we obtain

$$f(x, y, z) = x(yz) \leq \frac{1}{2} 75 = 37.5.$$

This shows that the maximum of f on the closed and bounded set $\{(x, y, z) \in S; x, y, z \leq 150\}$, which exists by the continuity of f and is at least 62.5, is a global maximum.

Afternote

It is important to understand that we cannot take the domain D in the preceding example as $(\mathbb{R}_0^+)^3$, because the Lagrange Multiplier Theorem requires D to be open. Alternatively, the Lagrange Multiplier Theorem applies only to points in $S \cap D^\circ$, and any points in $S \cap \partial D$ need to be checked separately for the extremum property, just like those points $\mathbf{x} \in S \cap D^\circ$ for which $\mathbf{J}_g(\mathbf{x})$ doesn't have full row rank.

Example

We complete the determination of the extrema of $f(x, y) = x^3 + y^3 - xy$ on the closed unit disk. From the preceding considerations we know already that the extrema are located on the boundary S^1 , i.e., they satisfy the constraint $g(x, y) = x^2 + y^2 - 1 = 0$.

The Lagrange multiplier theorem gives the equations

$$3x^2 - y - \lambda x = 0$$

$$3y^2 - x - \lambda y = 0$$

$$x^2 + y^2 - 1 = 0$$

Multiplying the 1st equation by y , the 2nd by x and subtracting, we get $3x^2y - y^2 = 3xy^2 - x^2$, which simplifies to

$$(3xy + x + y)(x - y) = 0.$$

This is also the condition that the determinant of the 2×2 matrix formed from ∇f and ∇g is zero.

Clearly the solutions with $x = y$ of the 3rd equation (and hence all three equations) are $P_1 = \left(\frac{1}{2}\sqrt{2}, \frac{1}{2}\sqrt{2}\right)$, $P_2 = \left(-\frac{1}{2}\sqrt{2}, -\frac{1}{2}\sqrt{2}\right)$.

Example (cont'd)

Any remaining solution must satisfy $3xy + x + y = 0$, i.e., $y = -\frac{x}{3x+1}$, and $x \neq y$. Substituting this into the 3rd equation gives

$$9x^4 + 6x^3 - 7x^2 - 6x - 1 = 0.$$

This quartic equation can be solved by standard methods (but see below for an easy solution) and turns out to have 4 real roots

$$x_1 = \frac{1}{6} \left(-1 - \sqrt{5} - \sqrt{12 - 4\sqrt{5}} \right) \approx -0.831,$$

$$x_2 = \frac{1}{6} \left(-1 - \sqrt{5} + \sqrt{12 - 4\sqrt{5}} \right) \approx -0.248,$$

$$x_3 = \frac{1}{6} \left(-1 + \sqrt{5} - \sqrt{12 + 4\sqrt{5}} \right) \approx -0.557,$$

$$x_4 = \frac{1}{6} \left(-1 + \sqrt{5} + \sqrt{12 + 4\sqrt{5}} \right) \approx 0.969$$

One can check that $3x_1x_3 + x_1 + x_3 = 3x_2x_4 + x_2 + x_4 = 0$ (and also that $\sqrt{12 \pm 4\sqrt{5}} = \sqrt{10} \pm \sqrt{2}$). These equations, which must hold for two pairs of roots (since if (x, y) is a solution, so is (y, x)), make it easy to compute first $x_1 + x_3$, $x_2 + x_4$ and then x_1, x_2, x_3, x_4 .

Example (cont'd)

For this use

$$(x_1 + x_3) + (x_2 + x_4) = x_1 + x_2 + x_3 + x_4 = -6/9,$$

$$(x_1 + x_3)(x_2 + x_4) = (-3x_1x_3)(-3x_2x_4) = 9x_1x_2x_3x_4 = -1.$$

This gives 4 further candidate points, so that in all we have 6:

$$P_1 = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \approx (0.707, 0.707), \quad f(P_1) = \frac{-1+\sqrt{2}}{2} \approx 0.207,$$

$$P_2 = -\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \quad f(P_2) = \frac{-1-\sqrt{2}}{2} \approx -1.207,$$

$$P_3 = (x_1, x_3) \approx (-0.831, -0.557), \quad f(P_3) \approx -1.208,$$

$$P_4 = (x_3, x_1) \approx (-0.557, -0.831), \quad f(P_4) \approx -1.208,$$

$$P_5 = (x_2, x_4) \approx (-0.248, 0.969), \quad f(P_5) \approx 1.134,$$

$$P_6 = (x_4, x_2) \approx (0.969, -0.248), \quad f(P_6) \approx 1.134$$

From this we see (but you have to look carefully!) that P_3, P_4 are the global minima and P_5, P_6 are the global maxima of our function $f(x, y) = x^3 - xy + y^3$ on $S^1 = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 = 1\}$.

The extrema in the points P_1, P_2 are only local, as our earlier plot of $t \mapsto \cos^3 t - \cos t \sin t + \sin^3 t$ had suggested.

Example (cont'd)

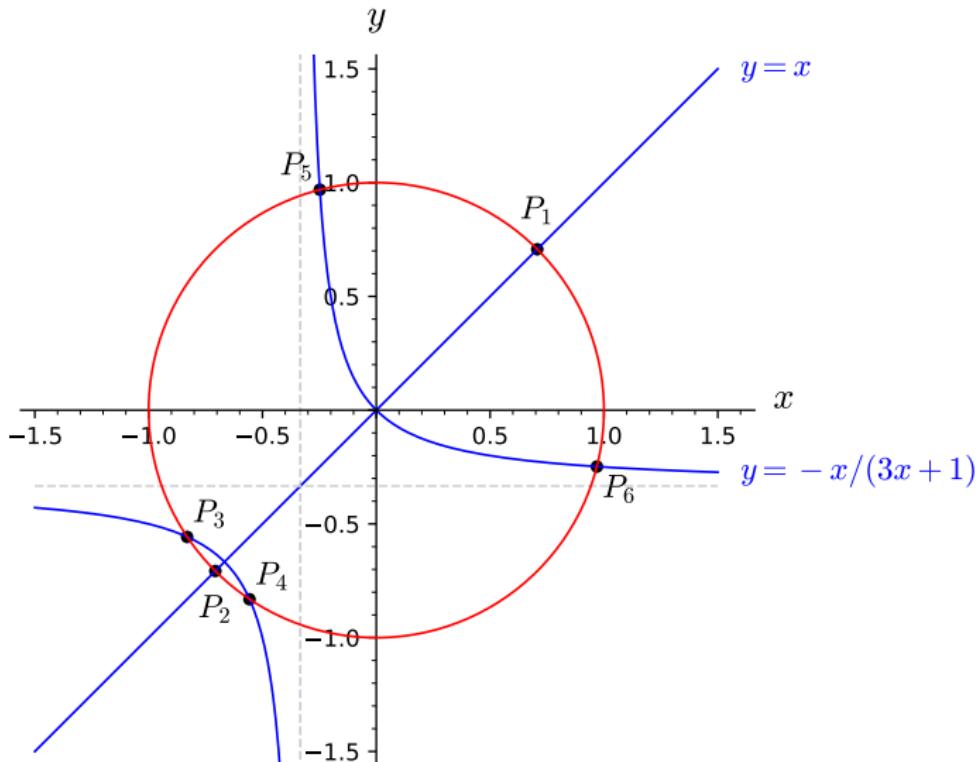


Figure: The geometry behind these computations

In this section proofs of the most important theorems of 1-variable calculus are collected. In [Ste21] these are mostly omitted, because they are ultimately based on the completeness property of \mathbb{R} (*every non-empty, bounded-from-above subset of \mathbb{R} has a supremum (least upper bound)*), whose discussion is avoided in [Ste21]. The mother of all these theorems is the Bolzano-Weierstrass Theorem.

Theorem (BOLZANO-WEIERSTRASS)

Every bounded sequence in \mathbb{R} has a convergent subsequence.

Since a proof has been given in the lecture slides on parametric curves (in the section on arc length), we only recall the basic idea involved: By assumption, the sequence (x_n) is contained in an interval $I_0 = [-R, R]$, $R > 0$. By successively halving the interval, we can prove the existence of a nested sequence

$I_0 \supset I_1 \supset I_2 \supset \dots$ of closed intervals $I_k = [a_k, b_k]$ of length 2^{-k} and a subsequence $(x_{n_0}, x_{n_1}, x_{n_2}, \dots)$ of (x_n) satisfying $x_{n_k} \in I_k$ for all $k \in \mathbb{N}$. It is then easy to see that the subsequence converges to $\sup\{a_k; k \in \mathbb{N}\}$. (Note that the set $\{a_k; k \in \mathbb{N}\}$ is bounded as well, and hence the supremum exists.)

Theorem (Intermediate Value Theorem)

A continuous function $f: [a, b] \rightarrow \mathbb{R}$ attains every real number between $f(a)$ and $f(b)$ as a value.

Proof.

W.l.o.g. we may assume $f(a) < f(b)$. Given $y \in [f(a), f(b)]$, let

$$c = \sup\{x \in [a, b]; f(x) \leq y\}.$$

The supremum exists, since $\{x \in [a, b]; f(x) \leq y\}$ contains at least one element, viz. a , and clearly $c \in [a, b]$. We claim that $f(c) = y$. Assume, by contradiction, that this is false.

If $f(c) < y$, we must have $c < b$, and by using the continuity of f in c we can find a response $\delta > 0$ (to the challenge $\epsilon = y - f(c)$) such that $c + \delta \leq b$ and $f(x) < y$ for all $x \in [c, c + \delta]$. This contradicts the definition of c .

Similarly, if $f(c) > y$, we can find $\delta > 0$ such that $c - \delta \geq a$ and $f(x) > y$ for all $x \in (c - \delta, c]$. But then we have $f(x) > y$ for all $x \in (c - \delta, b]$. Thus $c - \delta$ would be an upper bound for $\{x \in [a, b]; f(x) \leq y\}$, contradicting again the definition of c . □

Theorem (Extreme Value Theorem)

Suppose $f: [a, b] \rightarrow \mathbb{R}$ is continuous. Then there exist $c_1, c_2 \in [a, b]$ such that $f(c_1) \leq f(x) \leq f(c_2)$ for all $x \in [a, b]$.

In this situation we say that f has a *minimum (absolute minimum, global minimum)* in c_1 and a *maximum (absolute maximum, global maximum)* in c_2 .

Proof.

We show the existence of a maximum. (The existence of a minimum is proved, mutatis mutandis, in the same way.)

Let $M = \sup\{f(x); x \in [a, b]\}$, with the convention that $M = +\infty$ if $\{f(x); x \in [a, b]\}$ is not bounded from above. Then there exists a sequence (x_n) in $[a, b]$ such that $\lim_{n \rightarrow \infty} f(x_n) = M$. (In the case $M \in \mathbb{R}$ such a sequence is found by choosing, for every $n \in \mathbb{N}$, a point $x_n \in [a, b]$ such that $f(x_n) > M - 1/n$.)

By the Bolzano-Weierstrass Theorem, (x_n) has a convergent subsequence (x_{n_k}) . Setting $c_2 = \lim_{k \rightarrow \infty} x_{n_k}$, we have $c_2 \in [a, b]$ (since $[a, b]$ is closed) and $\lim_{k \rightarrow \infty} f(x_{n_k}) = M$ (since subsequences of convergent sequences have the same limit).

Further, $x_{n_k} \rightarrow c_2$ implies $f(x_{n_k}) \rightarrow f(c_2)$ for $k \rightarrow \infty$, since f is continuous in c_2 . Thus $M = f(c_2)$ (in particular $M \in \mathbb{R}$), and $f(c_2) \geq f(x)$ for $x \in [a, b]$. □

Recall that f is said to have a *local minimum* (or *relative minimum*) in x_0 if there exists $\delta > 0$ such that $f(x) \geq f(x_0)$ for all $x \in (x_0 - \delta, x_0 + \delta)$ for which $f(x)$ is defined, and similarly for *local (relative) maximum*.

Theorem

Suppose $f: (a, b) \rightarrow \mathbb{R}$ is differentiable and has a local extremum in x_0 . Then necessarily $f'(x_0) = 0$.

Proof.

Asssuming w.l.o.g. that the extremum is a minimum, we have

$$\frac{f(x) - f(x_0)}{x - x_0} \begin{cases} \geq 0 & \text{for } x \in (x_0, x_0 + \delta), \\ \leq 0 & \text{for } x \in (x_0 - \delta, x_0). \end{cases}$$

Hence, if $\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$ exists, it can only be zero. □

Theorem (ROLLE's Theorem)

Suppose $f: [a, b] \rightarrow \mathbb{R}$ is continuous. If f is differentiable in (a, b) and $f(a) = f(b) = 0$ there exists $x_0 \in (a, b)$ such that $f'(x_0) = 0$.

Proof.

Rolle's Theorem is an easy consequence of the preceding theorem and the Extreme Value Theorem: If $f(x) \equiv 0$, the assertion is trivially true. Otherwise either the minimum or the maximum of f , which exist by the Extreme Value Theorem, cannot be zero and hence is attained at a point $x_0 \in (a, b)$. The preceding theorem then yields $f'(x_0) = 0$. □

Theorem (Mean Value Theorem)

Suppose $f: [a, b] \rightarrow \mathbb{R}$ is continuous. If f is differentiable in (a, b) , there exists $\xi \in (a, b)$ such that

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

Proof.

Apply Rolle's Theorem to $g(x) := f(x) - f(a) - \frac{f(b)-f(a)}{b-a}(x - a)$, i.e., the difference of f and the linear function interpolating f at a, b . □

Theorem

Suppose f is defined on an interval I and $f'(x) = 0$ for all $x \in I$. Then f is constant.

Proof.

For $a, b \in I$, $a < b$, the Mean Value Theorem gives $f(b) - f(a) = f'(\xi)(b - a) = 0$.

The following theorem, also referred to as *First Mean Value Theorem of Integral Calculus*, is not directly related to the Ordinary Mean Value Theorem, as the name suggests. Rather it is a direct consequence of the Intermediate Value Theorem.

Theorem (Mean Value Theorem of Integral Calculus)

Suppose f, p are integrable over $[a, b]$, $p(x) \geq 0$ for $x \in [a, b]$, and f is continuous on $[a, b]$. Then there exists $c \in [a, b]$ such that

$$\int_a^b f(x)p(x) dx = f(c) \int_a^b p(x) dx.$$

Proof.

Denote by m, M the minimum/maximum of f on $[a, b]$. Since $p(x) \geq 0$, this implies $mp(x) \leq f(x)p(x) \leq Mp(x)$, and hence

$$m \int_a^b p(x) dx = \int_a^b mp(x) dx \leq \int_a^b f(x)p(x) dx \leq \int_a^b Mp(x) dx = M \int_a^b p(x) dx.$$

If $\int_a^b p(x) dx = 0$ then $\int_a^b f(x)p(x) dx = 0$, and the assertion holds trivially. Otherwise $\int_a^b p(x) dx > 0$, and we can divide the inequality by $\int_a^b p(x) dx$ to conclude that

$y = \int_a^b f(x)p(x) / \int_a^b p(x) dx \in [m, M]$, and apply the Intermediate Value Theorem to f and y . □

Theorem (Fundamental Theorem of Calculus)

If $f: [a, b] \rightarrow \mathbb{R}$ is continuous, the function $F: [a, b] \rightarrow \mathbb{R}$ defined by

$$F(x) = \int_a^x f(t) dt \quad \text{for } x \in [a, b]$$

is differentiable with $F'(x) = f(x)$ for all $x \in (a, b)$, and has one-sided derivatives $F'_+(a) = f(a)$, $F'_-(b) = f(b)$ in the end points a, b .

Proof.

For $x, x + h \in (a, b)$ we have

$$\frac{F(x+h) - F(x)}{h} - f(x) = \frac{1}{h} \int_x^{x+h} f(t) - f(x) dt.$$

If $\delta > 0$ is such that $|f(t) - f(x)| < \epsilon$ for $t \in (x - \delta, x + \delta)$ then for $|h| < \delta$ we have

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| \leq \frac{1}{|h|} |h| \epsilon = \epsilon.$$

This shows $\lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = f(x)$. The remaining assertions can be inferred from this, setting $x = a, h > 0$, respectively, $x = b, h < 0$. □

Notes

- FTC says in particular that a continuous function $f: [a, b] \rightarrow \mathbb{R}$ has an *antiderivative*, i.e., a function $F: [a, b] \rightarrow \mathbb{R}$ satisfying $F' = f$ (with one-sided derivatives meant in a, b). Any two such antiderivatives F_1, F_2 differ by a constant, since $(F_1 - F_2)' = F'_1 - F'_2 = f - f = 0$.
- FTC is often stated in the form $F(x) = F(a) + \int_a^x F'(t) dt$, reflecting the fact that F is the unique anti-derivative of F' having the value $F(a)$ at the point a .

Theorem (integration by parts)

Suppose $f: [a, b] \rightarrow \mathbb{R}$ is continuous and $g: [a, b] \rightarrow \mathbb{R}$ is continuously differentiable. Then, denoting by F any antiderivative of f , we have

$$\int_a^b f(x)g(x) dx = F(b)g(b) - F(a)g(a) - \int_a^b F(x)g'(x) dx.$$

Proof.

Applying FTC to $(Fg)(x) = F(x)g(x)$, we have

$$(Fg)(b) = (Fg)(a) + \int_a^b (Fg)'(x) dx = (Fg)(a) + \int_a^b (fg)(x) dx + \int_a^b (Fg')(x) dx.$$



Taylor Approximation

Suppose f is defined in a neighborhood of $a \in \mathbb{R}$ and the derivatives $f'(a), f''(a), \dots, f^{(n)}(a)$ exist. Recall that the polynomial (function)

$$\begin{aligned} T_n(x) &= T_n(x; f, a) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x - a)^k \\ &= f(a) + f'(a)(x - a) + \frac{f''(a)}{2}(x - a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x - a)^n \end{aligned}$$

is called *Taylor polynomial* of f of order n at the point a .

Theorem (Taylor's Formula)

Suppose f is of class C^{n+1} (i.e., the derivatives $f', f'', \dots, f^{(n+1)}$ exist, and $f^{(n+1)}$ is continuous). Then the remainder

$R_n(x) = f(x) - T_n(x)$ has the following properties:

$$① R_n(x) = \frac{1}{n!} \int_a^x (x - t)^n f^{(n+1)}(t) dt;$$

$$② R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} (x - a)^{n+1} \text{ for some } \xi \text{ between } a \text{ and } x.$$

Notes

- The expression for $R_n(x)$ in Part (1) is called *integral form of the remainder term*, while that in Part (2) is known as *Lagrange form of the remainder Term*.
- In the special case $n = 0$ Part (1) says that any C^1 -function satisfies $f(x) - f(a) = \int_a^x f'(t) dt$, and is thus equivalent to the Fundamental Theorem of Calculus.
- In the special case $n = 0$ Part (2) says that any C^1 -function satisfies $f(x) - f(a) = f'(\xi)(x - a)$ for some ξ between a and x , and is thus equivalent to the Mean Value Theorem (except that the Mean Value Theorems allows functions defined on $[a, b]$ to be non-differentiable at a, b).
- “Taylor’s Inequality”, stated on p. 798 of [Ste21], is an immediate consequence of Part (2).

Proof.

We do only the cases $n = 1$ and $n = 2$ of Part (1). The proof in the general case, which uses mathematical induction, can be inferred from this. The main tool is the “integration by parts” formula, which is a corollary to the FTC.



Proof cont'd.

$n = 1$: Using the already known case $n = 0$ and integration by parts, which requires the continuity of f'' , we obtain

$$\begin{aligned} f(x) - f(a) &= \int_a^x f'(t) dt = [(t-x)f'(t)]_a^x - \int_a^x (t-x)f''(t) dt \\ &= (x-a)f'(a) + \int_a^x (x-t)f''(t) dt. \end{aligned}$$

This is the formula for $n = 1$.

$n = 2$: Similarly, using the result for $n = 1$ and the continuity of f''' , we obtain

$$\begin{aligned} f(x) - T_1(x) &= \int_a^x (x-t)f''(t) dt \\ &= \left[-\frac{(x-t)^2}{2} f''(t) \right]_a^x + \int_a^x (x-t)^2 f'''(t) dt \\ &= \frac{(x-a)^2}{2} f''(a) + \int_a^x (x-t)^2 f'''(t) dt. \end{aligned}$$

Thus $f(x) - T_2(x) = f(x) - T_1(x) - \frac{(x-a)^2}{2} f''(a) = \int_a^x (x-t)^2 f'''(t) dt$, which is the formula for $n = 2$.

Proof cont'd.

Part (2) follows from Part (1) by applying the Mean Value Theorem of Integral Calculus to the interval $[a, x]$ (in place of $[a, b]$) and the functions $t \mapsto (x - t)^n$, $t \mapsto f^{(n+1)}(t)$ (in place of p , f), and noting that

$$\int_a^x (x - t)^n dt = \left[-\frac{(x - t)^{n+1}}{n+1} \right]_a^x = \frac{(x - a)^{n+1}}{n+1}.$$

Corollary (Qualitative Taylor Formula)

If f is of class C^n , we have

$$f(x) = T_n(x) + o((x - a)^n) \quad \text{for } x \rightarrow a.$$

Proof.

Applying Part (2) of the theorem with $n - 1$ in place of n and using the continuity of $f^{(n)}$ gives for $x \rightarrow a$ (and hence also $\xi \rightarrow a$)

$$\begin{aligned} f(x) &= T_{n-1}(x) + \frac{f^{(n)}(\xi)}{n!}(x - a)^n = T_{n-1}(x) + \frac{f^{(n)}(a)(1 + o(1))}{n!}(x - a)^n \\ &= T_{n-1}(x) + \frac{f^{(n)}(a)}{n!}(x - a)^n + o(1)(x - a)^n = T_n(x) + o((x - a)^n). \end{aligned}$$

Math 241
Calculus III

Thomas
Honold

Sylvester's
Inertia
Theorem

Quadrics and
their
Classification

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Sylvester's Inertia Theorem

2 Quadrics and their Classification

Today's Lecture: Quadratic Forms and Quadrics

Positive Definite Quadratic Forms and Sylvester's Inertia Theorem

Recall the following

Definition

- ① A quadratic form $q(\mathbf{x}) = q(x_1, \dots, x_n) = \sum_{i \leq j} q_{ij}x_i x_j$ with coefficients $q_{ij} \in \mathbb{R}$ is *positive definite* if $q(\mathbf{x}) > 0$ for all nonzero vectors $\mathbf{x} \in \mathbb{R}^n$.
- ② A symmetric matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n \times n}$ is *positive definite* if the associated quadratic form $q_{\mathbf{A}}(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i,j=1}^n a_{ij} x_i x_j$ has this property; equivalently $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for all nonzero vectors $\mathbf{x} \in \mathbb{R}^n$.

Question

How to decide algorithmically whether a given matrix/quadratic form is positive definite?

The answer is provided by the proof of Sylvester's Theorem. It uses—guess what?—Gaussian Elimination.

Coordinate Changes

If $\mathbf{b}_1, \dots, \mathbf{b}_n$ form a basis of \mathbb{R}^n , we can express every vector $\mathbf{x} \in \mathbb{R}^n$ uniquely as a linear combination

$$\mathbf{x} = x'_1 \mathbf{b}_1 + \cdots + x'_n \mathbf{b}_n = \mathbf{S}\mathbf{x}' \quad \text{with } \mathbf{S} = (\mathbf{b}_1 | \dots | \mathbf{b}_n).$$

The (linear) coordinate change $\mathbf{x} = \mathbf{S}\mathbf{x}'$ changes $q_A(\mathbf{x})$ into

$$q_A(\mathbf{S}\mathbf{x}') = (\mathbf{S}\mathbf{x}')^\top \mathbf{A} \mathbf{S}\mathbf{x}' = \mathbf{x}'^\top \mathbf{S}^\top \mathbf{A} \mathbf{S} \mathbf{x}' = q_{\mathbf{S}^\top \mathbf{A} \mathbf{S}}(\mathbf{x}').$$

Definition

Tow quadratic forms q_1 and q_2 on \mathbb{R}^n are said to be *equivalent* (notation $q_1 \sim q_2$) if there exists an invertible matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that $q_2(\mathbf{x}) = q_1(\mathbf{S}\mathbf{x})$.

This is indeed an equivalence relation (easy exercise) and corresponds to the equivalence relation

$$\mathbf{A} \sim \mathbf{B} : \iff \exists \mathbf{S} \in \mathbb{R}^{n \times n} \text{ such that } \text{rk}(\mathbf{S}) = n \text{ and } \mathbf{B} = \mathbf{S}^\top \mathbf{A} \mathbf{S}$$

on $\mathbb{R}^{n \times n}$.

Equivalent quadratic forms take the same values (on possibly different vectors). Hence all quadratic forms (or matrices) in a fixed equivalence class are positive definite, or none of them.

Theorem (SYLVESTER's Inertia Theorem)

For every quadratic form q on \mathbb{R}^n there are unique integers $r, s, t \geq 0$ with $r + s + t = n$ such that

$$q(x_1, \dots, x_n) \sim x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots - x_{r+s}^2. \quad (\text{S})$$

Note that $t = n - r - s$, and $t \geq 0$ is equivalent to $r + s \leq n$.

Observations

- 1 The matrix representing the quadratic form on the right-hand side of (S) (called *canonical form* of q) is a diagonal matrix of the special form

$$\begin{pmatrix} \mathbf{I}_r & & \\ & -\mathbf{I}_s & \\ & & \mathbf{0}_t \end{pmatrix}.$$

- 2 From the canonical form, say $q_{r,s}(x_1, \dots, x_n) = \sum_{i=1}^r x_i^2 - \sum_{i=r+1}^{r+s} x_i^2$, we can easily decide whether q is positive definite: Clearly $q_{r,s}$ is positive definite iff it is a sum of squares and all variables are involved, i.e., iff $r = n$ and $s = t = 0$. Since $q \sim q_{r,s}$, this criterion applies to q as well.

Observations cont'd

- ③ Using the form $q(\mathbf{x}) = q(\mathbf{Sx}') = \sum_{i=1}^r x_i'^2 - \sum_{i=r+1}^{r+s} x_i'^2$ and inverting, viz. $\mathbf{x}' = \mathbf{S}^{-1}\mathbf{x}$ or $x'_i = \sum_{j=1}^n t_{ij}x_j$ with $\mathbf{S}^{-1} = (t_{ij})$, we see that an “equivalent” formulation of Sylvester’s Theorem is the following: Every quadratic form on \mathbb{R}^n can be expressed as a sum of squares of linearly independent linear forms in the variables x_1, \dots, x_n with coefficients ± 1 .

Idea of the proof

We prove the matrix-equivalent of Sylvester's Theorem (more precisely, its existence part): There exists an invertible matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ such that $\mathbf{S}^T \mathbf{A} \mathbf{S}$ has the form $\begin{pmatrix} \mathbf{I}_r & & \\ & -\mathbf{I}_s & \\ & & \mathbf{0}_t \end{pmatrix}$.

In the special case where $\mathbf{S} = \mathbf{E}$ is an elementary matrix, the transformation $\mathbf{A} \mapsto \mathbf{E}^T \mathbf{A} \mathbf{E}$ corresponds to applying the elementary operation corresponding to \mathbf{E} both to the rows and columns of \mathbf{A} .

In the general case, since every invertible matrix \mathbf{S} is a product of elementary matrices, we can obtain the transformation $\mathbf{A} \mapsto \mathbf{S}^T \mathbf{A} \mathbf{S}$ as a sequence of such elementary row+column operations.

A suitable variant of the Gaussian Elimination algorithm can be used to transform, by means of row+column operations, every symmetric matrix into a diagonal matrix with entries $\pm 1, 0$ on the main diagonal.

Finally, elementary row+column operations of Type I ("interchange two rows and the corresponding columns") can be used to regroup the diagonal entries into the desired order. (Such an operation just interchanges the corresponding diagonal entries.)

Example (the 2×2 case)

Earlier we had transformed $q(h_1, h_2) = Ah_1^2 + 2Bh_1h_2 + Ch_2^2$ under the assumption $A \neq 0$ into $Ah_1'^2 + \frac{AC-B^2}{A}h_2'^2$ by means of the variable change $h'_1 \equiv h_1 + \frac{B}{A}h_2$, $h'_2 \equiv h_2$. This says in particular that

$$Ax_1^2 + 2Bx_1x_2 + Cx_2^2 \sim Ax_1^2 + \frac{AC - B^2}{A}x_2^2.$$

Now we are going to reproduce this result by applying suitable row+column operations to the coefficient matrix $\mathbf{A} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}$.

$$\begin{pmatrix} A & B \\ B & C \end{pmatrix} \xrightarrow{R2=R2-\frac{B}{A}R1} \begin{pmatrix} A & B \\ 0 & C - \frac{B^2}{A} \end{pmatrix} \xrightarrow{C2=C2-\frac{B}{A}C1} \begin{pmatrix} A & 0 \\ 0 & C - \frac{B^2}{A} \end{pmatrix}$$

This is the same as above in matrix terms.

In the case where \mathbf{A} is positive definite ($A > 0 \wedge AC - B^2 > 0$), the algorithm would then continue as

$$\xrightarrow{\begin{array}{l} R1=\frac{1}{\sqrt{A}}R1 \\ R2=\frac{1}{\sqrt{C-\frac{B^2}{A}}}R2 \end{array}} \begin{pmatrix} \sqrt{A} & 0 \\ 0 & \sqrt{C - \frac{B^2}{A}} \end{pmatrix} \xrightarrow{\begin{array}{l} C1=\frac{1}{\sqrt{A}}C1 \\ C2=\frac{1}{\sqrt{C-\frac{B^2}{A}}}C2 \end{array}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Example (cont'd)

We can keep track of the variable changes by performing the corresponding column operations (but not the row operations!) on the identity matrix.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{C_2 = C_2 - \frac{B}{A} C_1} \begin{pmatrix} 1 & -\frac{B}{A} \\ 0 & 1 \end{pmatrix} \xrightarrow{\begin{array}{l} C_1 = \frac{1}{\sqrt{A}} C_1 \\ C_2 = \frac{1}{\sqrt{C - \frac{B^2}{A}}} C_2 \end{array}} \begin{pmatrix} \frac{1}{\sqrt{A}} & -\frac{B}{A\sqrt{C - \frac{B^2}{A}}} \\ 0 & \frac{1}{\sqrt{C - \frac{B^2}{A}}} \end{pmatrix}$$

The last matrix is the change-of-variables matrix $\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} \\ 0 & s_{22} \end{pmatrix}$, i.e., $q(\mathbf{S}\mathbf{x}) = A(s_{11}x_1 + s_{12}x_2)^2 + 2B(s_{11}x_1 + s_{12}x_2)(s_{22}x_2) + (s_{22}x_2)^2 = x_1^2 + x_2^2$. The inverse change-of-variables matrix \mathbf{S}^{-1} is

$$\begin{pmatrix} \sqrt{A} & 0 \\ 0 & \sqrt{C - \frac{B^2}{A}} \end{pmatrix} \begin{pmatrix} 1 & \frac{B}{A} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \sqrt{A} & \frac{B}{\sqrt{A}} \\ 0 & \sqrt{C - \frac{B^2}{A}} \end{pmatrix}$$

and affords the transformation $x_1^2 + x_2^2 \mapsto q(x_1, x_2)$. Indeed, we have

$$Ax_1^2 + 2Bx_1x_2 + Cx_2^2 = \left(\sqrt{A}x_1 + \frac{B}{\sqrt{A}}x_2 \right)^2 + \left(\sqrt{C - \frac{B^2}{A}}x_2 \right)^2.$$

Proof of Sylvester's Theorem.

The main step is to prove that every symmetric matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n \times n}$ can be transformed into a diagonal matrix by a sequence of elementary row-column transformations. Scaling of the nonzero diagonal entries to ± 1 can be done as in the 2×2 case.

Case 1: $a_{11} \neq 0$

Here we use elementary row operations with a_{11} as pivot to turn \mathbf{A} into a matrix $\mathbf{A}' = (a'_{ij})$ with $a'_{21} = a'_{31} = \cdots = a'_{n1} = 0$.

The corresponding column operations, which may be performed afterwards, do not destroy these zeros and turn \mathbf{A}' into a matrix $\mathbf{A}'' = (a''_{ij})$ with $a''_{12} = a''_{13} = \cdots = a''_{1n} = 0$.

\Rightarrow The matrix \mathbf{A}'' has the form

$$\mathbf{A}'' = \begin{pmatrix} a_{11} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \quad \text{with} \quad \mathbf{B} \in \mathbb{R}^{(n-1) \times (n-1)}, \quad \mathbf{B} = \mathbf{B}^T.$$

The proof can then be finished by induction on n .

Case 2: $a_{11} = 0, a_{ii} \neq 0$ for some $2 \leq i \leq n$

This case is reduced to Case 1 by interchanging Row 1 with Row i and Column 1 with Column i .

Proof cont'd.

Case 3: $a_{ii} = 0$ for $1 \leq i \leq n$, $a_{i1} \neq 0$ for some $2 \leq i \leq n$

In this case we can add Row i to Row 1, which doesn't change a_{1i} , and then Column i to Column 1. The new matrix \mathbf{A}'' has $a''_{11} = 2a_{11} \neq 0$, so that we are in Case 1.

Case 4: $a_{i1} = a_{1i} = 0$ for $1 \leq i \leq n$

In this case \mathbf{A} has the form

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \quad \text{with} \quad \mathbf{B} \in \mathbb{R}^{(n-1) \times (n-1)}, \quad \mathbf{B} = \mathbf{B}^T,$$

and the proof can be finished by induction on n as in Case 1.

This completes the proof of the existence part of Sylvester's Theorem. It remains to prove the uniqueness of r, s, t .

Clearly, the rank of the resulting diagonal matrix, and hence the rank of \mathbf{A} , is equal to $r + s$. Thus $t = n - \text{rk}(\mathbf{A})$ is uniquely determined by \mathbf{A} .

Proof cont'd.

In order to determine r (and hence s) we can argue as follows:

The subspace $U = \langle \mathbf{e}_1, \dots, \mathbf{e}_r \rangle$ has dimension r and

$q_{r,s}(\mathbf{x}) = x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots - x_{r+s}^2$ is positive definite on U .

If V is a subspace of \mathbb{R}^n with $\dim(V) > r$ then V must intersect $\langle \mathbf{e}_{r+1}, \dots, \mathbf{e}_n \rangle$ nontrivially and hence contain a nonzero vector \mathbf{v}

such that $q_{r,s}(\mathbf{x}) \leq 0$. It follows that r is equal to the largest dimension of a subspace of \mathbb{R}^n on which $q_{r,s}$ is positive definite.

But clearly this quantity is invariant under a change-of-variables, so that r is also the largest dimension of a subspace of \mathbb{R}^n on which $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ is positive definite. Thus r is uniquely determined by \mathbf{A} . □

Illustration of the First Three Cases

In each case it is assumed that $a \neq 0$.

$$\textcircled{1} \quad \begin{pmatrix} a & * & * \\ * & * & * \\ * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} a & * & * \\ 0 & * & * \\ 0 & * & * \end{pmatrix} \rightarrow \begin{pmatrix} a & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{pmatrix}$$

$$\textcircled{2} \quad \begin{pmatrix} 0 & * & * \\ * & * & * \\ * & * & a \end{pmatrix} \rightarrow \begin{pmatrix} * & * & a \\ * & * & * \\ 0 & * & * \end{pmatrix} \rightarrow \begin{pmatrix} a & * & * \\ * & * & * \\ * & * & 0 \end{pmatrix}$$

$$\textcircled{3} \quad \begin{pmatrix} 0 & a & * \\ a & 0 & * \\ * & * & 0 \end{pmatrix} \rightarrow \begin{pmatrix} a & a & * \\ a & 0 & * \\ * & * & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 2a & a & * \\ a & 0 & * \\ * & * & 0 \end{pmatrix}$$

Note that, in order to continue, the transformed matrices should again be symmetric. Here is the proof in the general case:

$$\mathbf{A} = \mathbf{A}^T \implies \mathbf{B}^T = (\mathbf{S}^T \mathbf{A} \mathbf{S})^T = \mathbf{S}^T \mathbf{A}^T \mathbf{S}^T = \mathbf{S}^T \mathbf{A} \mathbf{S} = \mathbf{B}.$$

Quadrics and their Classification

Definition

- 1 An (affine) *quadric* Q in \mathbb{R}^n is the solution set of an equation $f(\mathbf{x}) = f(x_1, \dots, x_n) = 0$, where f is a polynomial of degree 2 with coefficients in \mathbb{R} .

Examples

- 1 All affine subspaces of \mathbb{R}^n except \mathbb{R}^n itself are quadrics. The hyperplane with equation $a_1x_1 + \dots + a_nx_n = b$ is the solution set of $(a_1x_1 + \dots + a_nx_n - b)^2 = 0$. A general affine subspace is the solution of a system of linear equations, and we can take f as the corresponding sum of squares.
- 2 In \mathbb{R}^3 unions of two distinct planes are quadrics, since they are solution sets of quadratic equations of the form $(a_1x_1 + a_2x_2 + a_3x_3 - b)(\alpha_1x_1 + \alpha_2x_2 + \alpha_3x_3 - \beta) = 0$.
- 3 Conic sections in \mathbb{R}^2 define “cylindrical” quadrics in \mathbb{R}^3 (with the z -direction $(0, 0, 1)$ as axis) by interpreting the corresponding polynomial $f(x, y)$ as $f(x, y, z)$.

All these examples are degenerate in some sense. The following is more like what we want.

Example (cont'd)

In \mathbb{R}^3 consider the quadratic equations

$$x^2 + y^2 + z^2 = 1, \tag{1}$$

$$x^2 + y^2 - z^2 = 1, \tag{2}$$

$$x^2 - y^2 - z^2 = 1. \tag{3}$$

The corresponding quadrics are

- ① the *unit sphere*,
- ② a *hyperboloid of one sheet*,
obtained by rotating the hyperbola $x^2 - z^2 = 1$ around the z -axis; the horizontal traces $z = k$ are circles, viz.
 $x^2 + y^2 = 1 + k^2$, for all $k \in \mathbb{R}$);
- ③ a *hyperboloid of two sheets*,
obtained by rotating the hyperbola $x^2 - z^2 = 1$ around the x -axis; the (vertical) traces $x = k$ are circles for $|k| > 1$, viz.
 $y^2 + z^2 = k^2 - 1$, and empty for $|k| < 1$.

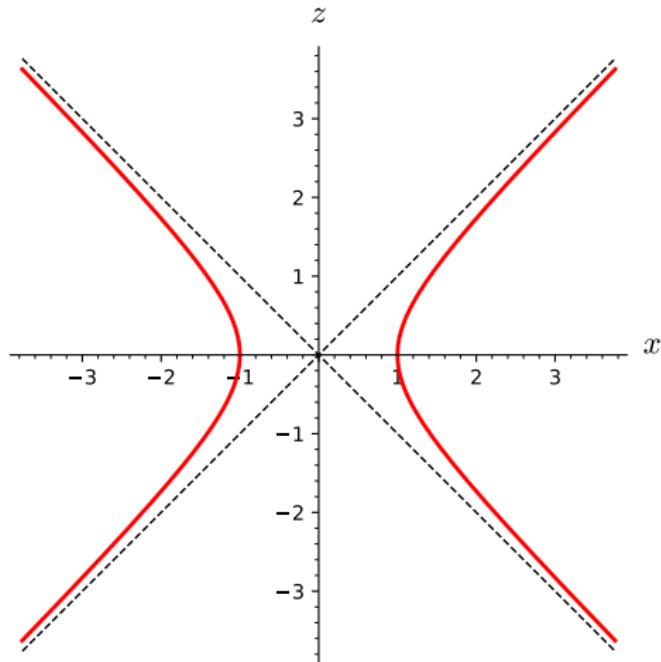


Figure: The hyperbola $x^2 - z^2 = 1$

When rotated around the z -axis, it yields a hyperboloid of one sheet (viz. $x^2 + y^2 - z^2 = 1$).

When rotated around the x -axis, it yields a hyperboloid of two sheets (viz. $x^2 - (y^2 + z^2) = 1$).

Mnemonic: The number of sheets of a hyperboloid is the number of minus signs in its standard equation.

Using the coordinate changes $x' = x/a$, $y' = y/b$, $z' = z/c$ with $a, b, c > 0$ (“stretching” the surface along the coordinate axes), we can produce from these quadratics further quadratics, which look similar but are metrically different.

For example the unit sphere generates in this way all *ellipsoids* with equations

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

General Quadrics

Every quadric $Q \subseteq \mathbb{R}^n$ is given by an equation

$$\mathbf{x}^T \mathbf{A} \mathbf{x} + 2\mathbf{b}^T \mathbf{x} + c = 0,$$

with a nonzero symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, a vector $\mathbf{b} \in \mathbb{R}^n$ and a real number c .

An affine coordinate change $\mathbf{x} = \mathbf{S}\mathbf{x}' + \mathbf{v}$ ($\mathbf{S} \in \mathbb{R}^{n \times n}$ invertible, $\mathbf{v} \in \mathbb{R}^n$), changes the equation to

$$\begin{aligned} & (\mathbf{S}\mathbf{x}' + \mathbf{v})^T \mathbf{A} (\mathbf{S}\mathbf{x}' + \mathbf{v}) + 2\mathbf{b}^T (\mathbf{S}\mathbf{x}' + \mathbf{v}) + c = 0 \\ \iff & \mathbf{x}'^T (\mathbf{S}^T \mathbf{A} \mathbf{S}) \mathbf{x}' + (2\mathbf{v}^T \mathbf{A} \mathbf{S} + 2\mathbf{b}^T \mathbf{S}) \mathbf{x}' + \mathbf{v}^T \mathbf{A} \mathbf{v} + 2\mathbf{b}^T \mathbf{v} + c = 0 \end{aligned}$$

If \mathbf{A} is invertible, we can eliminate the linear term by setting $\mathbf{S} = \mathbf{I}_n$ and solving $\mathbf{v}^T \mathbf{A} + \mathbf{b}^T = \mathbf{0}$ (which is equivalent to $\mathbf{A}\mathbf{v} = -\mathbf{b}$) for \mathbf{v} . In this case Q is said to be *central*, and \mathbf{v} is called the *center* of Q .

Example

The three example quadrics (the unit sphere and the two hyperboloids), as well as all quadrics obtained from these by “stretching”, are central with the origin $(0, 0, 0)$ as center.

The Classification of Space Quadrics

We will restrict ourselves to the non-degenerate case.

Definition

A quadric $Q \subseteq \mathbb{R}^n$ with equation $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2 \mathbf{b}^T \mathbf{x} + c = 0$ is said to be *non-degenerate* if

$\begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}$ is invertible (and *degenerate* otherwise).

Remarks

This is equivalent to the requirement that the so-called *projective extension* \overline{Q} of Q , which is the projective quadric defined by

$\bar{f}(\mathbf{x}, x_{n+1}) = \mathbf{x}^T \mathbf{A} \mathbf{x} + 2(\mathbf{b}^T \mathbf{x})x_{n+1} + cx_{n+1}^2 = 0$, has no singular points: The matrix representing \bar{f} (which is a quadratic form in $n+1$ variables) is $\bar{\mathbf{A}} = \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix}$. Since $\nabla \bar{f}(\bar{\mathbf{x}}) = 2\bar{\mathbf{A}}\bar{\mathbf{x}}$ for $\bar{\mathbf{x}} = (\mathbf{x}, x_{n+1}) \in \mathbb{R}^{n+1}$, invertibility of $\bar{\mathbf{A}}$ amounts to $\nabla \bar{f}(\bar{\mathbf{x}}) \neq \mathbf{0}$ for all $\bar{\mathbf{x}} \in \mathbb{R}^{n+1} \setminus \{\mathbf{0}\}$ and hence for all points on \overline{Q} .

It may also be restated as “ \bar{f} is not equivalent to a quadratic form in fewer than $n+1$ variables” (the case $t=0$ in Sylvester’s Theorem) and excludes cylinder- and cone-shaped quadrics.

Examples

- 1 The quadric C in \mathbb{R}^3 with equation $x^2 + y^2 - z^2 = 0$ is called a *cone*. Since the equation is equivalent to $z = \pm\sqrt{x^2 + y^2}$, C consists of two “copies” of the graph of the length function $(x, y) \mapsto \sqrt{x^2 + y^2}$, which have their cusps in $(0, 0, 0)$ joined. Since $f(x, y, z) = x^2 + y^2 - z^2$ has $\mathbf{b} = \mathbf{0}$ and $c = 0$, C is degenerate. Since $\nabla f(0, 0, 0) = \mathbf{0}$, the origin is a singular point of C (i.e., C has no tangent plane there).
- 2 A union of two planes in \mathbb{R}^3 forms a degenerate quadric, since the corresponding matrix \mathbf{A} is easily seen to have rank 1; cf. the proof of the subsequent classification theorem.
(If \mathbf{a}, \mathbf{b} are normal vectors for the planes then $\mathbf{A} = \mathbf{a}\mathbf{b}^T$.)
Geometrically this can be seen as follows: If the planes are not parallel then their line of intersection consist of singular points (the union has no tangent plane there). But if the planes are parallel (and distinct), there are no singular points in \mathbb{R}^3 and we must look at the projective extension:
3-dimensional (real) projective space may be thought of as \mathbb{R}^3 together with a plane at infinity. Parallel planes intersect in a line of this plane (their “line at infinity”), which consequently consists of singular points.

Examples (cont'd)

- ③ The cylinder surface C in \mathbb{R}^3 with equation $x^2 + y^2 = 1$ is degenerate, because the corresponding matrix \mathbf{A} has rank 2 and $\mathbf{b} = \mathbf{0}$, which gives $\text{rk} \begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^\top & c \end{pmatrix} = 3$.

The surface C is generated by parallel lines (the vertical lines $(\cos \phi, \sin \phi, 0) + \mathbb{R}(0, 0, 1)$, $0 \leq \phi < 2\pi$). In the projective space these lines meet at a point at infinity, so that the cylinder surface can be seen as a cone with its cusp at infinity. This is visible in the equations $x^2 + y^2 = z^2$ of the standard cone and $x^2 + y^2 = w^2$ of the projective extension \overline{C} . Just make the variable change $z \rightarrow w$.

- ④ The sphere $x^2 + y^2 + z^2 = 1$ and the hyperboloids $x^2 + y^2 - z^2 = 1$, $x^2 - y^2 - z^2 = 1$ are non-degenerate. In all three cases $\begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^\top & c \end{pmatrix}$ is a diagonal matrix with four entries ± 1 on the diagonal.

Exercise

Show that a quadric Q in \mathbb{R}^n with a singular point is degenerate.

Note: If Q has equation $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A} \mathbf{x} + 2 \mathbf{b}^\top \mathbf{x} + c = 0$, a point $\mathbf{x}_0 \in Q$ is said to be *singular* if $\nabla f(\mathbf{x}_0) = \mathbf{0}$.

Theorem

Let Q be a non-degenerate quadric in \mathbb{R}^3 with equation $\mathbf{x}^T \mathbf{A} \mathbf{x} + 2 \mathbf{b}^T \mathbf{x} + c = 0$. Then there are the following possibilities:

- ① $\text{rk}(\mathbf{A}) = 3$. In this case Q is central and, provided that $Q \neq \emptyset$, there is an affine coordinate change $\mathbf{x} = \mathbf{S}\mathbf{x}' + \mathbf{v}$ transforming Q into a sphere or a hyperboloid of one or two sheets.
- ② $\text{rk}(\mathbf{A}) = 2$. In this case Q is non-central, and there is an affine coordinate change transforming Q into either the quadric with equation $z = x^2 + y^2$ (elliptic paraboloid) or the quadric with equation $z = x^2 - y^2$ (hyperbolic paraboloid).

Proof of the theorem.

If $\text{rk}(\mathbf{A}) \leq 1$ then $\begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix}$, which is obtained from \mathbf{A} by adding one column and one row, has rank at most 3. Since Q is non-degenerate, this case cannot occur.

Case 1: $\text{rk}(\mathbf{A}) = 3$

We have seen that in this case Q is central and a suitable translation $\mathbf{x} \mapsto \mathbf{x} - \mathbf{v}$ transforms Q into a quadric with equation $\mathbf{x}^T \mathbf{A} \mathbf{x} + c = 0$. Since this quadric is non-degenerate, we must have $c \neq 0$. Further, Sylvester's Interia Theorem yields an invertible matrix $\mathbf{S} \in \mathbb{R}^{3 \times 3}$ such that $\mathbf{S}^T \mathbf{A} \mathbf{S}$ is one of

$$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}, \begin{pmatrix} 1 & & \\ & 1 & \\ & & -1 \end{pmatrix}, \begin{pmatrix} 1 & & \\ & -1 & \\ & & -1 \end{pmatrix}, \begin{pmatrix} -1 & & \\ & -1 & \\ & & -1 \end{pmatrix}.$$

The overall coordinate change $\mathbf{x} = \mathbf{S}\mathbf{x}' + \mathbf{v}$ then transforms Q into one of the quadrics with equations $x^2 + y^2 + z^2 = -c$, $x^2 + y^2 - z^2 = -c$, $x^2 - y^2 - z^2 = -c$, $-x^2 - y^2 - z^2 = -c$.

Finally we can achieve $c = \pm 1$ by applying the coordinate change $x' = x/\sqrt{|c|}$, $y' = y/\sqrt{|c|}$, $z' = z/\sqrt{|c|}$.

Now $x^2 + y^2 + z^2 = \pm 1$ represent \emptyset and a sphere, $x^2 + y^2 - z^2 = \pm 1$ the two hyperboloids, and the last two equations are redundant.

Proof cont'd.

Case 2: $\text{rk}(\mathbf{A}) = 2$

Here Sylvester's Theorem gives a coordinate change $\mathbf{x} = \mathbf{S}\mathbf{x}'$ that transforms Q into a quadric Q' with equation

$$\pm x^2 \pm y^2 + 2b_1x + 2b_2y + 2b_3z + c = 0.$$

Since

$$\begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{pmatrix} = \begin{pmatrix} \pm 1 & & b_1 \\ & \pm 1 & b_2 \\ b_1 & b_2 & b_3 & c \end{pmatrix}$$

is invertible, we must have $b_3 \neq 0$.

The affine coordinate change $z' = 2b_1x + 2b_2y + 2b_3z + c = 0$ (and $x' = x$, $y' = y$) then transforms Q' into one of the two quadrics with equations $z = x^2 \pm y^2$ (possibly with an additional sign change of z'). □

Remarks

- 1 The classification of space quadrics stated in [Ste21], Chapter 12.6 (see Table 1) comprises 6 types, since it includes the cone with equation $x^2 + y^2 = z^2$. Since the cone is degenerate, it is not included in our theorem.
- 2 Sylvester's Theorem also gives the classification of projective space quadrics, which are sets of projective zeros of quadratic forms $q(x, y, z, w)$. (A point in 3-dimensional real projective space has the form $\mathbb{R}(x, y, z, w)$ with $(x, y, z, w) \neq \mathbf{0}$. If $w \neq 0$, we can normalize the generator to $(x, y, z, 1)$. This accounts for the “finite” points in \mathbb{R}^3 . The points with $w = 0$ form the plane E_∞ at infinity.)

There are 3 types of non-degenerate projective quadrics. These have equations $x^2 + y^2 + z^2 + w^2 = 0$ (the empty quadric), $x^2 + y^2 + z^2 - w^2 = 0$ (the sphere; to see this, set $w = 1$) and $x^2 + y^2 - z^2 - w^2 = 0$ (the hyperboloid of one sheet). All 5 types of affine quadrics arise from the sphere and the hyperboloid by choosing the plane at infinity in a certain way. (For example, an elliptic paraboloid comes from a sphere if E_∞ is chosen as a tangent plane of the sphere.)

The following theorem, which is harder to prove than Sylvester's Inertia Theorem, allows for the refined classification of space (and higher-dimensional) quadrics up to isometry.

Theorem (Principal Axes Theorem)

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be symmetric. Then there exists an orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ such that

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

is a diagonal matrix.

The numbers $\lambda_1, \dots, \lambda_n$ are the so-called *eigenvalues* of \mathbf{A} . Their sign distribution must be the same as in Sylvester's normal form (i.e., r positive eigenvalues, s negative eigenvalues, and t times the eigenvalue 0). The columns $\mathbf{q}_1, \dots, \mathbf{q}_n$ of \mathbf{Q} (or the corresponding lines $\mathbb{R}\mathbf{q}_1, \dots, \mathbb{R}\mathbf{q}_n$) are called *principal axes*, since the coordinate change $\mathbf{x} = x'_1 \mathbf{q}_1 + \dots + x'_n \mathbf{q}_n$ turns the quadratic form $q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ into $q(\mathbf{x}) = q(\mathbf{Qx}') = \lambda_1 {x'}_1^2 + \dots + \lambda_n {x'}_n^2$.

Combining the Principal Axes Theorem with the previous classification theorem, we obtain that every non-degenerate quadric in \mathbb{R}^3 is metrically equivalent (isometric) to one of the following five types:

- ① $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1, \quad a \geq b \geq c > 0;$
- ② $x^2/a^2 + y^2/b^2 - z^2/c^2 = 1, \quad a \geq b > 0, c > 0;$
- ③ $x^2/a^2 - y^2/b^2 - z^2/c^2 = 1, \quad a > 0, b \geq c > 0;$
- ④ $z/c = x^2/a^2 + y^2/b^2, \quad a \geq b > 0, c > 0;$
- ⑤ $z/c = x^2/a^2 - y^2/b^2, \quad a, b, c > 0.$

Exercise

Determine the non-degenerate types of space quadrics that contain a line; cf. Exercise 51 in [Ste21], Chapter 12.6.

Notes: Surfaces containing a line through each of its points are called *ruled surfaces* and are important in architectural/engineering design. Since affine coordinate changes preserve this property, it suffices to check the 5 affine normal forms $x^2 + y^2 + z^2 = 1$, $x^2 + y^2 - z^2 = 1$, $x^2 - y^2 - z^2 = 1$, $z = x^2 + y^2$, $z = x^2 - y^2$ derived in the classification theorem.

Example

We determine the type of the quadric Q in \mathbb{R}^3 defined by

$$xy + xz + yz + x + y + z + 1 = 0.$$

Multiplying the equation by 2 turns it into

$(x, y, z)\mathbf{A}(x, y, z)^T + 2\mathbf{b}^T(x, y, z)^T + c = 0$ with an integral matrix \mathbf{A} , viz.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad c = 2.$$

Since $\text{rk}(\mathbf{A}) = 3$, Q is central. The center \mathbf{v} is obtained by solving $\mathbf{Av} = -\mathbf{b} = (-1, -1, -1)^T$, which gives $\mathbf{v} = (-\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2})^T$.

Since

$$\mathbf{v}^T \mathbf{A} \mathbf{v} + 2\mathbf{b}^T \mathbf{v} + c = \frac{6}{4} - 3 + 2 = \frac{1}{2},$$

Q is affinely equivalent to the quadric with equation $\mathbf{x}^T \mathbf{A} \mathbf{x} + \frac{1}{2} = 0$ (or $xy + xz + yz + \frac{1}{4} = 0$).

Example (cont'd)

Next we apply the algorithm in the proof of Sylvester's Inertia Theorem to \mathbf{A} :

$$\begin{array}{c}
 \left(\begin{array}{ccc} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{array} \right) \xrightarrow{R1=R1+R2} \left(\begin{array}{ccc} 1 & 1 & 2 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{array} \right) \xrightarrow{C1=C1+C2} \left(\begin{array}{ccc} 2 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{array} \right) \\
 \\
 \xrightarrow{\substack{R3=R3-R1 \\ R2=R2-\frac{1}{2}R1}} \left(\begin{array}{ccc} 2 & 1 & 2 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -2 \end{array} \right) \xrightarrow{\substack{C3=C3-C1 \\ C2=C2-\frac{1}{2}C1}} \left(\begin{array}{ccc} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -2 \end{array} \right) \\
 \\
 \rightarrow \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{array} \right)
 \end{array}$$

$\implies Q$ is affinely equivalent to the quadric with equation $x^2 - y^2 - z^2 = -\frac{1}{2}$, which is a hyperboloid of one sheet.

(To see this, scale the variables by $\sqrt{2}$ to turn the equation into $x^2 - y^2 - z^2 = -1$, and then multiply by -1 to obtain $-x^2 + y^2 + z^2 = 1$.)

Example (cont'd)

Let us remark that Q is the graph of the function

$$z = f(x, y) = \frac{-xy - x - y - 1}{x + y + 1}, \quad x + y \neq -1,$$

with the two vertical lines defined by $(x, y) = (-1, 0)$ and $(x, y) = (0, -1)$ adjoined. This indicates already that Q is a ruled surface (and hence cannot be an ellipsoid, elliptic paraboloid, or hyperboloid of two sheets).

At the end we outline how the principal axes of Q (or \mathbf{A}) can be found. The basic idea is to rewrite the equation

$$\mathbf{Q}^{-1} \mathbf{A} \mathbf{Q} = \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix} \quad \text{as} \quad \mathbf{A} \mathbf{Q} = \mathbf{Q} \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix},$$

which is equivalent to $\mathbf{A} \mathbf{q}_i = \lambda_i \mathbf{q}_i$ for $i = 1, 2, 3$.

Thus we need to find vectors in \mathbb{R}^3 that are mapped to scalar multiples of themselves by $\mathbf{x} \mapsto \mathbf{Ax}$ (so-called *eigenvectors* of \mathbf{A}).

Example (cont'd)

Here we can easily guess such a vector: Since \mathbf{A} has constant row sums 2, we have $\mathbf{A}(1, 1, 1)^T = 2(1, 1, 1)^T$. Normalizing to unit length preserves this property, so that $\mathbf{q}_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^T$ also satisfies $\mathbf{A}\mathbf{q}_1 = 2\mathbf{q}_1$.

Since \mathbf{Q} is supposed to be orthogonal, we extend \mathbf{q}_1 to an orthonormal basis $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$ of \mathbb{R}^3 and take $\mathbf{Q} = (\mathbf{q}_1 | \mathbf{q}_2 | \mathbf{q}_3)$, e.g.,

$$\tilde{\mathbf{Q}} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & -2 \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \end{pmatrix}.$$

Then we compute $\mathbf{A}\mathbf{q}_2$ and $\mathbf{A}\mathbf{q}_3$:

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} = -\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix} = \begin{pmatrix} -1/\sqrt{6} \\ -1/\sqrt{6} \\ 2/\sqrt{6} \end{pmatrix} = -\begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{pmatrix}.$$

Example

Thus $\mathbf{A}\mathbf{q}_1 = 2\mathbf{q}_1$, $\mathbf{A}\mathbf{q}_2 = -\mathbf{q}_2$, $\mathbf{A}\mathbf{q}_3 = -\mathbf{q}_3$, implying that

$$\mathbf{Q}^T \mathbf{A} \mathbf{Q} = \mathbf{Q}^{-1} \mathbf{A} \mathbf{Q} = \begin{pmatrix} 2 & & \\ & -1 & \\ & & -1 \end{pmatrix}.$$

It follows that Q is isometric to the quadric with equation $2x^2 - y^2 - z^2 + \frac{1}{2} = 0$.

The principal axes of Q are the lines $\mathbf{v} + \mathbb{R}\mathbf{q}_i$, and Q is rotation-symmetric with axis $\mathbf{v} + \mathbb{R}\mathbf{q}_1 = \mathbb{R}(1, 1, 1)$.

Writing the equation in the more convenient form

$-4x^2 + 2y^2 + 2z^2 = 1$, we can see that Q arises from rotating the hyperbola $-4x'^2 + 2y'^2 = 1$ in the plane $\mathbf{v} + \mathbb{R}\mathbf{q}_1 + \mathbb{R}\mathbf{q}_2$ around $\mathbb{R}(1, 1, 1)$ (Coordinates in the plane are relative to \mathbf{v} and the ordered basis $\mathbf{q}_1, \mathbf{q}_2$.)

Two Final Notes

- In this particular example we were lucky to find \mathbf{Q} from only one eigenvalue of \mathbf{A} . This is due to the fact that \mathbf{A} has only two eigenvalues, $\lambda_1 = 2$ of multiplicity 1 and $\lambda_2 = \lambda_3 = -1$ of multiplicity 2. In general, one needs to find a second eigenvalue of a 3×3 matrix before \mathbf{Q} can be computed. But there is a standard way to compute the eigenvalues of an $n \times n$ matrix, which uses the so-called characteristic polynomial of the matrix.
- For $q(x, y, z) = 2xy + 2xz + 2yz = \mathbf{x}^T \mathbf{A} \mathbf{x}$, $\mathbf{x} = (x, y, z)$, we have established the identity $q(\mathbf{Q}\mathbf{x}) = 2x^2 - y^2 - z^2 = q'(\mathbf{x})$, say. Since $\mathbf{Q}^{-1} = \mathbf{Q}^T$, this identity can be easily inverted to

$$\begin{aligned} q(\mathbf{x}) &= q'(\mathbf{Q}^T \mathbf{x}) = 2 \left(\frac{x+y+z}{\sqrt{3}} \right)^2 - \left(\frac{x-y}{\sqrt{2}} \right)^2 - \left(\frac{x+y-2z}{\sqrt{6}} \right)^2 \\ &= \frac{2}{3}(x+y+z)^2 - \frac{1}{2}(x-y)^2 - \frac{1}{6}(x+y-2z)^2, \quad \text{or} \\ xy + xz + yz &= \frac{1}{3}(x+y+z)^2 - \frac{1}{4}(x-y)^2 - \frac{1}{12}(x+y-2z)^2. \end{aligned}$$

Would you have found this without the machinery developed?

Math 241
Calculus III

Thomas
Honold

Riemann's vs
Darboux's
Definition

Continuous
Functions

Problems

Higher-
Dimensional
Riemann
Integrals

Iterated
Integrals

New Problems

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Riemann's vs Darboux's Definition

2 Continuous Functions

3 Problems

4 Higher-Dimensional Riemann Integrals

5 Iterated Integrals

6 New Problems

Riemann's vs
Darboux's
Definition

Continuous
Functions

Problems

Higher-
Dimensional
Riemann
Integrals

Iterated
Integrals

New Problems

Today's Lecture: Introduction to Multivariate Integration

Riemann Integral

First recall the definition of the 1-dimensional Riemann integral from Calculus I.

Definition (Riemann)

A function $f: [a, b] \rightarrow \mathbb{R}$ is *Riemann integrable* with $\int_a^b f(x) dx = V$ if for every “error bound” $\epsilon > 0$ there is a “response” $\delta > 0$ such that for every partition $a = x_0 < x_1 < \dots < x_N = b$ with subintervals $[x_{i-1}, x_i]$ of lengths $< \delta$ and any choice of “sample points” $x_i^* \in [x_{i-1}, x_i]$ we have

$$\left| \sum_{i=1}^N f(x_i^*)(x_i - x_{i-1}) - V \right| < \epsilon.$$

This definition makes the intuitive requirement that the *Riemann sums* $\sum_{i=1}^N f(x_i^*)(x_i - x_{i-1})$ should approach the integral of f when the mesh size tends to zero precise.

Darboux's Definition

Darboux sums

Suppose $f: [a, b] \rightarrow \mathbb{R}$ is bounded (from above and from below). For a partition $P: a = x_0 < x_1 < \dots < x_N = b$ and $i \in \{1, \dots, N\}$ let

$$m_i = \inf\{f(x); x_{i-1} \leq x \leq x_i\}, \quad M_i = \sup\{f(x); x_{i-1} \leq x \leq x_i\},$$

and define the *lower* and *upper Darboux sum* of f with respect to P as

$$\underline{S}(P; f) = \sum_{i=1}^N m_i(x_i - x_{i-1}), \quad \overline{S}(P; f) = \sum_{i=1}^N M_i(x_i - x_{i-1})$$

In the following we view a partition of $[a, b]$ as a finite subset $P \subset [a, b]$ with $a, b \in P$. Writing $|P| = N + 1$, P can be ordered as $P = \{x_0, x_1, \dots, x_N\}$ with $a = x_0 < x_1 < \dots < x_N = b$. The set of all partitions of $[a, b]$ will be denoted by \mathcal{P} .

Properties of Darboux Sums

- 1 If $Q \in \mathcal{P}$ arises from $P \in \mathcal{P}$ by adding one or more points (set-theoretically, $Q \supset P$) then

$$\underline{S}(P; f) \leq \underline{S}(Q; f) \leq \overline{S}(Q; f) \leq \overline{S}(P; f).$$

To prove this, it suffices to consider the case $Q = P \cup \{x^*\}$ and prove the inequality $\overline{S}(Q; f) \leq \overline{S}(P; f)$. If $P = \{x_0, \dots, x_N\}$ and $x^* \in [x_{i-1}, x_i]$, we have

$$\overline{S}(P; f) - \overline{S}(Q; f) = M_i(x_i - x_{i-1}) - M'(x^* - x_{i-1}) - M''(x_i - x^*),$$

where $M' = \sup\{f(x); x_{i-1} \leq x \leq x^*\}$, $M'' = \sup\{f(x); x^* \leq x \leq x_i\}$. But $M_i = \max\{M', M''\}$ and hence $M'(x^* - x_{i-1}) + M''(x_i - x^*) \leq M_i(x^* - x_{i-1}) + M_i(x_i - x^*) = M_i(x_i - x_{i-1})$, as desired.

- 2 For all $P, Q \in \mathcal{P}$ we have $\underline{S}(P; f) \leq \overline{S}(Q; f)$.

This is clear in the case $P = Q$ and follows in the general case by squeezing in the Darboux sums for the *common refinement* $P \cup Q$ according to Property (1):

$$\underline{S}(P; f) \leq \underline{S}(P \cup Q; f) \leq \overline{S}(P \cup Q; f) \leq \overline{S}(Q; f).$$

Definition (Darboux)

A bounded function $f: [a, b] \rightarrow \mathbb{R}$ is Riemann integrable with $\int_a^b f(x) dx = V$ if the *lower* and *upper Darboux integrals*

$$\underline{\int_a^b} f(x) dx = \sup \{ \underline{S}(P; f); P \in \mathcal{P} \},$$

$$\overline{\int_a^b} f(x) dx = \inf \{ \overline{S}(P; f); P \in \mathcal{P} \}$$

have the same value V .

Notes

- The inequality $\underline{\int_a^b} f(x) dx \leq \overline{\int_a^b} f(x) dx$ holds in general, as follows easily from $\underline{S}(P; f) \leq \overline{S}(Q; f)$ for $P, Q \in \mathcal{P}$.
For non-integrable functions there is a gap between these two quantities. (We will see an example later.)

Notes cont'd

- Darboux's definition is much easier to handle because it makes no reference to "sample points" and "mesh size". To show that $\int_a^b f(x) dx$ exists, it suffices to find for each $\epsilon > 0$ a partition P of $[a, b]$ satisfying

$$\overline{S}(P; f) - \underline{S}(P; f) < \epsilon,$$

as is clear from the following chain of inequalities:

$$\underline{S}(P; f) \leq \underline{\int_a^b f(x) dx} \leq \overline{\int_a^b f(x) dx} \leq \overline{S}(P; f)$$

(Finding for each $\epsilon > 0$ two possibly different partitions $P, Q \in \mathcal{P}$ satisfying $\overline{S}(Q; f) - \underline{S}(P; f) < \epsilon$ would also do the job.)

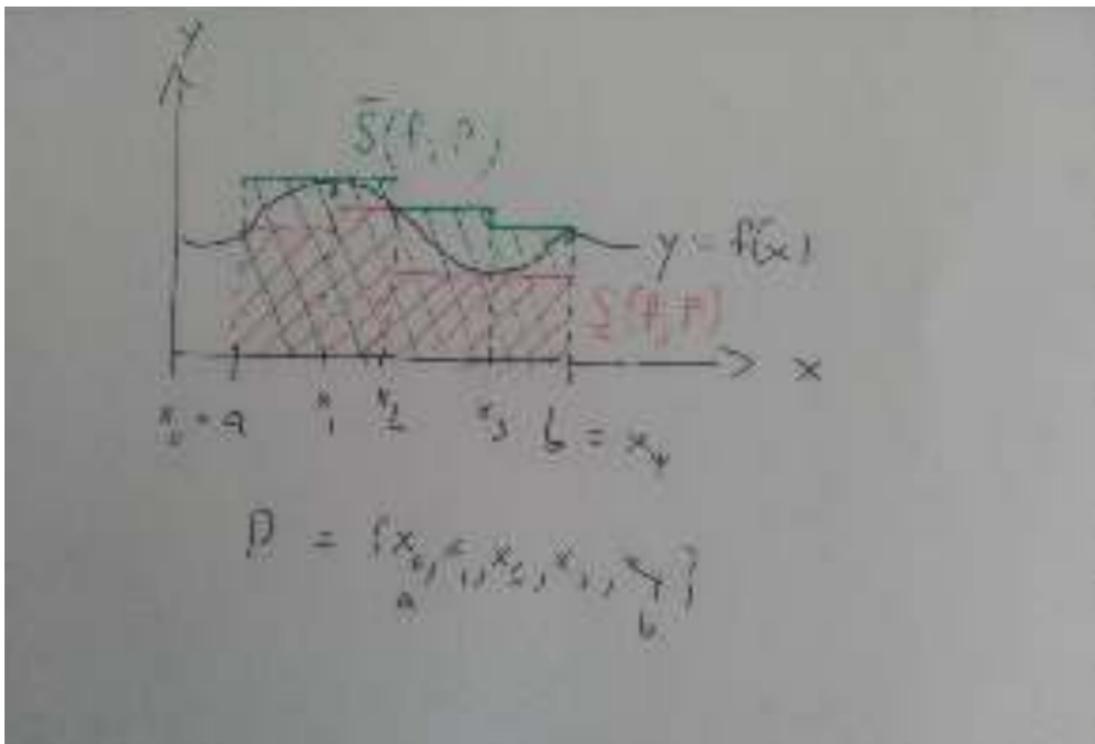


Figure: Illustration of lower and upper Darboux sums

Equivalence of the two Definitions

Theorem

A bounded function $f: [a, b] \rightarrow \mathbb{R}$ is integrable according to

Riemann's original definition if and only if $\underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx$.

If this is the case then $\underline{\int_a^b} f(x) dx = \underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx$.

Proof.

\Rightarrow : Writing $V = \underline{\int_a^b} f(x) dx$, let $\epsilon > 0$ be given.

There exists a partition $P: a = x_0 < x_1 < \dots < x_N = b$ such that for any choice of sample points $x_i^* \in [x_{i-1}, x_i]$ we have

$$\sum_{i=1}^N f(x_i^*)(x_i - x_{i-1}) \geq V - \epsilon.$$

This inequality must remain valid if we replace each $f(x_i^*)$ by $m_i = \inf\{f(x); x \in [x_{i-1}, x_i]\}$, showing that $\underline{S}(P; f) \geq V - \epsilon$.

In the same way one proves $\overline{S}(P; f) \leq V + \epsilon$.

Both inequalities together clearly imply $\underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx = V$.

Proof con't.

\iff : Writing $V = \underline{\int_a^b} f(x) dx = \overline{\int_a^b} f(x) dx$, we can find partitions P_1 and P_2 of $[a, b]$ such that

$$V - \epsilon/2 < \underline{S}(P_1; f) \leq V \leq \overline{S}(P_2; f) < V + \epsilon/2.$$

Denoting by P_3 the common refinement of P_1 and P_2 (set-theoretically, $P_3 = P_1 \cup P_2$), we have $\underline{S}(P_3; f) \geq \underline{S}(P_1; f)$, $\overline{S}(P_3; f) \leq \overline{S}(P_2; f)$, and hence also

$$V - \epsilon/2 < \underline{S}(P_3; f) \leq V \leq \overline{S}(P_3; f) < V + \epsilon/2.$$

\implies For the particular partition P_3 , say

$a = x_0 < x_1 < \dots < x_N = b$, also every Riemann sum $\sum_{i=1}^N f(x_i^*)(x_i - x_{i-1})$ satisfies the inequality

$$V - \epsilon/2 < \sum_{i=1}^N f(x_i^*)(x_i - x_{i-1}) < V + \epsilon/2.$$

The same is true for all refinements $P'_3 \supset P_3$, since

$\underline{S}(P_3; f) \leq \underline{S}(P'_3; f) \leq \overline{S}(P'_3; f) \leq \overline{S}(P_3; f)$ and the corresponding Riemann sums are between $\underline{S}(P'_3; f)$ and $\overline{S}(P'_3; f)$.

Proof cont'd.

Question: Does this hold for all partitions of $[a, b]$ whose mesh size is less than or equal to that of P_3 ?

Answer: No in general, but if we make the threshold δ for the mesh size sufficiently small then the Darboux sums $\underline{S}(P; f)$, $\overline{S}(P; f)$ (and hence also the corresponding Riemann sums) for partitions P of mesh size $< \delta$ will be close enough to $\underline{S}(P_3; f)$ and $\overline{S}(P_3; f)$, respectively.

A suitable choice is

$$\delta = \frac{\epsilon}{2N(M - m)}, \quad \text{where} \quad m = \inf_{x \in [a,b]} f(x), \quad M = \sup_{x \in [a,b]} f(x).$$

(The case $M = m$, i.e. f is constant, is trivial.)

To verify the claim, check that, e.g., the upper Darboux sum

$\overline{S}(Q; f)$ of the refinement $Q = P \cup P_3$, which satisfies

$\overline{S}(Q; f) \leq \overline{S}(P_3; f)$, differs from $\overline{S}(P; f)$ by no more than $\epsilon/2$: The difference has at most $N - 1$ terms (one for each subinterval of P that contains a point of P_3 in its interior) and the size of each such term is bounded by $\delta(M - m)$.) This implies

$$\overline{S}(P; f) \leq \overline{S}(Q; f) + \epsilon/2 \leq \overline{S}(P_3; f) + \epsilon/2 \leq V + \epsilon. \quad \square$$

Example

We compute $\int_a^b x^n dx$ for positive integers n and $0 < a < b$ using Darboux's definition. (If you have ever been asked to compute $\int_1^2 x^n dx$ for $n = 2, 3, 4$, say, using partitions with points $x_i = 1/N$, you will appreciate this example and Darboux's definition.)

Consider the partitions $P_N = \{aq^i; 0 \leq i \leq N\}$ of $[a, b]$ with $q = \sqrt[N]{b/a}$. The points $x_i = aq^i$ of P_N are in *geometric progression* (i.e., $x_i/x_{i-1} = q$ is constant), which greatly facilitates the computation of the corresponding Darboux sums.

Clearly $f(x) = x^n$ attains its infimum/supremum in $[x_{i-1}, x_i] = [aq^{i-1}, aq^i]$ at the left/right endpoint.

$$\begin{aligned}\implies \underline{S}(P_N; f) &= \sum_{i=1}^N (aq^{i-1})^n (aq^i - aq^{i-1}) \\ &= a^{n+1} (q-1) \sum_{i=1}^N q^{(i-1)(n+1)} = a^{n+1} (q-1) \frac{q^{N(n+1)} - 1}{q^{n+1} - 1} \\ &= \frac{b^{n+1} - a^{n+1}}{(q^{n+1} - 1)/(q-1)} \rightarrow \frac{b^{n+1} - a^{n+1}}{n+1} \quad \text{for } N \rightarrow \infty,\end{aligned}$$

since $q = \sqrt[N]{b/a} \rightarrow 1$.

Example (cont'd)

Further we have

$$\overline{S}(P_N; f) = \sum_{i=1}^N (aq^i)^n (aq^i - aq^{i-1}) = q^n \cdot \underline{S}(P_N; f),$$

and hence

$$\overline{S}(P_N; f) - \underline{S}(P_N; f) = (q^n - 1)\underline{S}(P_N; f) \rightarrow 0 \cdot \frac{b^{n+1} - a^{n+1}}{n+1} = 0$$

for $N \rightarrow \infty$.

Using Darboux's definition, this shows that $\int_a^b x^n dx$ exists and has the value

$$\int_a^b x^n dx = \frac{b^{n+1} - a^{n+1}}{n+1}.$$

This is well-known, of course, but the usual proof requires the Fundamental Theorem of Calculus.

Continuous Functions are Integrable

Lemma

Suppose $f: [a, b] \rightarrow \mathbb{R}$ is continuous, i.e., for every $\epsilon > 0$ and every point $x_0 \in [a, b]$ there exists a “response” $\delta = \delta(x_0) > 0$ such that $|f(x) - f(x_0)| < \epsilon$ whenever $x \in [a, b]$ and $|x - x_0| < \delta$.

Then f is uniformly continuous, i.e., for every $\epsilon > 0$ there exists a response $\delta > 0$ that works for all $x_0 \in [a, b]$ simultaneously:

$$|f(x) - f(x_0)| < \epsilon \quad \text{whenever } x, x_0 \in [a, b] \text{ and } |x - x_0| < \delta.$$

Notes

- This property is usually stated as “ $|f(x) - f(y)| < \epsilon$ whenever $x, y \in [a, b]$ and $|x - y| < \delta$.”
- It is essential that the domain of f is closed and bounded. Continuous functions on arbitrary intervals are not necessarily uniformly continuous.
- The lemma generalizes easily to continuous maps $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$. The domain D must be closed and bounded (but may have more general shape than a cuboid), and the absolute value on \mathbb{R} is replaced by the Euclidean length functions on \mathbb{R}^n and \mathbb{R}^m .

Proof of the Lemma.

A proof based on the Bolzano-Weierstrass Theorem was given earlier. The following alternative proof uses the HEINE-BOREL covering property of compact intervals.

For every point $x \in [a, b]$ choose a response $\delta(x)$ which works for that particular point. Then the open intervals

$$I(x) = (x - \delta(x)/2, x + \delta(x)/2), \quad x \in [a, b],$$

cover $[a, b]$. We claim that there exists a finite subcover, i.e., points x_1, \dots, x_N such that

$$[a, b] \subseteq \bigcup_{i=1}^N I(x_i).$$

Assume, by contradiction, that this is false. Divide $I_0 = [a, b]$ into two closed intervals of equal length, viz. $[a, (a+b)/2]$ and $[(a+b)/2, b]$. One of these intervals, call it I_1 , does not have a finite subcover as well. Continuing in this way, we find a sequence $I_0 \supset I_1 \supset I_2 \supset \dots$ of closed intervals such that I_n has length $(b-a)/2^n$ and no finite subcover. The nested intervals principle then gives $\bigcap_{n=0}^{\infty} I_n = \{s\}$ for some $s \in \mathbb{R}$.

Proof of the lemma cont'd.

$\implies I_n \subset I(s)$ for sufficiently large n , and thus I_n is covered by a single interval from the family $I(x)$, $x \in [a, b]$. Contradiction!

Now we can easily finish the proof. Suppose $x, y \in [a, b]$ satisfy $|x - y| < \delta/2$. By what we have just shown, $x \in I(x_i)$ for some $1 \leq i \leq N$, which implies

$$|x - x_i| < \delta/2 < \delta,$$

$$|y - x_i| \leq |y - x| + |x - x_i| < \delta/2 + \delta/2 = \delta$$

$$\implies |f(x) - f(y)| \leq |f(x) - f(x_i)| + |f(x_i) - f(y)| < 2\epsilon$$

Changing ϵ to $\epsilon/2$ everywhere in the proof (that is, the initial response $\delta(x_0)$ should be that for $\epsilon/2$) completes the proof. □

Theorem

Every continuous function $f: [a, b] \rightarrow \mathbb{R}$ is Riemann integrable.

Proof.

Given $\epsilon > 0$, choose $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. This is possible, since f is uniformly continuous.

Consider now a partition $P: a = x_0 < x_1 < \dots < x_N = b$ of mesh size $< \delta$. Clearly such partitions exist. (For example we can take $x_i = a + \frac{i}{N}(b - a)$ with $N > (b - a)/\delta$.)

For $x, y \in [x_{i-1}, x_i]$ we then have $|f(x) - f(y)| < \epsilon$ and hence

$$M_i - m_i = \sup\{f(x); x \in [x_{i-1}, x_i]\} - \inf\{f(y); y \in [x_{i-1}, x_i]\} \leq \epsilon.$$

$$\begin{aligned} \implies \bar{S}(P; f) - \underline{S}(P; f) &= \sum_{i=1}^N (M_i - m_i)(x_i - x_{i-1}) \\ &\leq \epsilon \sum_{i=1}^N (x_i - x_{i-1}) = \epsilon(b - a). \end{aligned}$$

This shows that the difference $\bar{S}(P; f) - \underline{S}(P; f)$ can be made arbitrarily small and implies that $\underline{\int_a^b f(x) dx} = \overline{\int_a^b f(x) dx}$. □

Problems with the Riemann Integral

Riemann's vs
Darboux's
Definition

Continuous
Functions

Problems

Higher-
Dimensional
Riemann
Integrals

Iterated
Integrals

New Problems

- ① The class of functions which are Riemann integrable is too restricted.
- ② There is no satisfactory theory describing the integration of sequences/series of functions.

These defects make the Riemann integral unsuitable for applications in several modern areas of advanced mathematics, where integration is an indispensable tool.

Example

Consider the *Dirichlet function* $f: [0, 1] \rightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \notin \mathbb{Q}. \end{cases}$$

Enumerating \mathbb{Q} as q_1, q_2, q_3, \dots and setting $f_n(x) = 1$ if $x = q_n$ and $f_n(x) = 0$ otherwise (the domain of f_n is the same as for f), we have

$$f(x) = \sum_{n=1}^{\infty} f_n(x) \quad \text{for every } x \in [0, 1].$$

Clearly each f_n is Riemann integrable with $\int_0^1 f_n(x) dx = 0$. In this situation we would like to have

$$\int_0^1 f(x) dx = \int_0^1 \left(\sum_{n=1}^{\infty} f_n(x) \right) dx = \sum_{n=1}^{\infty} \int_0^1 f_n(x) dx = \sum_{n=1}^{\infty} 0 = 0$$

as well. But, unfortunately, f is not Riemann integrable. (Check that $\underline{S}(P; f) = 0$, $\overline{S}(P; f) = 1$ for every partition of $[0, 1]$.)

Example

Consider the function $g: [-1, 1] \rightarrow \mathbb{R}$ defined by

$$g(x) = \begin{cases} \frac{1}{\sqrt{|x|}} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

Since g is unbounded, g is not Riemann integrable (at least not when following the definition of Darboux).

We have learned in Calculus I or II how to do this so-called *improper integral* nevertheless:

$$\begin{aligned} \int_{-1}^1 \frac{dx}{\sqrt{|x|}} &= \lim_{\epsilon_1, \epsilon_2 \downarrow 0} \left(\int_{-1}^{-\epsilon_1} \frac{dx}{\sqrt{-x}} + \int_{\epsilon_2}^1 \frac{dx}{\sqrt{x}} \right) \\ &= \lim_{\epsilon_1, \epsilon_2 \downarrow 0} \left([-2\sqrt{-x}]_{-1}^{-\epsilon_1} + [2\sqrt{x}]_{\epsilon_2}^1 \right) \\ &= \lim_{\epsilon_1, \epsilon_2 \downarrow 0} (-2\sqrt{\epsilon_1} + 2 + 2 - 2\sqrt{\epsilon_2}) \\ &= 4. \end{aligned}$$

Wouldn't it be nice to have this as a properly defined integral?

The Higher-Dimensional Case

There is no difficulty in extending Riemann's and Darboux's definition of $\int_a^b f(x) dx$ to dimensions $n > 1$.

- 1-dimensional intervals $[a, b] \subset \mathbb{R}$ are replaced by n -dimensional intervals (cuboids)

$$\begin{aligned} [\mathbf{a}, \mathbf{b}] &= \{\mathbf{x} \in \mathbb{R}^n; a_i \leq x_i \leq b_i \text{ for } 1 \leq i \leq n\} \\ &= [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]. \end{aligned}$$

- Single partitions $a = x_0 < x_1 < \cdots < x_N = b$ are replaced by cartesian products $P_1 \times \cdots \times P_n$ of partitions $P_i: a_i = x_0^{(i)} < x_1^{(i)} < \cdots < x_{N_i}^{(i)} = b_i$, decomposing $[\mathbf{a}, \mathbf{b}]$ into $N_1 N_2 \cdots N_n$ smaller intervals in a grid-like fashion.
- The length $b - a$ of a 1-dimensional interval (and similarly for its subintervals $[x_{i-1}, x_i]$) is replaced by the volume

$$\text{vol}([\mathbf{a}, \mathbf{b}]) = \prod_{i=1}^n (b_i - a_i).$$

- n -dimensional Riemann and Darboux sums, Riemann integrability, the integral $\int_{[\mathbf{a}, \mathbf{b}]} f(\mathbf{x}) d^n \mathbf{x}$, and lower/upper Darboux integrals are defined in the obvious way—except, maybe, for the mesh size of (P_1, \dots, P_n) , which is taken as the maximum length of any subinterval appearing in one of P_1, \dots, P_n .
- Darboux's definition of the n -dimensional Riemann integral is again equivalent to Riemann's definition and simplifies proofs considerably.
- In the n -dimensional case we also have that continuous functions $f: [\mathbf{a}, \mathbf{b}] \rightarrow \mathbb{R}$ are Riemann-integrable. The proof uses the uniform continuity of a continuous function on the closed and bounded n -dimensional interval $[\mathbf{a}, \mathbf{b}]$. If $\delta > 0$ is such that $|f(\mathbf{x}) - f(\mathbf{y})| < \epsilon$ whenever $\mathbf{x}, \mathbf{y} \in [\mathbf{a}, \mathbf{b}]$ and $\mathbf{y} \in B_\delta(\mathbf{x})$, then

$$0 \leq \overline{S}(f; P_1 \times \dots \times P_n) - \underline{S}(f; P_1 \times \dots \times P_n) \leq \epsilon \operatorname{vol}([\mathbf{a}, \mathbf{b}])$$

for any family (P_1, \dots, P_n) of partitions with mesh size $< \delta/\sqrt{n}$. For this use that $|x_i - y_i| < \delta/\sqrt{n}$ implies

$$|\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} < \sqrt{n(\delta^2/n)} = \delta, \text{ i.e., } \mathbf{y} \in B_\delta(\mathbf{x}).$$

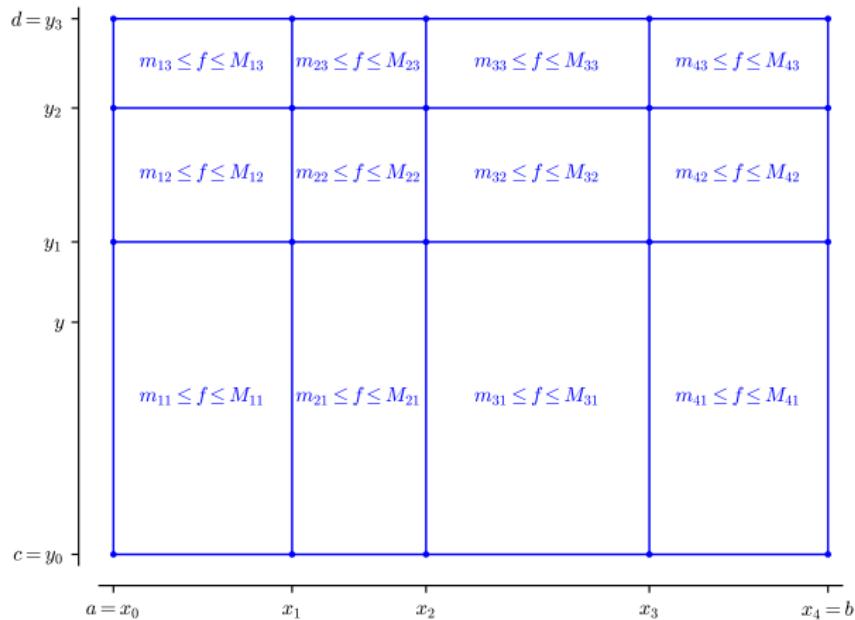


Figure: Illustration of 2-dimensional partitions for bounded functions $f: [a, b] \times [c, d] \rightarrow \mathbb{R}$.

Here $P = \{x_0, x_1, x_2, x_3, x_4\}$, $Q = \{y_0, y_1, y_2, y_3\}$, $P \times Q = \{(x_i, y_j); 0 \leq i \leq 4, 0 \leq j \leq 3\}$, and with $M_{ij} = \sup\{f(x, y); x_{i-1} \leq x \leq x_i, y_{j-1} \leq y \leq y_j\}$, the corresponding upper Darboux sum is $\overline{S}(f; P \times Q) = \sum_{i=1}^4 \sum_{j=1}^3 M_{ij}(x_i - x_{i-1})(y_j - y_{j-1})$; similarly for $\underline{S}(f; P \times Q)$.

Example

We compute the 2-dimensional integral $\int_{[0,1]^2} xy \, d^2(x,y)$ using Darboux's definition.

To this end we decompose $[0, 1]$ into N equally spaced subintervals, i.e., $P: 0 = 0/N < 1/N < 2/N < \dots < N/N = 1$, and use the product partition

$$P \times P = \{(i/N, j/N); 0 \leq i, j \leq N\}$$

of $[0, 1]^2 = \{(x, y) \in \mathbb{R}^2; 0 \leq x \leq 1, 0 \leq y \leq 1\}$.

Any of the N^2 two-dimensional subintervals (squares) determined by $P \times P$ has area $1/N^2$, and the lower and upper Darboux sums of $f(x, y) = xy$ are

$$\underline{S}(f; P \times P) = \frac{1}{N^2} \sum_{i,j=1}^N m_{ij}, \quad \overline{S}(f; P \times P) = \frac{1}{N^2} \sum_{i,j=1}^N M_{ij},$$

where $m_{ij} = \min\left\{xy; \frac{i-1}{N} \leq x \leq \frac{i}{N}, \frac{j-1}{N} \leq y \leq \frac{j}{N}\right\} = \frac{(i-1)(j-1)}{N^2}$ and $M_{ij} = \max\left\{xy; \frac{i-1}{N} \leq x \leq \frac{i}{N}, \frac{j-1}{N} \leq y \leq \frac{j}{N}\right\} = \frac{ij}{N^2}$.

It follows that

$$\begin{aligned}\underline{S}(f; P \times P) &= \frac{1}{N^4} \sum_{i,j=1}^N (i-1)(j-1) = \frac{1}{N^4} \left(\sum_{i=1}^N (i-1) \right)^2 \\ &= \frac{N^2(N-1)^2}{4N^4} = \frac{(N-1)^2}{4N^2},\end{aligned}$$

$$\overline{S}(f; P \times P) = \frac{1}{N^4} \sum_{i,j=1}^N ij = \frac{1}{N^4} \left(\sum_{i=1}^N i \right)^2 = \frac{(N+1)^2}{4N^2}.$$

Since

$$\overline{S}(f; P \times P) - \underline{S}(f; P \times P) = \frac{1}{N} \rightarrow 0 \quad \text{for } N \rightarrow \infty,$$

f is Riemann integrable over $[0, 1]^2$ according to Darboux's definition, and

$$\int_{[0,1]^2} f(x, y) d^2(x, y) = \sup \left\{ \frac{(N-1)^2}{4N^2}; N \in \mathbb{N} \right\} = \lim_{N \rightarrow \infty} \frac{(N-1)^2}{4N^2} = \frac{1}{4}.$$

Notes

- The preceding computation vividly shows the advantage of Darboux's definition over Riemann's. To show that f is integrable, it suffices to find, given $\epsilon > 0$, one product partition $P_1 \times P_2$ that satisfies $\bar{S}(f; P_1 \times P_2) - \underline{S}(f; P_1 \times P_2) < \epsilon$. We are free to choose this partition as nice as possible. In our case the partitions have the form $P \times P$ with the (subdivision) points of P in arithmetic progression; points in geometric progression could also be used (and would facilitate the direct computation of $\int_{[0,1]^2} x^m y^n d^2(x, y)$ for any m, n).
- You may conjecture from the preceding computation that for functions f of the special form $f(x, y) = g(x)h(y)$ we have in general

$$\int_{[a,b] \times [c,d]} f(x, y) d^2(x, y) = \int_a^b g(x) dx \int_c^d h(y) dy,$$

and hence in particular $\int_{[0,1]^2} x^m y^n d^2(x, y) = \frac{1}{(m+1)(n+1)}$. Both conjectures are true.

Notes cont'd

- As its 1-dimensional specialization, the n -dimensional integral is \mathbb{R} -linear, i.e., for integrable functions f_1, f_2 and constants $c_1, c_2 \in \mathbb{R}$ we have

$$\int c_1 f_1(\mathbf{x}) + c_2 f_2(\mathbf{x}) d^n \mathbf{x} = c_1 \int f_1(\mathbf{x}) d^n \mathbf{x} + c_2 \int f_2(\mathbf{x}) d^n \mathbf{x}.$$

(That the domain of integration is missing here, is not an oversight and will be explained soon.)

Together with the preceding note this shows, for example, how to integrate polynomial functions

$$p(x, y) = \sum_{m,n=0}^d p_{mn} x^m y^n$$

over $[0, 1]^2$ or, a little more general, over $[a, b] \times [c, d]$.

Iterated Integrals

Like differentiation of n -variable functions, n -dimensional integration can be reduced to the 1-dimensional case.

The key to this simplification is FUBINI's Theorem, which in its simplest form is the following

Theorem (Little Fubini)

Suppose $f: [a, b] \times [c, d] \rightarrow \mathbb{R}$, $(x, y) \mapsto f(x, y)$ is continuous. For $y \in [c, d]$ define $F(y) = \int_a^b f(x, y) dx$. Then $F: [c, d] \rightarrow \mathbb{R}$ is Riemann integrable and satisfies

$$\int_{[a,b] \times [c,d]} f(x, y) d^2(x, y) = \int_c^d F(y) dy = \int_c^d \left(\int_a^b f(x, y) dx \right) dy.$$

Interchanging the variables x, y we obtain the more general form

$$\int_{[a,b] \times [c,d]} f(x, y) d^2(x, y) = \int_a^b \left(\int_c^d f(x, y) dy \right) dx = \int_c^d \left(\int_a^b f(x, y) dx \right) dy.$$

Proof.

Since continuous one-variable functions are integrable over closed and bounded intervals, and since $[a, b] \rightarrow \mathbb{R}$, $x \mapsto f(x, y)$ is continuous, $\int_a^b f(x, y) dx$ exists for every $y \in [c, d]$, so that F is well defined.

Our goal is to squeeze the Darboux sums for $\int_c^d F(y) dy$ between the Darboux sums for $\int_{[a,b] \times [c,d]} f(x, y) d^2(x, y)$ in the following sense:

For any partitions P of $[a, b]$ and Q of $[c, d]$ we have

$$\underline{S}(f; P \times Q) \leq \underline{S}(F; Q) \leq \overline{S}(F; Q) \leq \overline{S}(f; P \times Q). \quad (*)$$

Since $\int_{[a,b] \times [c,d]} f(x, y) d^2(x, y)$ is known to exist, this implies the theorem.

For the proof of $(*)$ suppose $P = \{x_0, x_1, \dots, x_M\}$, $Q = \{y_0, y_1, \dots, y_N\}$ are partitions of $[a, b]$ and $[c, d]$, respectively, where as usual $x_0 = a$, $x_M = b$, $y_0 = c$, $y_N = d$. Let

$$M_{ij} = \sup\{f(x, y); (x, y) \in [x_{i-1}, x_i] \times [y_{j-1}, y_j]\},$$

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

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

Proof cont'd.

Then we have

$$\underline{S}(f(\cdot, y); P) \equiv \sum_{i=1}^M m_i(y)(x_i - x_{i-1}),$$

$$\overline{S}(f(\cdot, y); P) = \sum_{i=1}^M M_i(y)(x_i - x_{i-1}),$$

$$\underline{S}(f(\cdot, y); P) \leq F(y) \leq \overline{S}(f(\cdot, y); P),$$

since the Darboux sums are those for $\int_a^b f(x, y) dx = F(y)$.
 Further, for $y \in [y_{j-1}, y_j]$ we have $M_i(y) \leq M_{ij}$ and hence

$$\sup\{F(y); y \in [y_{j-1}, y_j]\} \leq \sum_{i=1}^M M_{ij}(x_i - x_{i-1}).$$

$$\begin{aligned} \implies \overline{S}(F; Q) &\leq \sum_{j=1}^N \left(\sum_{i=1}^M M_{ij}(x_i - x_{i-1}) \right) (y_j - y_{j-1}) \\ &= \overline{S}(f; P \times Q). \end{aligned}$$

Similarly, $\underline{S}(F; Q) \geq \underline{S}(f; P \times Q)$, completing the proof. □

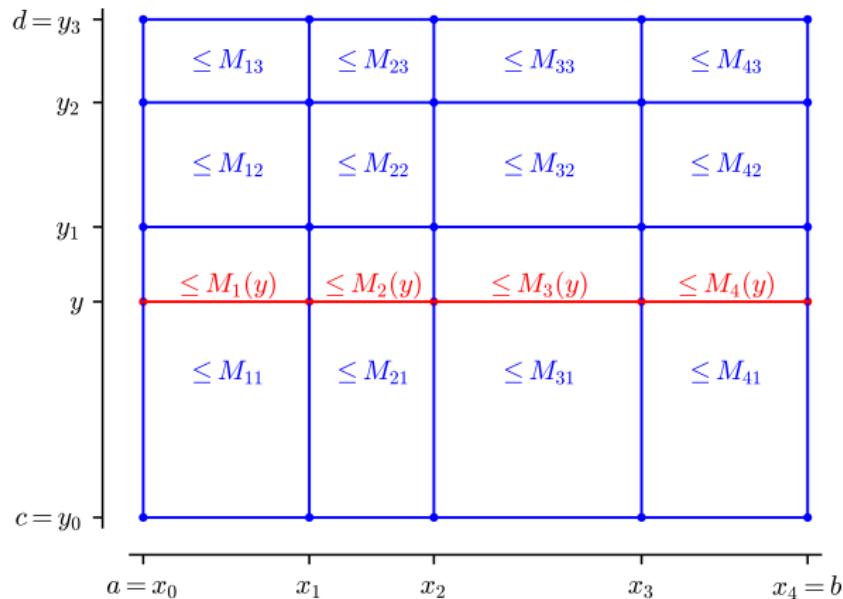


Figure: Illustration of the proof of Fubini's Little Theorem

Here $P = \{x_0, x_1, x_2, x_3, x_4\}$, $Q = \{y_0, y_1, y_2, y_3\}$,
 $P \times Q = \{(x_i, y_j); 0 \leq i \leq 4, 0 \leq j \leq 3\}$,
 $P \times \{y\} = \{(x_0, y), (x_1, y), (x_2, y), (x_3, y), (x_4, y)\}$,
and only the upper bounds for f on the various squares and horizontal line segments are indicated.

Notes

- Continuity of f was only used to prove the existence of $\int_a^b f(x, y) dx$ for $y \in [c, d]$. Hence the conclusion of Fubini's Little Theorem remains valid under the assumptions that f is integrable and all integrals $\int_a^b f(x, y) dx$, $y \in [c, d]$, exist.
- The functions $F: [c, d] \rightarrow \mathbb{R}$, $y \mapsto \int_a^b f(x, y) dx$, and $G: [a, b] \rightarrow \mathbb{R}$, $x \mapsto \int_c^d f(x, y) dy$ appearing in Fubini's Little Theorem are in fact continuous, but this is not at all obvious from their definition. One has, e.g., to bound $F(y) - F(y_0) = \int_a^b f(x, y) - f(x, y_0) dx$, and for this a bound for $|f(x, y) - f(x, y_0)|$ that holds uniformly in x is needed. We will return to such *parameter integrals* later.

Application

As a corollary to Fubini's Little Theorem we now prove the conjecture about integrals of “factorable” functions made earlier.

Corollary

If $f: [a, b] \times [c, d] \rightarrow \mathbb{R}$ factorizes as $f(x, y) = g(x)h(y)$ for some continuous functions $g: [a, b] \rightarrow \mathbb{R}$ and $h: [c, d] \rightarrow \mathbb{R}$, then we have

$$\int_{[a,b] \times [c,d]} f(x, y) d^2(x, y) = \left(\int_a^b g(x) dx \right) \left(\int_c^d h(y) dy \right).$$

This follows easily from Fubini's Little Theorem:

$$\begin{aligned} \int_{[a,b] \times [c,d]} f(x, y) d^2(x, y) &= \int_a^b \left(\int_c^d g(x)h(y) dy \right) dx \\ &= \int_a^b g(x) \left(\int_c^d h(y) dy \right) dx = \left(\int_c^d h(y) dy \right) \int_a^b g(x) dx. \end{aligned}$$

Example

The volume of the unit ball

$$B_1(\mathbf{0}) = \{(x, y, z) \in \mathbb{R}^3; x^2 + y^2 + z^2 \leq 1\}$$

in \mathbb{R}^3 is equal to $2 \times$ the integral of the function $f: [-1, 1]^2 \rightarrow \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} \sqrt{1 - x^2 - y^2} & \text{if } x^2 + y^2 \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Since f is continuous on $[-1, 1]^2$, we can apply the Little Fubini Theorem to compute this volume:

$$\begin{aligned} \text{vol}(B_1(\mathbf{0})) &= 2 \int_{-1}^1 \left(\int_{-1}^1 f(x, y) dx \right) dy \\ &\stackrel{(!)}{=} 2 \int_{-1}^1 \left(\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \sqrt{1-x^2-y^2} dx \right) dy \\ &\stackrel{(!!)}{=} 2 \int_{-1}^1 (1-y^2)\pi/2 dy = \pi \int_{-1}^1 (1-y^2) dy = \frac{4}{3}\pi. \end{aligned}$$

Example (cont'd)

The steps marked by (!) and (!!!) perhaps require further explanation:

- (!) In order to determine the inner integrand $f(\cdot, y)$, we need to solve the equations defining $f(x, y)$ for x :

$$x \mapsto f(x, y) = \begin{cases} \sqrt{1 - x^2 - y^2} & \text{if } -\sqrt{1 - y^2} \leq x \leq \sqrt{1 - y^2}, \\ 0 & \text{otherwise.} \end{cases}$$

This reduces the inner integral to one over the shorter interval $[-\sqrt{1 - y^2}, \sqrt{1 - y^2}]$.

- (!!) The graph of $x \mapsto \sqrt{1 - x^2 - y^2} = \sqrt{1 - y^2 - x^2}$ is a half-circle of radius $r = \sqrt{1 - y^2}$. The value of the inner integral is just the area of the corresponding half-disk, which is $r^2\pi/2 = (1 - y^2)\pi/2$.

Effectively we have computed the volume of the half-ball B by determining the areas $A(B_b)$ of all y -sections

$B_b = B \cap \{y = b\}$, $-1 \leq b \leq 1$, and integrating the resulting function $b \mapsto A(B_b)$ (\rightarrow Principle of CAVALIERI).

Notes

The preceding example also serves to illustrate the computation of integrals $\int_D f(x, y) d^2(x, y)$ for functions f whose domain isn't a rectangle of the form $[a, b] \times [c, d]$ (i.e., with sides parallel to the coordinate axes):

Enclose D into a larger rectangle R of this special form and extend f to R by setting $f(x, y) = 0$ for $(x, y) \in R \setminus D$. Then define $\int_D f(x, y) d^2(x, y) := \int_R f(x, y) d^2(x, y)$.

With this definition (which doesn't depend on the particular choice of R , as can be shown), the integral is still a measure for the volume of the region

$$\{(x, y, z) \in \mathbb{R}^3; (x, y) \in D \wedge 0 \leq z \leq f(x, y)\}$$

under the “true” graph of f , provided f is non-negative.

In the example, which corresponds to

$D = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 1\}$, we were “lucky” that the extended function is still continuous and hence Fubini's Little Theorem can be applied directly. Usually extending f destroys its continuity.

Further Examples

Example

Let Δ be the triangle in \mathbb{R}^2 with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$.

Compute the integral $\int_{\Delta} (y - x^2) d^2(x, y)$.

Here the domain $\Delta = \{(x, y) \in \mathbb{R}^2; x, y \geq 0, x + y \leq 1\}$ of integration is not a rectangle and hence Fubini's Little Theorem cannot be applied directly. As noted, the proper way to cure this problem is to extend $f(x, y) = y - x^2$ to $[0, 1]^2$ by setting $f(x, y) = 0$ for $(x, y) \in [0, 1]^2 \setminus \Delta$. But the extended function f is no longer continuous.

However, f turns out to be (Riemann) integrable, and Fubini's Theorem remains valid in this case, as we have noted, provided that the iterated integrals are defined. This will be apparent from the following computation.

For this we will use Fubini's Little Theorem in the form

$$\int_{\Delta} f(x, y) d^2(x, y) = \int_{x \in [0, 1]} \int_{y \in \Delta_x} f(x, y) dy dx$$

with $\Delta_x = \{y \in \mathbb{R}; (x, y) \in \Delta\}$ (vertical sections or traces of Δ).

Example (cont'd)

We obtain

$$\begin{aligned} \int_{\Delta} y - x^2 d^2(x, y) &= \int_{x=0}^1 \int_{y=0}^{1-x} y - x^2 dy dx = \int_{x=0}^1 \left[\frac{y^2}{2} - x^2 y \right]_{y=0}^{1-x} dx \\ &= \int_{x=0}^1 \frac{1}{2}(1-x)^2 - x^2(1-x) dx = \int_0^1 x^3 - \frac{1}{2}x^2 - x + \frac{1}{2} dx \\ &= \left[\frac{1}{4}x^4 - \frac{1}{6}x^3 - \frac{1}{2}x^2 + \frac{1}{2}x \right]_0^1 = \frac{1}{4} - \frac{1}{6} - \frac{1}{2} + \frac{1}{2} = \frac{1}{12} \end{aligned}$$

We can also integrate the other way round:

$$\begin{aligned} \int_{\Delta} y - x^2 d^2(x, y) &= \int_{y=0}^1 \int_{x=0}^{1-y} y - x^2 dx dy = \int_{y=0}^1 \left[xy - \frac{x^3}{3} \right]_{x=0}^{1-y} dy \\ &= \int_0^1 (1-y)y - \frac{1}{3}(1-y)^3 dy = \int_0^1 \frac{1}{3}y^3 - 2y^2 + 2y - \frac{1}{3} dy \\ &= \left[\frac{1}{12}y^4 - \frac{2}{3}y^3 + y^2 - \frac{1}{3}y \right]_0^1 = \frac{1}{12}. \end{aligned}$$

Riemann's vs
Darboux's
Definition

Continuous
Functions

Problems

Higher-
Dimensional
Riemann
Integrals

Iterated
Integrals

New Problems

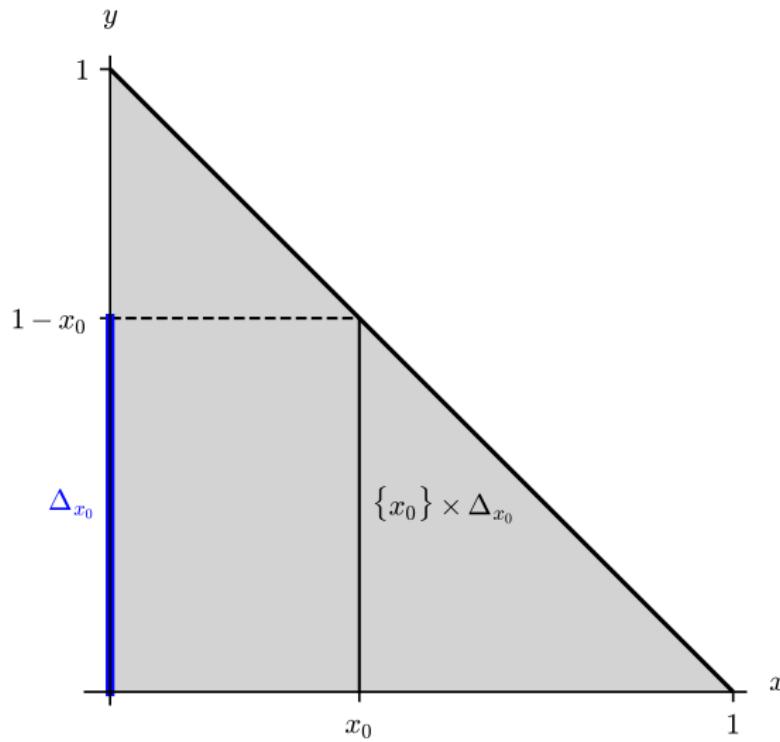


Figure: Integration over Δ

Fubini says $\int_{\Delta} f(x, y) d^2(x, y) = \int_0^1 \left(\int_{y \in \Delta_x} f(x, y) dy \right) dx.$

Example

We compute $\int_D y \, d^2(x, y)$, where D is the bounded region between the two parabolas $y = x^2$ and $y = \frac{1}{2}x^2 + \frac{1}{2}$.

The parabolas intersect in $(-1, 1)$ and $(1, 1)$, and for $-1 \leq x \leq 1$ we have $x^2 \leq \frac{1}{2}x^2 + \frac{1}{2}$. $\Rightarrow D_x = [x^2, \frac{1}{2}x^2 + \frac{1}{2}]$ for $-1 \leq x \leq 1$

$$\begin{aligned}\int_D y \, d^2(x, y) &= \int_{-1}^1 \int_{x^2}^{\frac{1}{2}x^2 + \frac{1}{2}} y \, dy \, dx = \int_{-1}^1 \left[\frac{y^2}{2} \right]_{x^2}^{\frac{1}{2}x^2 + \frac{1}{2}} \, dx \\ &= \frac{1}{2} \int_{-1}^1 \frac{1}{4}(x^2 + 1)^2 - x^4 \, dx = \int_0^1 -\frac{3}{4}x^4 + \frac{1}{2}x^2 + \frac{1}{4} \, dx \\ &= -\frac{3}{20} + \frac{1}{6} + \frac{1}{4} = \frac{4}{15}\end{aligned}$$

Again we interchange the order of integration, both for illustration and for checking correctness of the previous computation.

The computation shows that $\int_D y \, d^2(x, y) = 2 \int_{D^+} y \, d^2(x, y)$, where D^+ denotes the part of D in the half plane $x \geq 0$. This is also clear from the symmetry of D and the integrand $f(x, y) = y$, which satisfies $f(-x, y) = f(x, y)$.

Example (cont'd)

Now we need to calculate $D_y^+ = \{x \in \mathbb{R}; (x, y) \in D\}$. (Note that the same notation is now used for horizontal sections of D^+ , e.g., “ $D_{0.5}^+$ ” has changed its meaning.)

We have the equivalences $y \geq x^2 \iff |x| \leq \sqrt{y}$ and $y \leq \frac{1}{2}x^2 + \frac{1}{2} \iff y \leq \frac{1}{2} \vee |x| \geq \sqrt{2y - 1}$.

$$\Rightarrow D_y^+ = \begin{cases} [0, \sqrt{y}] & \text{for } y \in [0, \frac{1}{2}], \\ [\sqrt{2y - 1}, \sqrt{y}] & \text{for } y \in [\frac{1}{2}, 1]. \end{cases}$$

$$\begin{aligned} \Rightarrow \int_{D^+} y \, d^2(x, y) &= \int_{y=0}^1 \int_{x \in D_y^+} y \, dx \, dy \\ &= \int_0^{1/2} \int_0^{\sqrt{y}} y \, dx \, dy + \int_{1/2}^1 \int_{\sqrt{2y-1}}^{\sqrt{y}} y \, dx \, dy \\ &= \int_0^{1/2} y\sqrt{y} \, dy + \int_{1/2}^1 y \left(\sqrt{y} - \sqrt{2y-1} \right) \, dy \\ &= \left[\frac{2}{5}y^{5/2} \right]_0^1 - \left[\frac{1}{3}y(2y-1)^{3/2} - \frac{1}{15}(2y-1)^{5/2} \right]_{1/2}^1 \\ &= \frac{2}{5} - \frac{1}{3} + \frac{1}{15} = \frac{2}{15}, \quad \text{as expected.} \end{aligned}$$

Question

Why didn't I choose $\int_D x \, d^2(x, y)$ for the computation?

Answer

Because for symmetry reasons this integral is zero:

$g(x, y) = x$ satisfies $g(-x, y) = -g(x, y)$ and D is symmetric with respect to the y -axis.

$$\implies \int_{D^-} x \, d^2(x, y) = - \int_{D^+} x \, d^2(x, y)$$

$$\implies \int_D x \, d^2(x, y) = \int_{D^-} x \, d^2(x, y) + \int_{D^+} x \, d^2(x, y) = 0.$$

Remark

The integrals just computed are related to the centroid of D , which is defined as the point $\mathbf{s} = \frac{1}{\text{vol}_2(D)} (\int_D x \, d^2(x, y), \int_D y \, d^2(x, y))$,

where $\text{vol}_2(D) = \int_D 1 \, d^2(x, y)$ denotes the 2-dimensional volume ("area") of D . More generally, for $A \subseteq \mathbb{R}^n$ the numbers

$s_i = \int_A x_i \, d^n \mathbf{x} / \int_A 1 \, d^n \mathbf{x}$ represent the coordinates of the *centroid* \mathbf{s} of A , provided that the integrals exist.

Riemann's vs

Darboux's
Definition

Continuous
Functions

Problems

Higher-
Dimensional
Riemann
Integrals

Iterated
Integrals

New Problems

Remark (cont'd)

In our example we have

$$\begin{aligned}\text{vol}_2(D) &= \int_{-1}^1 \int_{x^2}^{\frac{1}{2}x^2 + \frac{1}{2}} dy dx = \int_{-1}^1 \frac{1}{2}x^2 + \frac{1}{2} - x^2 dx \\ &= \int_0^1 1 - x^2 dx = \frac{2}{3}\end{aligned}$$

and hence $\mathbf{s} = (0, \frac{4}{15}/\frac{2}{3}) = (0, \frac{2}{5})$.

Since D is bounded by the graphs of two one-variable functions, viz. $g_1(x) = x^2$ and $g_2(x) = \frac{1}{2}x^2 + \frac{1}{2}$, we can also calculate $\text{vol}_2(D)$ using 1-dimensional integrals:

$$\text{vol}_2(D) = \int_{-1}^1 \frac{1}{2}x^2 + \frac{1}{2} dx - \int_{-1}^1 x^2 dx = \dots = \frac{2}{3}.$$

After one step the computations are identical.

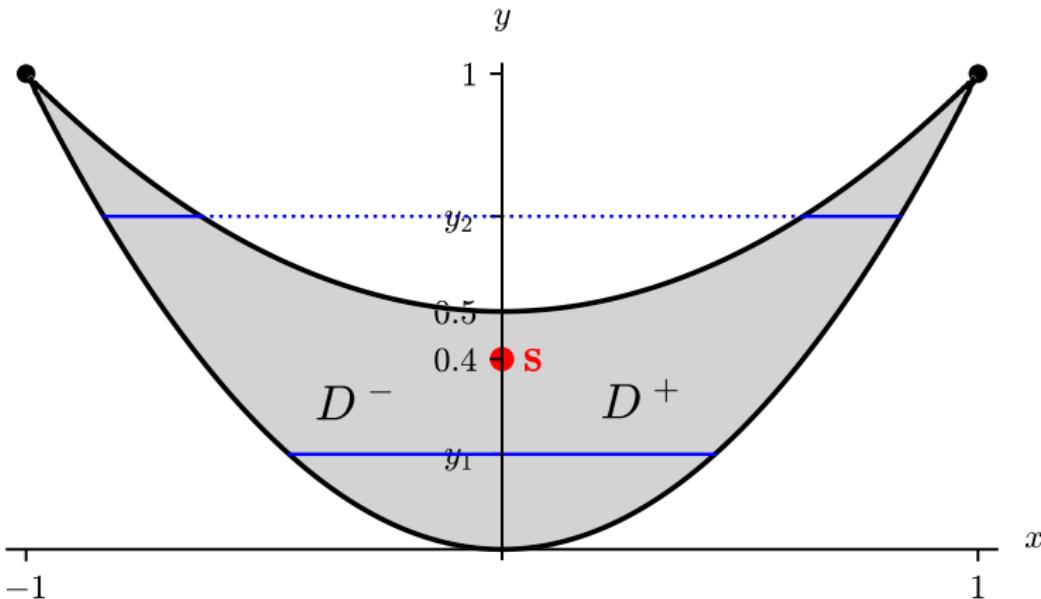


Figure: The region D , its centroid $\mathbf{s} = (0, \frac{2}{5})$, and horizontal sections $D_{y_1} \times \{y_1\}$, $D_{y_2} \times \{y_2\}$ for $y_1 < 0.5 < y_2$

Example

As an example for a triple integral we compute the centroid

$\mathbf{s} = (s_1, s_2, s_3)$ of the half-bowl

$$B^+ = \{(x, y, z) \in \mathbb{R}^3; x^2 + y^2 + z^2 \leq 1, z \geq 0\}.$$

By symmetry we have $s_1 = s_2 = 0$.

For the computation of s_3 we use the fact that for $0 \leq z \leq 1$ the z -sections (z -traces) B_z^+ are disks of radius $\sqrt{1 - z^2}$. This follows by rewriting the first condition as $x^2 + y^2 \leq 1 - z^2$.

The 3-dimensional version of Fubini's Little Theorem gives

$$\begin{aligned} \int_{B^+} z \, d^3(x, y, z) &= \int_{z=0}^1 \int_{(x,y) \in B_z^+} z \, d^2(x, y) \, dz \\ &= \int_0^1 z \operatorname{vol}_2(B_z^+) \, dz = \int_0^1 z(1 - z^2)\pi \, dz = \frac{\pi}{4}. \end{aligned}$$

$$\implies \mathbf{s} = (0, 0, \frac{\pi/4}{2\pi/3}) = (0, 0, \frac{3}{8}).$$

The volume of B^+ can be computed in the same way (compare this with our earlier “2-dimensional” computation):

$$\operatorname{vol}_3(B^+) = \int_{B^+} d^3(x, y, z) = \int_0^1 \operatorname{vol}_2(B_z^+) \, dz = \int_0^1 (1 - z^2)\pi \, dz = \frac{2\pi}{3}.$$

Why is $f(x, y) = y - x^2$ integrable over Δ ?

Answer: The extended function $f: [0, 1]^2 \rightarrow \mathbb{R}$,

$$f(x, y) = \begin{cases} y - x^2 & \text{if } (x, y) \in \Delta, \\ 0 & \text{if } (x, y) \in [0, 1]^2 \setminus \Delta \end{cases}$$

is continuous except for a set of points in $[0, 1]^2$ that is negligible for integration.

More precisely, the set of discontinuities of f is contained in the line $y = 1 - x$. Now consider “grid” partitions $P \times P$ of $[0, 1]^2$ into N^2 small squares of side length $1/N$. For sufficiently large N the contribution of a small square \square to $\bar{S}(f; P \times P) - \underline{S}(f; P \times P)$ is

- ① $\leq \epsilon/N^2$ if \square is contained in Δ (since f is uniformly continuous on Δ);
- ② $= 0$ if \square is disjoint from Δ ;
- ③ $\leq 2/N^2$ if \square meets the line $y = 1 - x$

Since there are only N squares of Type 3, we obtain

$$\bar{S}(f; P \times P) - \underline{S}(f; P \times P) \leq N^2(\epsilon/N^2) + N(2/N^2) = \epsilon + 2/N,$$

which can be made arbitrarily small.

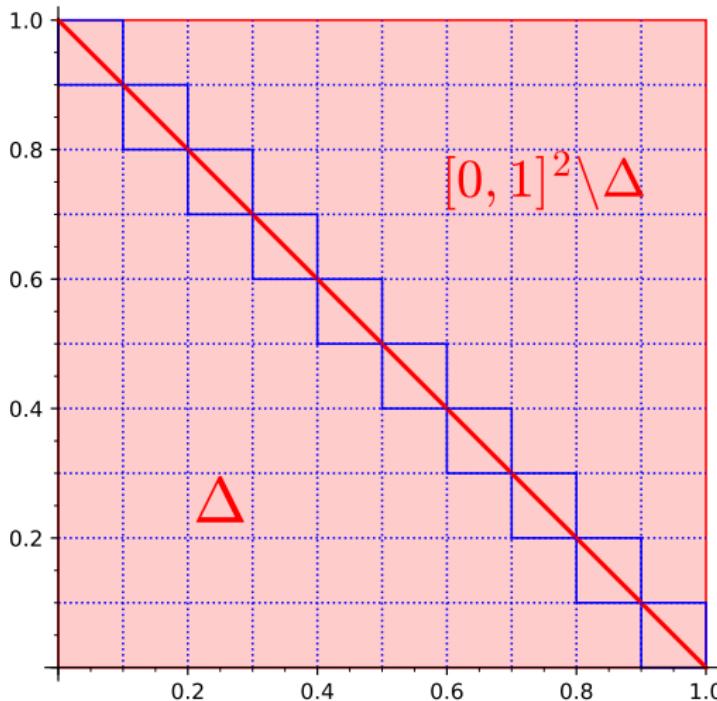


Figure: For integration over $[0, 1]^2$ the diagonal is negligible

Reason: For any $N \in \mathbb{N}$, the diagonal can be covered by N squares of total area $1/N$ (in the picture $N = 10$).
 \Rightarrow The diagonal has 2-dimensional volume (area) zero.

Further Problems with the Riemann Integral

The n -dimensional Riemann integral, as we have defined it, poses further questions:

- 1 We would like to integrate functions $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, say, over domains, which aren't rectangles $[a, b] \times [c, d]$. For example, when computing $\text{vol}(B_1(\mathbf{0}))$ it would be much more natural to integrate $f(x, y) = \sqrt{1 - x^2 - y^2}$ over the unit disk $B = B_1(0, 0) = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 1\}$.
- 2 Following $\text{vol}([b, a]) = b - a = \int_a^b 1 \, dx$, it is natural to define the volume of a set $A \subset \mathbb{R}^n$ as

$$\text{vol}(A) = \int_A 1 \, d^n \mathbf{x}, \quad \text{cf. Part (1).}$$

Such integrals should be defined for a large class of subsets $A \subseteq \mathbb{R}^n$, not only for intervals (cuboids).

- ③ Volumes should be invariant under Euclidian motions. In particular, for an n -dimensional interval $I = [\mathbf{a}, \mathbf{b}]$, an orthogonal matrix $\mathbf{A} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^n$ we should have, accepting the previous definition,

$$\int_{\mathbf{y} \in \mathbf{A} \cdot I + \mathbf{b}} 1 d^n \mathbf{y} = \text{vol}(\mathbf{A} \cdot I + \mathbf{b}) = \text{vol}(I) = \int_{\mathbf{x} \in I} 1 d^n \mathbf{x},$$

which equals $\prod_{i=1}^n (b_i - a_i)$. However, the image $\mathbf{A} \cdot I + \mathbf{b}$ can be a cuboid whose edges are not parallel to the coordinate axes, and hence our limited concept of integration ("integrate over rectangles") doesn't apply to the left-hand integral.

- ④ A generalization of The Little Fubini Theorem to not necessarily continuous integrable functions is needed. When using the Riemann integral, a serious obstruction to this generalization is the fact that, e.g., the values of $[a, b] \rightarrow \mathbb{R}$, $x \mapsto f(x, y)$ for one particular $y \in [c, d]$ do not matter at all for the integrability (and the integral) of f . Hence

$$F(y) = \int_a^b f(x, y) dx \text{ need not be defined for all } y \in [c, d].$$

At this point we cannot cure all 4 problems, but we can eliminate (formally at least) the "integration over complicated domains" problem. This is discussed in the next set of slides.

Math 241
Calculus III

Thomas
Honold

Integration
over \mathbb{R}^n

The Lebesgue
Integral

Lebesgue
Measure

Four Main
Theorems of
Lebesgue
Integration

Monotone
Convergence
Bounded
Convergence
Fubini's Theorem
Change of Variables

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Integration over \mathbb{R}^n

2 The Lebesgue Integral

3 Lebesgue Measure

4 Four Main Theorems of Lebesgue Integration

Monotone Convergence

Bounded Convergence

Fubini's Theorem

Change of Variables

Today's Lecture: Introduction to Lebesgue Integration

Characteristic Functions and Restriction

Definition

The *characteristic function* of a subset $A \subseteq \mathbb{R}^n$ is the function $\chi_A: \mathbb{R}^n \rightarrow \mathbb{R}$ defined by

$$\chi_A(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in A, \\ 0 & \text{if } \mathbf{x} \notin A. \end{cases}$$

Note that characteristic functions are $\{0, 1\}$ -valued, and that any $\{0, 1\}$ -valued function on \mathbb{R}^n is the characteristic function of a unique set $A \subseteq \mathbb{R}^n$.

Examples

The characteristic functions of \emptyset and \mathbb{R}^n , respectively, are the constants 0 and 1 (with domain \mathbb{R}^n and codomain \mathbb{R}). The Dirichlet function $f: [0, 1] \rightarrow \mathbb{R}$ defined by $f(x) = 1$ if $x \in \mathbb{Q}$ and $f(x) = 0$ if $x \notin \mathbb{Q}$ is not a characteristic function, because its domain is not equal to \mathbb{R} . The characteristic function $\chi_{\mathbb{Q}}$ of \mathbb{Q} has domain \mathbb{R} , but otherwise is defined in the same way as f .

Definition

- ① Suppose $A \subseteq \mathbb{R}^n$ and $f: \mathbb{R}^n \rightarrow \mathbb{R}$ (i.e., f is defined on the whole space \mathbb{R}^n). Then the integral of f over A is defined in terms of an integral over \mathbb{R}^n as

$$\int_A f(\mathbf{x}) d^n \mathbf{x} = \int_{\mathbb{R}^n} f(\mathbf{x}) \chi_A(\mathbf{x}) d^n \mathbf{x}.$$

In other words, we integrate the function that coincides with f on A and is zero outside A .

- ② Suppose $A \subseteq B \subseteq \mathbb{R}^n$ and $f: B \rightarrow \mathbb{R}$ (i.e., the domain of f includes A). Then the integral of f over A is defined in the same way as in (1), except that f is replaced by $F: \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto f(x)$ for $x \in B$ and $x \mapsto 0$ for $x \notin B$ (the so-called *trivial extension* of f to \mathbb{R}^n); so again we integrate the function that coincides with f on A and is zero outside A .

Notes

- The preceding definition reduces integration over subsets of \mathbb{R}^n to integration over \mathbb{R}^n and simplifies the introduction of an integral for multivariable functions considerably.

Notes cont'd

- Even if f is integrable over \mathbb{R}^n , the integrals $\int_A f(\mathbf{x}) d^n \mathbf{x}$ generally do not exist (since $f \chi_A$ may not be integrable over \mathbb{R}^n). But for “well-behaved” sets A (so-called measurable sets) the integrals do exist.

The Lebesgue Integral

Modern integration theory is largely based on the *Lebesgue Integral*, introduced by H. LEBESGUE (1875–1941) in 1901.

There are essentially two different, but equivalent approaches to the Lebesgue Integral:

- ① *The measure-theoretic approach:* Define Lebesgue measure for a large class of subsets of \mathbb{R}^n and use this to define the Lebesgue integral for so-called measurable functions.
- ② *The approach using the L^1 -seminorm:* Define an integral for step functions (similar to Riemann sums) and extend the definition “continuously” to functions that can be approximated by step functions in the L^1 -seminorm.

We will now briefly (without any proofs) discuss the 2nd approach, which (in my opinion) is easier to understand when encountering multi-variable integrals for the first time.

Step Functions

Definition (n -dimensional interval)

A set $Q \subset \mathbb{R}^n$ is said to be an (n -dimensional) interval if $Q = I_1 \times \cdots \times I_n$ is the cartesian product of bounded, non-empty intervals in \mathbb{R} .

Thus a factor I_i may have one of the forms $[a, b]$, $(a, b]$, $[a, b)$, (a, b) with $a, b \in \mathbb{R}$, $a < b$. For closed intervals $[a, b]$ we also admit $a = b$, i.e., $[a, a] = \{a\}$.

Definition (step function)

A function $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ is called a *step function* if there exist finitely many n -dimensional intervals Q_1, \dots, Q_r and constants $c_1, \dots, c_r \in \mathbb{R}$ such that

$$\phi(\mathbf{x}) = \sum_{i=1}^r c_i \chi_{Q_i}(\mathbf{x}) \quad \text{for } \mathbf{x} \in \mathbb{R}^n.$$

In Linear Algebra terms, a step function $\phi = \sum_{i=1}^r c_i \chi_{Q_i}$ is just an element in the linear span of the characteristic functions of n -dimensional intervals.

Integral for Step Functions

Definition (volume of an n -dimensional interval)

The volume of an n -dimensional interval $Q = I_1 \times \cdots \times I_n$ is defined as $\text{vol}(Q) = \prod_{i=1}^n (b_i - a_i)$, where $I_i = [a_i, b_i]$, $[a_i, b_i)$, $(a_i, b_i]$, or (a_i, b_i) .

In particular we have $\text{vol}(Q) = 0$ iff at least one interval I_i reduces to a single point.

Definition (integral for step functions)

The integral of a step function $\phi = \sum_{i=1}^r c_i \chi_{Q_i}$ is defined as

$$\int_{\mathbb{R}^n} \phi(\mathbf{x}) d^n \mathbf{x} = \sum_{i=1}^r c_i \text{vol}(Q_i). \quad (\text{ISF})$$

Other notations in use for the integral of a step function (and later for the Lebesgue integral in general) are $\int_{\mathbb{R}^n} \phi(\mathbf{x}) d^n(x_1, \dots, x_n)$, $\int_{\mathbb{R}^n} \phi$ or just $\int \phi$. (ISF) is equivalent to $\int \chi_Q = \text{vol}(Q)$ for all n -dimensional intervals Q and the linearity property
 $\int (c_1 \phi_1 + c_2 \phi_2) = c_1 \int \phi_1 + c_2 \int \phi_2$.

Problem

Is (ISF) a valid definition?

This problem arises, since a step function usually has many representations of the form $\sum_{i=1}^r c_i \chi_{Q_i}$.

Theorem

- 1 Every step function ϕ has a representation $\phi = \sum_{i=1}^r c_i \chi_{Q_i}$ with mutually disjoint n -dimensional intervals Q_i . For such a representation we have $\phi(\mathbf{x}) = c_i$ if $\mathbf{x} \in Q_i$ and $\phi(\mathbf{x}) = 0$ if $\mathbf{x} \in \mathbb{R}^n \setminus (Q_1 \cup \dots \cup Q_r)$. In particular the range of ϕ is $\{0, c_1, \dots, c_r\}$.
- 2 If $\phi = \sum_{i=1}^r c_i \chi_{Q_i} = \sum_{j=1}^s c'_j \chi_{Q'_j}$ are two representations of ϕ as a linear combination of characteristic functions of n -dimensional intervals, then

$$\sum_{i=1}^r c_i \text{vol}(Q_i) = \sum_{j=1}^s c'_j \text{vol}(Q'_j).$$

Hence the integral for step functions is well-defined.

Proof.

Omitted (but see the subsequent example). The main assertion is (2). For the proof of (2) it is convenient to use (1). □

Example

It is easy to see that a function $\phi: \mathbb{R} \rightarrow \mathbb{R}$ is a step function iff there exist real numbers $x_0 < x_1 < \dots < x_N$ such that $\phi(x) = 0$ for $x \notin [x_0, x_N]$ and the restriction of ϕ to any open subinterval (x_{i-1}, x_i) is a constant $c_i \in \mathbb{R}$.

Setting $Q_i = (x_{i-1}, x_i)$ for $1 \leq i \leq N$, $Q_{N+1} = \{x_0\}$, $Q_{N+2} = \{x_1\}$, \dots , $Q_{2N+1} = \{x_N\}$, we have

$$\phi = \sum_{i=1}^N c_i \chi_{Q_i} + \sum_{i=0}^N \phi(x_i) \chi_{Q_{N+1+i}},$$

and the intervals Q_1, \dots, Q_{2N+1} are mutually disjoint.

More specifically, let $Q_1 = [0, 2]$, $Q_2 = [1, 4]$, and $\phi = 2\chi_{Q_1} + \chi_{Q_2}$. Then

$$\phi(x) = \begin{cases} 2 & \text{if } 0 \leq x < 1, \\ 3 & \text{if } 1 \leq x \leq 2, \\ 1 & \text{if } 2 < x \leq 4, \\ 0 & \text{if } x < 0 \vee x > 4, \end{cases}$$

and a representation of ϕ with mutually disjoint intervals is

Example (cont'd)

$$\phi = 2 \chi_{[0,1]} + 3 \chi_{[1,2]} + \chi_{(2,4]}.$$

In sync with the theorem both decompositions yield the same value for the integral of ϕ :

$$\begin{aligned} 2 \text{vol}(Q_1) + \text{vol}(Q_2) &= 2 \text{vol}([0, 2]) + \text{vol}([1, 4]) \\ &= 2(\text{vol}([0, 1]) + \text{vol}([1, 2])) + \text{vol}([1, 2]) + \text{vol}((2, 4]) \\ &= 2 \text{vol}([0, 1]) + 3 \text{vol}([1, 2]) + \text{vol}((2, 4]). \end{aligned}$$

The second sum is over the range of ϕ (excluding zero) with coefficients equal to the volume of the corresponding preimage.

Remark

A function $f: [a, b] \rightarrow \mathbb{R}$ is Riemann integrable iff for every $\epsilon > 0$ there are step functions ϕ, ψ such that $\phi(x) \leq f(x) \leq \psi(x)$ for $x \in [a, b]$, $\phi(x) = \psi(x) = 0$ for $x \notin [a, b]$ and $\int (\psi - \phi) < \epsilon$.

This follows from the fact (easy to verify) that for an appropriately chosen partition P of $[a, b]$ the lower and upper Darboux sums $\underline{S}(f; P), \bar{S}(f; P)$ are squeezed between $\int \phi$ and $\int \psi$.

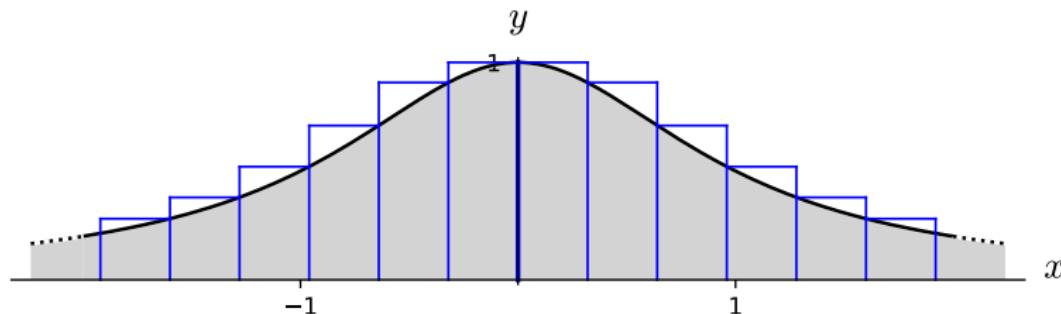


Figure: The ordinate set of $f(x) = \frac{1}{1+x^2}$, $x \in \mathbb{R}$ cannot be covered by finitely many rectangles. But with **countably many rectangles** we can do it.

Afternote

The picture doesn't show a covering by an **enveloping series** as defined on the following slide, because the intervals involved are closed, and removing the boundary would destroy the covering property. But we can remedy this by choosing slightly **overlapping intervals** (and rectangles). Since the overlap can be made arbitrarily small, the L^1 -seminorm isn't affected if we allow more general intervals in the definition of "enveloping series". The restriction to open intervals is only a technical assumption that simplifies the proofs of theorems about the Lebesgue integral.

The L^1 -seminorm

The L^1 -seminorm provides an outer measure for the ordinate set

$$O_f = \{\mathbf{x} \in \mathbb{R}^{n+1}; 0 \leq x_{n+1} \leq f(x_1, \dots, x_n)\}$$

of a function $f \geq 0$ (i.e., $f(x_1, \dots, x_n) \geq 0$ for all $(x_1, \dots, x_n) \in \mathbb{R}^n$) in much the same way as the upper Darboux integral, except that we permit infinitely many covering intervals $Q_i \times [0, c_i]$ and hence will be able to cover unbounded ordinate sets.

Definition (arithmetic in $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$)

We extend the usual arithmetic operations on \mathbb{R} by

$$+\infty + a = +\infty \quad \text{for } a \in \mathbb{R} \cup \{+\infty\},$$

$$-\infty + a = -\infty \quad \text{for } a \in \mathbb{R} \cup \{-\infty\},$$

$$+\infty \cdot 0 = 0,$$

$$+\infty \cdot a = \begin{cases} +\infty & \text{for } a \in \mathbb{R}^+ \cup \{+\infty\}, \\ -\infty & \text{for } a \in \mathbb{R}^- \cup \{-\infty\}, \end{cases}$$

$(-\infty)(-\infty) = +\infty$, and commutativity. (Note that $+\infty + (-\infty)$ remains undefined.) As usual, the ordering in $\bar{\mathbb{R}}$ is given by $-\infty < a < +\infty$ for $a \in \mathbb{R}$.

Definition (enveloping series)

A real (function) series $\Phi = \sum_{i=1}^{\infty} c_i \chi_{Q_i}$, is called an *enveloping series* for $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ if

- ① Each Q_i is an open n -dimensional interval and $c_i \geq 0$;
- ② $f(\mathbf{x}) \leq \Phi(\mathbf{x}) = \sum_{i=1}^{\infty} c_i \chi_{Q_i}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ (an inequality in $\bar{\mathbb{R}}$).

The *content* of Φ is defined as $I(\Phi) = \sum_{i=1}^{\infty} c_i \text{vol}(Q_i) \in [0, +\infty]$.

Definition (L^1 -seminorm)

The L^1 -seminorm of $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ is defined as

$$\|f\|_1 = \inf \{I(\Phi); \Phi \text{ is an enveloping series for } |f|\}.$$

Notes

- $\|f\|_1$ is well-defined since the series $\Phi = \sum_{i=1}^{\infty} \chi_{Q_i}$ with $Q_i = (-i, i)^n$ satisfies $\Phi(\mathbf{x}) = +\infty$ for all $\mathbf{x} \in \mathbb{R}^n$ and hence forms an enveloping series for f .
- $f \mapsto \|f\|_1$ satisfies all the usual properties of a distance function ("norm") except that $\|f\|_1 = +\infty$ is possible and $\|f\|_1 = 0$ does not imply $f = 0$ (i.e., $f(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathbb{R}^n$).

Definition

$f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ is said to be *Lebesgue integrable* (or *integrable*, for short) if there exists a sequence (ϕ_k) of step functions satisfying $\lim_{k \rightarrow \infty} \|f - \phi_k\|_1 = 0$. If this is the case then the (*Lebesgue*) *integral* of f is defined as

$$\int_{\mathbb{R}^n} f(\mathbf{x}) d^n \mathbf{x} = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} \phi_k(\mathbf{x}) d^n \mathbf{x} \quad (\text{which is in } \mathbb{R}).$$

Notes

- The limit is well-defined, since for $k, l \in \mathbb{N}$ the triangle inequality for $\|\cdot\|_1$ and $\|\phi\|_1 = \int |\phi|$ for step functions ϕ (another fact that requires some effort to prove) give

$$\begin{aligned} \left| \int \phi_k - \int \phi_l \right| &\leq \int |\phi_k - \phi_l| = \|\phi_k - \phi_l\|_1 \\ &\leq \|f - \phi_k\|_1 + \|f - \phi_l\|_1. \end{aligned}$$

This shows that $(\int \phi_k)_{k \in \mathbb{N}}$ is a Cauchy sequence and hence convergent in \mathbb{R} .

The Cauchy Test for Convergence

The most important convergence criterion

Definition

A sequence (a_n) of real numbers is said to be a *Cauchy sequence* if for every $\epsilon > 0$ there exists a response $N \in \mathbb{N}$ such that

$$|a_m - a_n| < \epsilon \quad \text{whenever } m, n > N.$$

It is easy to see that $\lim_{n \rightarrow \infty} a_n = a \in \mathbb{R}$ implies that (a_n) is a Cauchy sequence. (For this use the estimate $|a_m - a_n| \leq |a_m - a| + |a_n - a|$.) Conversely we have:

Theorem

Every Cauchy sequence converges in \mathbb{R} .

Proof.

Denoting the response to $\epsilon = 1$ by N , we have $-1 < a_m - a_{N+1} < 1$ for all $m > N$, which shows

$$\min\{a_1, \dots, a_N, a_{N+1} - 1\} \leq a_n \leq \max\{a_1, \dots, a_N, a_{N+1} + 1\}$$

for all n .

Proof cont'd.

\Rightarrow The sequence (a_n) is bounded.

$\Rightarrow (a_n)$ has a convergent subsequence a_{n_1}, a_{n_2}, \dots (by the Bolzano-Weierstrass Theorem).

Denote the limit of the subsequence by a and the response to $\epsilon/2$ in the Cauchy test for (a_n) by N . Then there exists k such that $n_k > N$ and $|a_{n_k} - a| < \epsilon/2$.

$$\Rightarrow |a_m - a| \leq |a_m - a_{n_k}| + |a_{n_k} - a| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $m > N$, i.e., N serves as response to ϵ in a proof of the convergence $a_n \rightarrow a$. □

Remarks

For a series $\sum_{n=0}^{\infty} a_n$ the Cauchy test requires to find N such that $|\sum_{k=m}^n a_k| < \epsilon$ for all $n \geq m > N$. Since $|\sum_{k=m}^n a_k| \leq \sum_{k=m}^n |a_k|$, the convergence of $\sum_{n=0}^{\infty} |a_n|$ (*absolute convergence*) implies the convergence of $\sum_{n=0}^{\infty} a_n$.

There are analogous Cauchy tests for limits of sequences in \mathbb{R}^n and for limits of functions $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x})$ with $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, $\mathbf{x}_0 \in D'$.

Notes on the Lebesgue integral cont'd

- If (ϕ_k) and (ψ_k) are sequences of step functions with $\lim_{k \rightarrow \infty} \|f - \phi_k\|_1 = \lim_{k \rightarrow \infty} \|f - \psi_k\|_1 = 0$ then

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} \phi_k(\mathbf{x}) d^n \mathbf{x} = \lim_{k \rightarrow \infty} \int_{\mathbb{R}^n} \psi_k(\mathbf{x}) d^n \mathbf{x},$$

and hence the Lebesgue integral of f is well-defined. This follows from $|\int \phi_k - \int \psi_k| \leq \|f - \phi_k\|_1 + \|f - \psi_k\|_1$; cp. the previous note.

- The Lebesgue integral of a step function $\phi = \sum_{i=1}^r c_i \chi_{Q_i}$ according to the new definition coincides with the earlier defined integral $\int \phi = \sum_{i=1}^r c_i \text{vol}(Q_i)$.
- Every Riemann integrable function $f: [a, b] \rightarrow \mathbb{R}$ (and similarly for functions defined on n -dimensional intervals) is also Lebesgue integrable over $[a, b]$, i.e., the trivial extension of f is Lebesgue integrable over \mathbb{R} , and both integrals have the same value.

Example

We show that the Dirichlet function $f: [0, 1] \rightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \notin \mathbb{Q}, \end{cases}$$

is Lebesgue integrable over $[0, 1]$ with $\int_0^1 f(x) dx = 0$.

According to the definition of Lebesgue integrals over subsets of \mathbb{R}^1 this is equivalent to the corresponding statements for the trivial extension $F: \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$F(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \cap [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

(F is the characteristic function of $\mathbb{Q} \cap [0, 1] \subset \mathbb{R}$.)

Given an enumeration q_1, q_2, \dots of the rational numbers in $[0, 1]$ and $\epsilon > 0$, let $I_k \subset \mathbb{R}$ be the open interval of length $\epsilon 2^{-k}$ centered at q_k , and consider the series $\Phi = \sum_{k=1}^{\infty} \chi_{I_k}$.

Example (cont'd)

For $k \in \mathbb{N}$ we have

$$\Phi(q_k) = \chi_{I_k}(q_k) + \sum_{l \neq k} \chi_{I_l}(q_k) \geq 1 = F(q_k)$$

and elsewhere $\Phi(x) \geq F(x) = 0$, so that Φ is an enveloping series for F .

$$I(\Phi) = \sum_{k=1}^{\infty} \text{vol}(I_k) = \sum_{k=1}^{\infty} \epsilon 2^{-k} = \epsilon$$

This shows that $\|F\|_1 = 0$ and hence that F is Lebesgue integrable with $\int_{\mathbb{R}} F(x) dx = 0$ (take $\phi_k = 0$ in the definition).

The key point is that $I(\Phi)$ can be made arbitrary small. In the picture on the following slide the first interval has length 2ϵ in place of $\epsilon/2$, but otherwise the construction is the same.

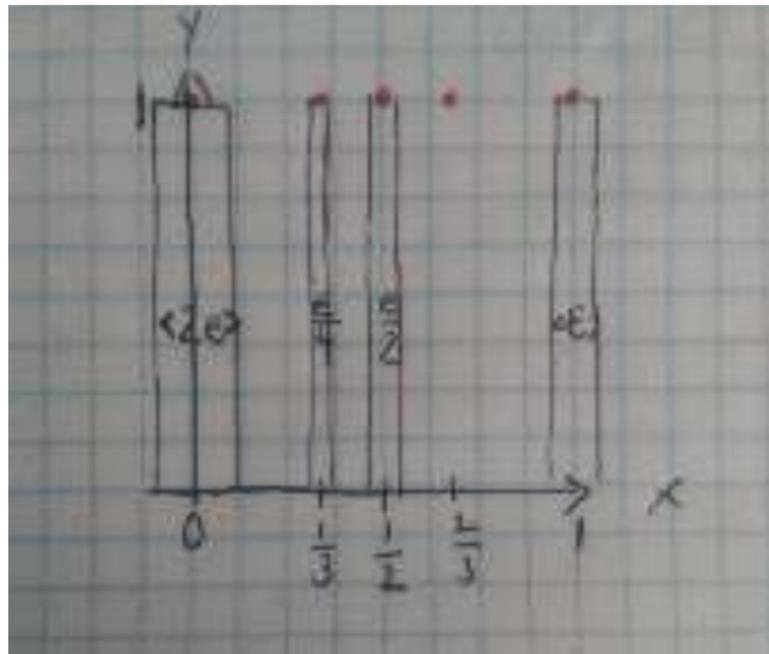


Figure: An enveloping series for Dirichlet's function

Here $\mathbb{Q} \cap [0, 1]$ is enumerated as $0, 1, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{1}{6}, \frac{5}{6}, \dots$; the 1st interval I_1 has width 2ϵ , the 2nd interval I_2 width ϵ , the 3rd interval I_3 width $\epsilon/2$, etc., so that $I(\Phi) = 4\epsilon$.

Elementary Properties of the Lebesgue Integral

- ① If f is integrable then $|f|$ is integrable; if applicable then

$$\left| \int f \right| \leq \int |f| = \|f\|_1.$$

- ② If f_1, f_2 are integrable then so are $c_1 f_1 + c_2 f_2$ for $c_1, c_2 \in \mathbb{R}$, and

$$\int (c_1 f_1 + c_2 f_2) = c_1 \int f_1 + c_2 \int f_2. \quad (\text{Linearity})$$

- ③ If $f_1 \leq f_2$ are integrable then $\int f_1 \leq \int f_2$. *(Monotonicity)*
- ④ If f_1, f_2 are integrable and one of f_1, f_2 is bounded then so is $f_1 f_2$.
- ⑤ If f_1, f_2 are integrable then so are $\max(f_1, f_2)$ and $\min(f_1, f_2)$.

The last property implies that with f also $f^+ = \max(f, 0)$ and $f^- = -\min(f, 0)$ are integrable.

Note that $f^+, f^- \geq 0$, $f = f^+ - f^-$, $|f| = f^+ + f^-$.

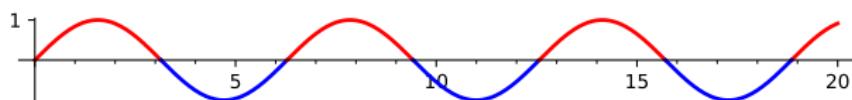
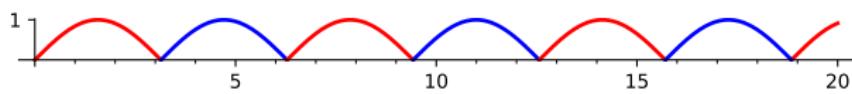
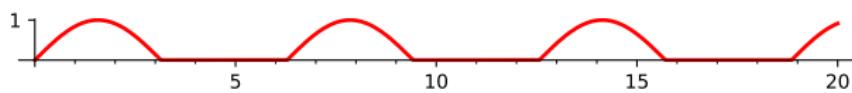
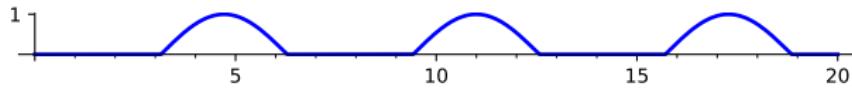
(a) f (b) $|f|$ (c) f^+ (d) f^-

Figure: The representations $f(x) = f^+(x) - f^-(x)$,
 $|f(x)| = f^+(x) + f^-(x)$ for $f(x) = \sin x$

Remarks

- 1 In an earlier version of the slides it was stated that f is integrable iff $|f|$ is integrable. This is not true, because, e.g., there exist sets $A \subseteq [0, 1]$ such that χ_A is not integrable over \mathbb{R} . (Such sets are said to be non-measurable; see the subsequent section on the Lebesgue measure.) Letting $B = [0, 1] \setminus A$ and $f = \chi_A - \chi_B$, we have $f^+ = \chi_A$, $f^- = \chi_B$, and $|f| = \chi_{[0,1]}$. Thus $|f|$ is integrable, but f , f^+ , f^- are not integrable.

In other words, the integrability of f implies the integrability of $|f|$, but not conversely.

It is true, however, that f is integrable iff both f^+ and f^- are integrable; cf. the note on the previous slide. But if neither f^+ nor f^- are integrable, $|f|$ may still be integrable.

The representation $f = f^+ - f^-$ can often be used to infer a property of the Lebesgue integral of an arbitrary integrable function from that of a non-negative integrable function (since $f^+, f^- \geq 0$). An example of this is the subsequent proof of the corollary to Beppo Levi's Theorem ("improper Lebesgue integration" or "integration by exhaustion").

Remarks (cont'd)

- 2 As an example for why the assumption “one of f_1, f_2 is bounded” in Property 4 is needed, take $f_1 = f_2 = f$ with

$$f(x) = \begin{cases} \frac{1}{\sqrt{x}} & \text{if } x \in (0, 1), \\ 0 & \text{if } x \in \mathbb{R} \setminus (0, 1). \end{cases}$$

As discussed on the next few slides, the function f is Lebesgue integrable, since $f \geq 0$ and the improper Riemann integral $\int_0^1 \frac{dx}{\sqrt{x}}$ exists. But, since $f(x)^2 = 1/x$ for $x \in (0, 1)$ and $\int_0^1 \frac{dx}{x} = +\infty$, the product $f_1 f_2 = f^2$ is not Lebesgue integrable.

Improper Riemann Integrals

We state the precise relation between improper 1-dimensional Riemann integrals and the Lebesgue integral only for the interval $I = \mathbb{R} = (-\infty, +\infty)$. For other types of intervals the statement holds mutatis mutandis.

Theorem

Suppose $f: \mathbb{R} \rightarrow \mathbb{R}$ is Riemann integrable over each closed and bounded interval $[a, b] \subset \mathbb{R}$. Then f is Lebesgue integrable iff the improper Riemann integral

$$\int_{-\infty}^{+\infty} f(x) dx = \lim_{\substack{a \rightarrow -\infty \\ b \rightarrow +\infty}} \int_a^b f(x) dx$$

converges absolutely. If applicable, both integrals have the same value.

Examples

- ① The function $f(x) = 1/x^2$ is Lebesgue integrable over $[1, +\infty]$ with

$$\int_{[1, +\infty]} f(x) dx = \int_1^{+\infty} \frac{dx}{x^2} = \left[\frac{-1}{x} \right]_1^{+\infty} = 1,$$

but not integrable over $[0, 1]$ since $\int_0^1 \frac{dx}{x^2} = +\infty$.

Likewise $\frac{\sin x}{x^2}$ is integrable over $[1, +\infty]$, $1/\sqrt{x}$ is integrable over $[0, 1]$, etc.

- ② $f(x) = \frac{\sin x}{x}$ is not Lebesgue integrable over \mathbb{R} , because the improper Riemann integral

$$\int_{-\infty}^{+\infty} \frac{\sin x}{x} dx,$$

which exists, does not converge absolutely (i.e., $\int_{-\infty}^{+\infty} \left| \frac{\sin x}{x} \right| dx$ does not exist).

Lebesgue Measure

In the L^1 -seminorm approach to the Lebesgue integral, Lebesgue measure on \mathbb{R}^n is defined in terms of the integral as follows:

Definition

A subset $A \subseteq \mathbb{R}^n$ is said to be *Lebesgue measurable* (or *measurable*, for short) if the characteristic function χ_A is integrable. If this is the case,

$$\text{vol}(A) = \int_{\mathbb{R}^n} \chi_A(\mathbf{x}) d^n \mathbf{x} = \int_A 1 d^n \mathbf{x} \in \mathbb{R}$$

is called the (n -dimensional) *measure* or *volume* of A .

Properties of the Lebesgue Measure

Obvious properties such as “ $\text{vol}(A) \geq 0$ ”, “ $A \subseteq B$ implies $\text{vol}(A) \leq \text{vol}(B)$ ” or that the volume of n -dimensional intervals remains the same w.r.t. Lebesgue measure are not listed.

- 1 If A_1, A_2, A_3, \dots is a sequence of measurable sets in \mathbb{R}^n then $\bigcap_{i=1}^{\infty} A_i$ is measurable.
- 2 If A_1, A_2, A_3, \dots is a sequence of measurable sets in \mathbb{R}^n with $\sum_{i=1}^{\infty} \text{vol}(A_i) < \infty$ then $\bigcup_{i=1}^{\infty} A_i$ is measurable. Moreover, if the sets A_i are pairwise disjoint then

$$\text{vol}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \text{vol}(A_i). \quad (\sigma\text{-additivity})$$

- 3 Every bounded closed set $A \subset \mathbb{R}^n$ is measurable.
- 4 Every bounded open set $U \subset \mathbb{R}^n$ is measurable.
- 5 There exist bounded sets which are not measurable.

Notes

- In some texts (in particular in those using the measure-theoretic approach) also certain unbounded sets are called “measurable” and assigned volume ∞ . With this extended definition, the set of measurable subsets of \mathbb{R}^n is closed with respect to denumerable unions and intersections and also with respect to the complementation map $A \mapsto \mathbb{R}^n \setminus A$. (It forms a so-called σ -algebra.)
- The key property of the Lebesgue measure is its σ -additivity.
- All countable subsets of \mathbb{R}^n (in particular all finite subsets) have measure zero. This follows from $\text{vol}(\{\mathbf{a}\}) = 0$ and σ -additivity, but we have also seen a proof for the case of \mathbb{Q} .
- All subsets of a set of measure zero are measurable (and then, of course, have measure zero as well).
- (4) is proved using the following property of open sets $U \subset \mathbb{R}^n$: There exists a sequence Q_1, Q_2, Q_3, \dots of n -dimensional intervals such that $U = \bigcup_{i=1}^{\infty} Q_i$. The Q_i can be selected such that their endpoints have coordinates of the form $k/2^s$ with $k \in \mathbb{Z}$. Similarly, every bounded closed set A has a representation $A = \bigcap_{i=1}^{\infty} F_i$ with F_i a finite union of Q_j 's.

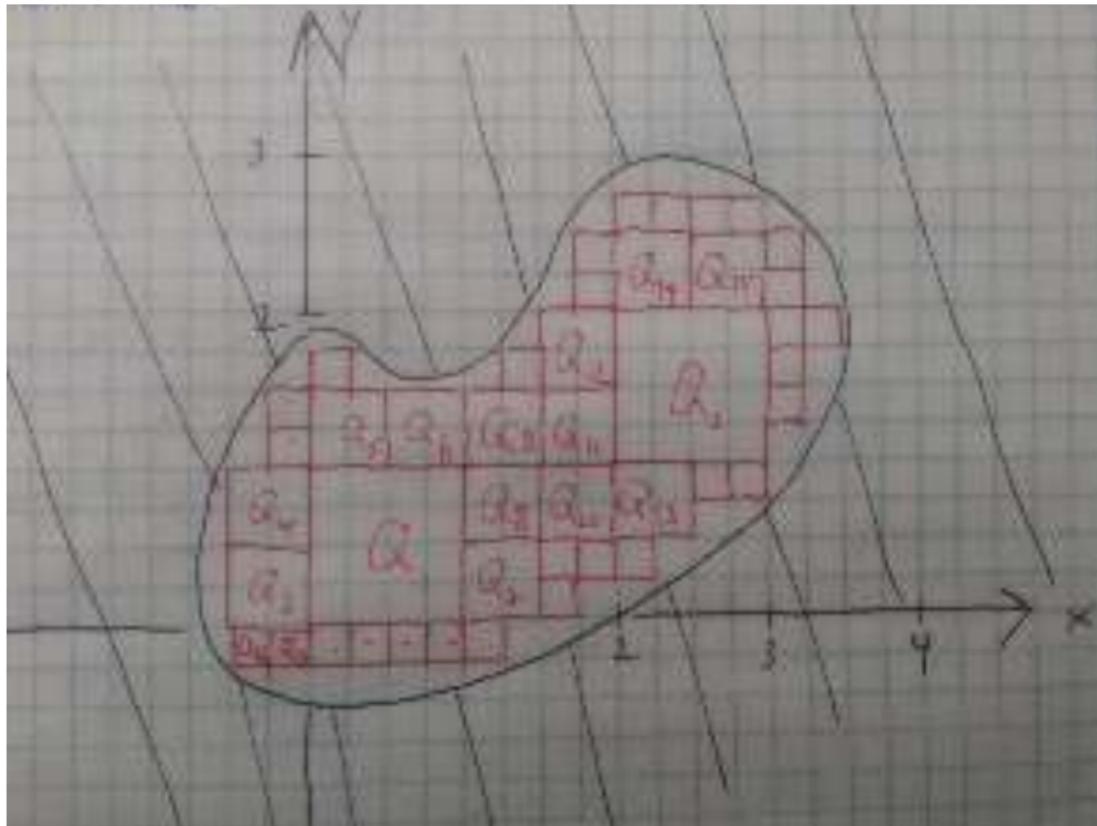


Figure: Every non-empty open set in \mathbb{R}^2 is the union of countably many closed squares Q_1, Q_2, Q_3, \dots with sides parallel to the coordinate axes and mutually disjoint interiors.

CANTOR's Ternary Set

Definition

Cantor's ternary set C is defined as follows:

$$C_0 = [0, 1] \text{ (the unit interval in } \mathbb{R}\text{);}$$

$$C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1];$$

$$C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{3}{9}] \cup [\frac{6}{9}, \frac{7}{9}] \cup [\frac{8}{9}, 1];$$

Recursively, C_{i+1} arises from C_i by removing from every interval in C_i the middle third (without the endpoints).

Finally, we set $C = \bigcap_{i=0}^{\infty} C_i$.

By induction, C_i consists of 2^i pairwise disjoint closed intervals of length 3^{-i} .

$$\implies \text{vol}(C_i) = \left(\frac{2}{3}\right)^i$$

$$\implies \text{vol}(C) = \lim_{i \rightarrow \infty} \left(\frac{2}{3}\right)^i = 0$$

Remark

It can be shown that C consists of all numbers that admit a ternary expansion $0.b_1b_2b_3\dots$ with $b_i \in \{0, 2\}$ (terminating or non-terminating). For example, $\frac{1}{3} = 0.1 = 0.0222\dots \in C$, $\frac{2}{3} = 0.2 \in C$, $1 = 0.222\dots \in C$, but $\frac{1}{2} = 0.111\dots \notin C$. Hence C has continuously many elements (as many as \mathbb{R} itself).



Figure: The construction of Cantor's Ternary Set $C = \bigcap_{i=0}^{\infty} C_i$

A Characterization of Sets of Measure Zero

Theorem

$N \subseteq \mathbb{R}^n$ has measure zero iff for every $\epsilon > 0$ there exists a sequence Q_1, Q_2, Q_3, \dots of n -dimensional intervals with $N \subseteq \bigcup_{k=1}^{\infty} Q_k$ and $\sum_{k=1}^{\infty} \text{vol}(Q_k) < \epsilon$.

Proof.

“ \Leftarrow ” is clear. “ \Rightarrow ” is proved by constructing an open set U with $N \subset U$ and $\text{vol}(U) < \epsilon$ (from an enveloping series Φ for χ_N with $I(\Phi) < \epsilon$), and a sequence (Q_k) of n -dimensional intervals with $\bigcup_{k=1}^{\infty} Q_k = U$. □

Example

A smooth hypersurface H in \mathbb{R}^n satisfies $\text{vol}(H) = 0$.

This follows from the fact that H is locally the graph of a continuous function $f: U' \rightarrow \mathbb{R}$ with $U' \subseteq \mathbb{R}^{n-1}$. The domain U' can be chosen as an $(n - 1)$ -dimensional closed interval.

Given $\epsilon > 0$, the interval U' can be partitioned into finitely many subintervals Q_1, Q_2, \dots, Q_N such that $|f(\mathbf{x}') - f(\mathbf{y}')| < \epsilon$ if \mathbf{x}', \mathbf{y}' are in the same subinterval Q_k .

Example (cont'd)

This gives a covering of G_f by n -dimensional intervals of the form $Q_k \times I_k$ with I_k of length ϵ and centered around $f(\mathbf{x}'_k)$, $\mathbf{x}'_k \in Q_k$. It follows that $\text{vol}(G_f) \leq \epsilon \text{ vol}(U')$ for every $\epsilon > 0$, i.e., $\text{vol}(G_f) = 0$. Further one can show—even if H is unbounded—that H is covered by countably many such graphs $G_{f_1}, G_{f_2}, G_{f_3}, \dots$. Since the union of countably many sets of volume zero has itself volume zero (by the σ -additivity of vol), we also have $\text{vol}(H) = 0$.

Notes

- Of course k -dimensional smooth surfaces in \mathbb{R}^{n-1} with $k < n - 1$ have volume zero as well.
- Level sets $N_f(k)$ of a C^1 -function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$ open, are guaranteed to have volume zero if f has no critical point \mathbf{x} with $f(\mathbf{x}) = k$ (and, consequently, $N_f(k)$ is smooth), but not in general. For this note, e.g., that we can extend the constant function $[0, 1] \rightarrow \mathbb{R}$, $x \mapsto 0$, to a C^1 -function $f: \mathbb{R} \rightarrow \mathbb{R}$ satisfying $N_f(0) = [0, 1]$ and hence $\text{vol}(N_f(0)) \neq 0$.

The Main Theorems of Lebesgue Integration

Definition

Let $P = P(\mathbf{x})$ be a property of $\mathbf{x} \in \mathbb{R}^n$. (In Discrete Mathematics terms, P is a unary predicate with domain \mathbb{R}^n .) We say that P holds *almost everywhere* if

$$\text{vol} (\{\mathbf{x} \in \mathbb{R}^n; P(\mathbf{x}) \text{ is false}\}) = 0.$$

Examples

- ① Almost every point in the unit interval $[0, 1]$ does not belong to Cantor's set C .
- ② Two functions $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$ agree almost everywhere if the set $\{\mathbf{x} \in \mathbb{R}^n; f(\mathbf{x}) \neq g(\mathbf{x})\}$ has volume zero.
- ③ The characteristic functions of denumerable subsets of \mathbb{R}^n are zero almost everywhere; likewise for χ_C (characteristic function of the Cantor set in \mathbb{R}) and the characteristic functions of smooth surfaces in \mathbb{R}^n .

Lemma

Suppose $f, g: \mathbb{R}^n \rightarrow \mathbb{R}$ satisfy $f(\mathbf{x}) = g(\mathbf{x})$ almost everywhere. Then f is integrable iff g is integrable; if this is the case then $\int f = \int g$.

Theorem (Monotone Convergence Theorem, B. LEVI)

Suppose that $(f_k)_{k \in \mathbb{N}}$ is a non-decreasing sequence (i.e., $f_k(\mathbf{x}) \leq f_{k+1}(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ and all $k \in \mathbb{N}$) of integrable functions on \mathbb{R}^n and that the sequence of integrals $(\int f_k)_{k \in \mathbb{N}}$ is bounded. Then $f(\mathbf{x}) = \lim_{k \rightarrow \infty} f_k(\mathbf{x})$ is finite almost everywhere, and the limit function $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ is integrable with

$$\int f = \lim_{k \rightarrow \infty} \int f_k.$$

Notes

- Since for $\mathbf{x} \in \mathbb{R}^n$ the sequence $(f_k(\mathbf{x}))_{k \in \mathbb{N}}$ is non-decreasing, it has a limit in $\bar{\mathbb{R}}$ (which can be either finite or $+\infty$). Thus the limit function $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ is well-defined.

Notes cont'd

- The limit function f is also known as *point-wise limit* of the function sequence (f_k) , because convergence can be tested for each argument x separately. There is also the notion of *uniform convergence* (as opposed to point-wise convergence), which roughly requires the “speed of convergence” to be the same for all arguments x .
- Since $f_k \leq f_{k+1}$ implies $\int f_k \leq \int f_{k+1}$, the sequence of integrals $(\int f_k)_{k \in \mathbb{N}}$ of a non-decreasing sequence of integrable functions is a non-decreasing sequence of real numbers and thus either bounded or $\lim_{k \rightarrow \infty} \int f_k = +\infty$.
- In general, integrable functions must be finite almost everywhere. This (not at all obvious) fact can be seen as follows: The set $N = \{\mathbf{x} \in \mathbb{R}^n; f(\mathbf{x}) = \pm\infty\}$ satisfies $\chi_N(\mathbf{x}) \leq \epsilon |f(\mathbf{x})|$ for all $\mathbf{x} \in \mathbb{R}^n$ and $\epsilon > 0$, because $\mathbf{x} \in N$ implies $|f(\mathbf{x})| = +\infty$ and $\epsilon \cdot (+\infty) = +\infty$. It follows that $\|\chi_N\|_1 \leq \|\epsilon |f|\|_1 = \epsilon \|f\|_1$ for all $\epsilon > 0$. Since $\|f\|_1$ is finite, this can only hold if $\|\chi_N\|_1 = 0$, i.e., N is measurable with $\text{vol}(N) = 0$.

Example

Consider the sequence of functions $f_k: \mathbb{R} \rightarrow \mathbb{R}$ ($k = 1, 2, \dots$) defined by

$$f_k(x) = \begin{cases} x - x^k & \text{for } 0 \leq x \leq 1, \\ 0 & \text{for } x < 0 \vee x > 1. \end{cases}$$

Since $x^{k+1} \leq x^k$ for $0 \leq x \leq 1$, we have $f_{k+1}(x) \geq f_k(x)$ for all $x \in \mathbb{R}$, $k \in \mathbb{N}$. The functions f_k are integrable with

$$\int_{\mathbb{R}} f_k(x) dx = \int_0^1 (x - x^k) dx = \left[\frac{x^2}{2} - \frac{x^{k+1}}{k+1} \right]_0^1 = \frac{1}{2} - \frac{1}{k+1}.$$

Since $\int f_k \leq 1/2$ for all k , B. Levi's Theorem applies and gives $\int \lim_{k \rightarrow \infty} f_k(x) dx = \lim_{k \rightarrow \infty} \int f_k(x) dx = 1/2$.

On the other hand, for $0 \leq x < 1$ we have $x - x^k \rightarrow x$ for $k \rightarrow \infty$, so that the limit function $f = \lim_{k \rightarrow \infty} f_k$ is

$$f(x) = \begin{cases} x & \text{for } 0 < x < 1, \\ 0 & \text{for } x \leq 0 \vee x \geq 1. \end{cases}$$

We have $\int f(x) dx = \int_0^1 dx = 1/2$, in sync with B. Levi's Theorem.

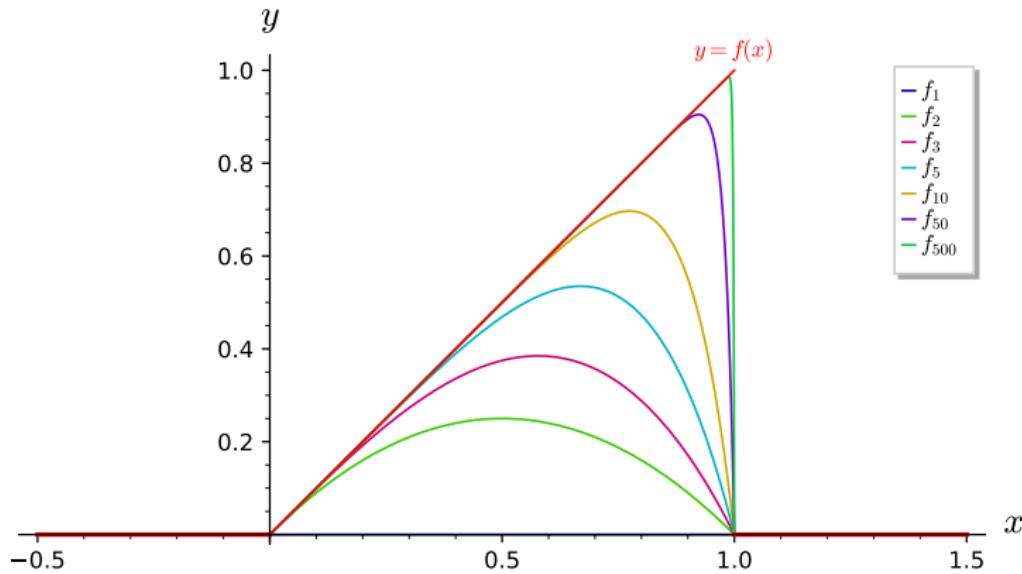


Figure: Illustration of the point-wise convergence $f_k \rightarrow f$

You can see that the speed of convergence of $f_k(x) \uparrow f(x)$ slows down for $x \uparrow 1$.

Example (cont'd)

Since all integrals involved are easily evaluated, this example doesn't show the real value of B. Levi's Theorem. But we can easily change the functions f_k and make their integration complicated, e.g., $f_k(x) = x^{1+1/k} - \ln(x)^2 x^k$ for $0 < x \leq 1$, $f_k(0) = 0$. Then $f_{k+1}(x) \geq f_k(x)$, $\lim_{k \rightarrow \infty} f_k(x) = x$ still hold, and from the inequality $f_k(x) \leq x$ we obtain the bound

$\int_0^1 f_k(x) dx \leq 1/2$, as before. Thus B. Levi's Theorem applies again and gives

$$\lim_{k \rightarrow \infty} \int_0^1 x^{1+1/k} - \ln(x)^2 x^k dx = 1/2.$$

Corollary (“Improper” Lebesgue Integration)

Suppose $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ is a nested sequence of measurable sets in \mathbb{R}^n , $A = \bigcup_{k=1}^{\infty} A_k$, and $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, is a function defined on A (i.e., $D \supseteq A$). Then f is integrable over A if and only if f is integrable over each A_k and the sequence

$\left(\int_{A_k} |f| \right)_{k \in \mathbb{N}}$ is bounded (from above). If this is the case, we have

$$\int_A f = \lim_{k \rightarrow \infty} \int_{A_k} f.$$

Proof.

Extend f to \mathbb{R}^n by setting $f(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathbb{R}^n \setminus D$, and consider first the sequence of functions $f_k = f^+ \chi_{A_k}$, which satisfy

$$f_k(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} \in A_k \text{ and } f(\mathbf{x}) \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

$$A_k \subseteq A_{k+1} \wedge f^+ \geq 0 \implies f_k \leq f_{k+1}$$

$$\bigcup_{k=1}^{\infty} A_k = A \implies \lim_{k \rightarrow \infty} f_k = f^+ \chi_A.$$

Proof cont'd.

\Leftarrow : With f also f^+ (and f^-) is integrable over each A_k , and hence $f_k = f^+ \chi_{A_k}$ is integrable over \mathbb{R}^n ;

$f^+ = \max(f, 0) \leq |f|$ implies $f_k = f^+ \chi_{A_k} \leq |f| \chi_{A_k}$ and hence $\int f_k \leq \int |f| \chi_{A_k} = \int_{A_k} |f|$. This shows that the sequence $(\int f_k)_{k \in \mathbb{N}}$ is bounded as well.

Thus the sequence (f_k) satisfies all assumptions of B. Levi's Theorem, and the theorem gives that $f^+ \chi_A$ is integrable over \mathbb{R}^n with

$$\int f^+ \chi_A = \lim_{k \rightarrow \infty} \int f_k = \lim_{k \rightarrow \infty} \int f^+ \chi_{A_k}.$$

This can also be written as $\int_A f^+ = \lim_{k \rightarrow \infty} \int_{A_k} f^+$ (in particular f^+ is integrable over A), and proves half of the assertion.

(Equivalently, it proves the assertion for non-negative functions f .)

For f^- the same reasoning can be used to prove

$$\int_A f^- = \lim_{k \rightarrow \infty} \int_{A_k} f^-.$$

Proof cont'd.

Putting both halves together finally gives that $f = f^+ - f^-$ is integrable over A and

$$\begin{aligned}\int_A f &= \int_A f^+ - \int_A f^- = \lim_{k \rightarrow \infty} \int_{A_k} f^+ - \lim_{k \rightarrow \infty} \int_{A_k} f^- \\ &= \lim_{k \rightarrow \infty} \int_{A_k} (f^+ - f^-) = \lim_{k \rightarrow \infty} \int_{A_k} f.\end{aligned}$$

\implies : Conversely, suppose that f is integrable over A . Then the same is true of $|f|$, and all functions $f \chi_{A_k}$ are integrable over \mathbb{R}^n (since the sets A_k are measurable and characteristic functions are trivially bounded.) Thus the first condition holds. In order to derive the bound, observe that $|f| \chi_{A_k} \leq |f| \chi_A$ gives

$$\int_{A_k} |f| = \int_{\mathbb{R}^n} |f| \chi_{A_k} \leq \int_{\mathbb{R}^n} |f| \chi_A = \int_A |f|.$$

Thus $\int_A |f|$ is the required bound for $(\int_{A_k} |f|)_{k \in \mathbb{N}}$. □

Example

We show that $f(x, y) = e^{-(x^2+y^2)}$ is integrable (over \mathbb{R}^2) and

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)} d^2(x, y) = \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2 = \left(\int_{-\infty}^{+\infty} e^{-x^2} dx \right)^2.$$

For $k \in \mathbb{N}$ let $A_k = [-k, k]^2 = \{(x, y) \in \mathbb{R}^2; -k \leq x, y \leq k\}$. Since A_k is a 2-dimensional interval and f is continuous, f is integrable over A_k , and we clearly have $A_k \subseteq A_{k+1}$ and $\mathbb{R}^2 = \bigcup_{k=1}^{\infty} A_k$. Fubini's Little Theorem gives

$$\begin{aligned} \int_{A_k} |f(x, y)| d^2(x, y) &= \int_{A_k} f(x, y) d^2(x, y) = \int_{[-k, k]^2} e^{-x^2} e^{-y^2} d^2(x, y) \\ &= \int_{[-k, k]} e^{-x^2} dx \int_{[-k, k]} e^{-y^2} dy = \left(\int_{[-k, k]} e^{-x^2} dx \right)^2. \end{aligned}$$

Hence, in order to apply the corollary, it suffices to show that the sequence $\left(\int_{[-k, k]} e^{-x^2} dx \right)_{k \in \mathbb{N}}$ is bounded.

Example (cont'd)

This follows, e.g., by using $e^{-x^2} \leq xe^{-x^2}$ for $x \geq 1$, which implies

$$\begin{aligned} \int_{[-k,k]} e^{-x^2} dx &\leq \int_{-1}^1 e^{-x^2} dx + 2 \int_1^k xe^{-x^2} dx \\ &= \int_{-1}^1 e^{-x^2} dx + \left[-e^{-x^2} \right]_1^k \\ &= \int_{-1}^1 e^{-x^2} dx + e^{-1} - e^{-k^2} \leq \int_{-1}^1 e^{-x^2} dx + e^{-1}, \end{aligned}$$

a bound that is independent of k .

Hence we can apply the corollary to conclude that

$(x, y) \mapsto e^{-(x^2+y^2)}$ and $x \mapsto e^{-x^2}$ are integrable over \mathbb{R}^2 , respectively, over \mathbb{R} , and that

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-(x^2+y^2)} d^2(x, y) &= \lim_{k \rightarrow \infty} \int_{A_k} e^{-(x^2+y^2)} d^2(x, y) \\ &= \lim_{k \rightarrow \infty} \left(\int_{[-k,k]} e^{-x^2} dx \right)^2 = \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2. \end{aligned}$$

Notes

- This example is a little “constructed”, since the general Fubini theorem (to be discussed later) and its converse (called Tonelli’s Theorem) yield that $(x, y) \mapsto e^{-(x^2+y^2)}$ is integrable over \mathbb{R}^2 (from the integrability of $x \mapsto e^{-x^2}$ over \mathbb{R}) and the formula just proved without resort to the sets A_k .
- Applying the corollary to the function $f = \chi_A$ itself shows that

$$\text{vol}(A) = \text{vol}\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k \rightarrow \infty} \text{vol}(A_k),$$

provided that the sequence $(\text{vol}(A_k))_{k \in \mathbb{N}}$ is bounded. If instead we are given a sequence of mutually disjoint sets B_k , we can set $A_k = \biguplus_{i=1}^k B_i$ and conclude in the same way that

$$\begin{aligned} \text{vol}\left(\biguplus_{k=1}^{\infty} B_k\right) &= \text{vol}\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k \rightarrow \infty} \text{vol}(A_k) \\ &= \lim_{k \rightarrow \infty} \sum_{i=1}^k \text{vol}(B_k) = \sum_{k=1}^{\infty} \text{vol}(B_k). \end{aligned}$$

Thus the σ -additivity of the Lebesgue measure is in a sense equivalent to B. Levi’s Theorem.

Example (The Gamma Function)

The Gamma function can be defined by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, \quad x \in (0, +\infty).$$

It satisfies $\Gamma(1) = \int_0^\infty e^{-t} dt = [-e^{-t}]_0^\infty = 1$ and the functional equation $\Gamma(x+1) = x \cdot \Gamma(x)$.

The latter can be proved using integration by parts:

$$\begin{aligned}\Gamma(x+1) &= \int_0^\infty t^x e^{-t} dt \\ &= [-t^x e^{-t}]_0^\infty - \int_0^\infty -xt^{x-1} e^{-t} dt \\ &\qquad\qquad\qquad (u(t) = t^x, v'(t) = e^{-t}) \\ &= 0 + x \int_0^\infty t^{x-1} e^{-t} dt = x \cdot \Gamma(x).\end{aligned}$$

By induction, the functional equation gives $\Gamma(n+1) = n!$ for $n \in \mathbb{N}$, so that Γ may be viewed as an analytic extension of the factorial function to the positive real axis \mathbb{R}^+ .

Example (cont'd)

But why are the integrals $\int_0^\infty t^{x-1}e^{-t} dt$, $x > 0$, defined in the first place?

Using the corollary to B. Levi's Theorem, we can show this as follows:

For $k \in \mathbb{N}$ set $A_k = [\frac{1}{k}, k] = \{x \in \mathbb{R}; \frac{1}{k} \leq x \leq k\}$. Then clearly $A_k \subseteq A_{k+1}$ and $\bigcup_{k=1}^\infty A_k = (0, +\infty)$. Hence, in order to apply the corollary and conclude that the integrals in question exist, we only need to find a bound for $\int_{1/k}^k t^{x-1}e^{-t} dt$ that is independent of k (but may depend on x).

Suppose $N \in \mathbb{N}$ satisfies $1/N < x < N$. (Clearly, given $x > 0$, we can find such an integer N .) Then

$$t^{x-1}e^{-t} \leq \begin{cases} t^{N-1}e^{-t} & \text{for } t \geq 1, \\ t^{1/N-1} & \text{for } 0 < t \leq 1. \end{cases}$$

$$\begin{aligned} \implies \int_{1/k}^k t^{x-1}e^{-t} dt &= \int_{1/k}^1 t^{x-1}e^{-t} dt + \int_1^k t^{x-1}e^{-t} dt \\ &\leq \int_{1/k}^1 t^{1/N-1} dt + \int_1^k t^{N-1}e^{-t} dt. \end{aligned}$$

Example (cont'd)

The first integral on the right-hand side can be bounded independently of k , since $\frac{1}{N} - 1 > -1$:

$$\int_{1/k}^1 t^{1/N-1} dt \leq \int_0^1 t^{1/N-1} dt = \left[Nt^{1/N} \right]_0^1 = N.$$

Similarly, the second integral on the right-hand side can be bounded independently of k by

$$\int_1^\infty t^{N-1} e^{-t} dt < \infty.$$

The existence of these integrals is a consequence of the fast growth of $t \mapsto e^t$ for $t \rightarrow \infty$.

Remark

The integrals $I_n = \int_1^\infty t^n e^{-t} dt$, $n = 0, 1, 2, \dots$, can also be evaluated directly using integration by parts, which yields the recurrence relation $I_n = e^{-1} + n I_{n-1}$. Together with

$I_0 = \int_1^\infty e^{-t} dt = e^{-1}$ one then finds that $I_n = e^{-1} \sum_{k=0}^n \frac{n!}{k!}$.

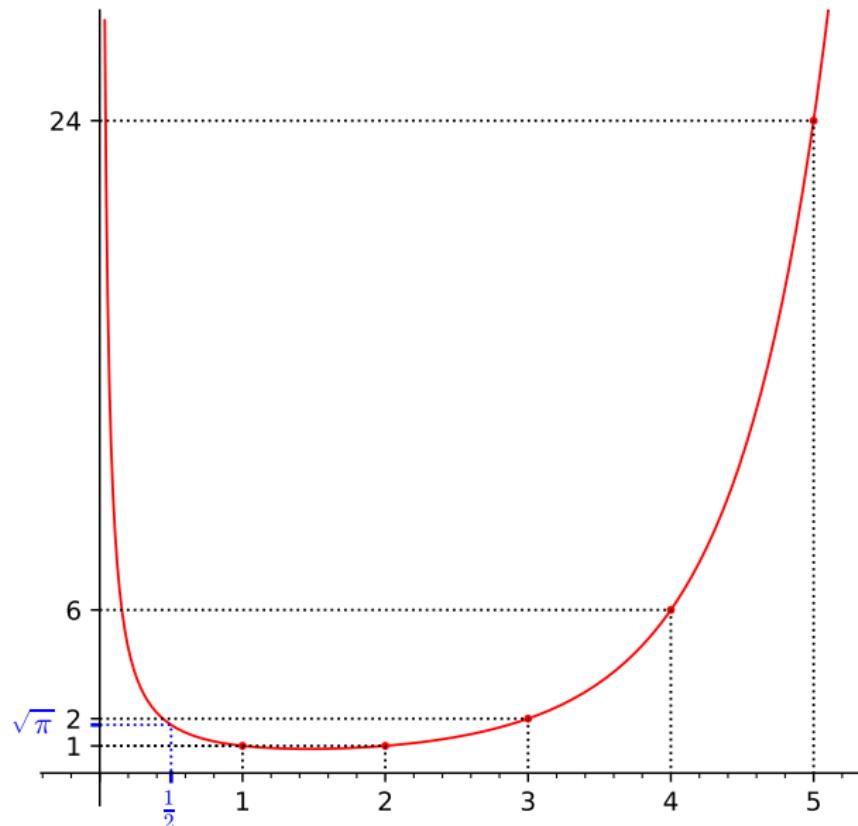


Figure: The Gamma function on $(0, \infty)$

Theorem (Bounded Convergence Theorem, H. LEBESGUE)

Suppose that $(f_k)_{k \in \mathbb{N}}$ is a sequence of integrable functions on \mathbb{R}^n converging almost everywhere and that there exists an integrable function (integrable “bound”) $\Phi \geq 0$ on \mathbb{R}^n such that $|f_k(\mathbf{x})| \leq \Phi(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$. Then the limit function $f(\mathbf{x}) = \lim_{k \rightarrow \infty} f_k(\mathbf{x})$ (extended to \mathbb{R}^n in any way) is integrable with

$$\int f = \lim_{k \rightarrow \infty} \int f_k.$$

Notes

- The theorem is also called “Dominated Convergence Theorem”.
- Compared with B. Levi's Theorem, the condition $f_k(\mathbf{x}) \leq f_{k+1}(\mathbf{x})$ is dropped and instead the existence of a limit function (at almost every point of the domain) is required. In the special case $0 \leq f_k \leq f_{k+1} \rightarrow f$ with f a known integrable function both theorems apply (Lebesgue's Theorem with $\Phi = f$).

Example

Consider the function sequence $f_k(x) = x/2 + (-1)^k x^k$ for $x \in [0, 1]$, $f_k(x) = 0$ for $x \in \mathbb{R} \setminus [0, 1]$.

The limit function f is defined for $x \neq 1$ and satisfies $f(x) = x/2$ for $x \in [0, 1)$, $f(x) = 0$ for $x < 0$ or $x > 1$.

$|f_k(x)| \leq 3/2$ for all k and $x \in [0, 1]$, so that we can take

$$\Phi(x) = \begin{cases} 3/2 & \text{if } 0 \leq x \leq 1, \\ 0 & \text{if } x < 0 \vee x > 1, \end{cases}$$

as integrable bound in Lebesgue's Theorem. It follows that

$$\lim_{k \rightarrow \infty} \int_0^1 x/2 + (-1)^k x^k \, dx = \int_0^1 x/2 \, dx = \frac{1}{4}.$$

Again this can easily be shown directly by evaluating $\int_0^1 f_k(x) \, dx$. But the remarks about more complicated sequences (f_k) apply here as well, of course.

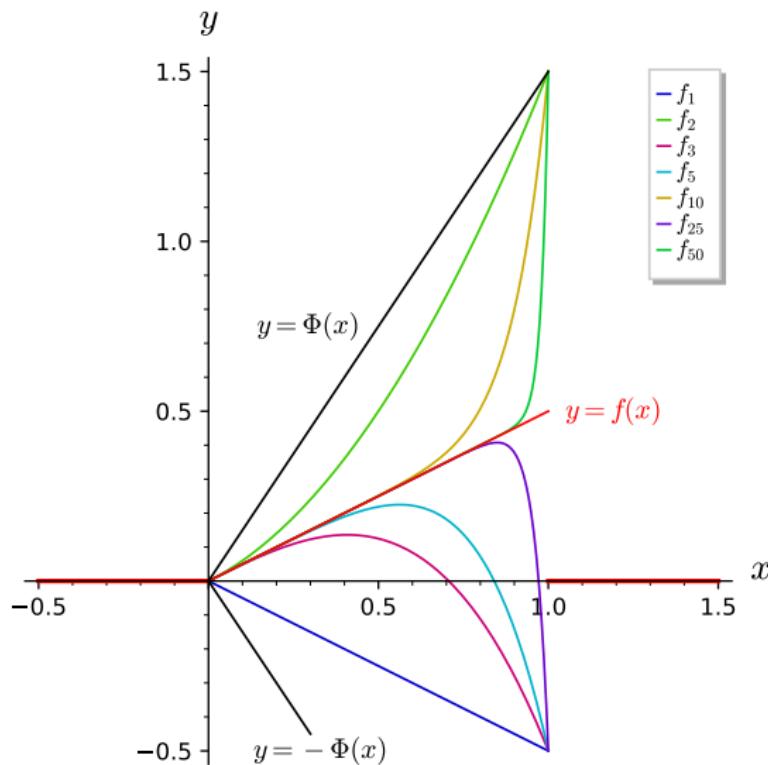


Figure: Illustration of the point-wise convergence $f_k \rightarrow f$ and the integrable bound Φ (In the example the subsequences (f_{2k}) and (f_{2k+1}) converge monotonically to f , which isn't the case in general.)

Example (wandering hill)

Consider $f_k: \mathbb{R} \rightarrow \mathbb{R}$ ($k = 0, 1, 2, \dots$) defined by

$$f_k(x) = \begin{cases} 1 & \text{if } k \leq x \leq k+1, \\ 0 & \text{if } x < k \vee x > k+1. \end{cases}$$

Here we have

$$\int_{\mathbb{R}} f_k(x) dx = \int_k^{k+1} 1 dx = 1 \quad \text{for all } k,$$

so that $\lim_{k \rightarrow \infty} \int_{\mathbb{R}} f_k(x) dx = 1$.

On the other hand we have, considering $x \in \mathbb{R}$ as fixed, $f_k(x) = 0$ for $k > \lceil x \rceil$ and hence $\lim_{k \rightarrow \infty} f_k(x) = 0$ for all $x \in \mathbb{R}$.

$$\implies \int_{\mathbb{R}} \lim_{k \rightarrow \infty} f_k(x) dx = \int_{\mathbb{R}} 0 dx = 0.$$

This doesn't contradict Lebesgue's Theorem: The smallest bound for f_k in the theorem is $\Phi(x) = 1$ for $x \geq 0$, $\Phi(x) = 0$ for $x < 0$, which is not integrable.

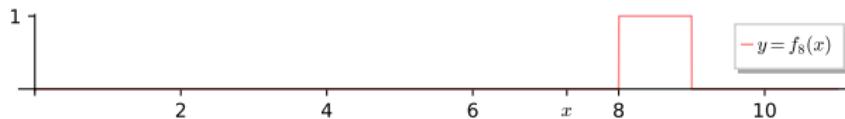
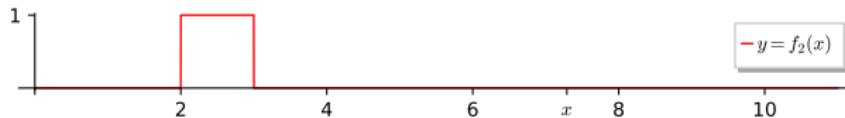


Figure: The wandering hill

Since $f_k(x) = 0$ for $k \geq \lceil x \rceil = 8$, we have $\lim_{k \rightarrow \infty} f_k(x) = 0$.

Parameter Integrals

An important application of H. Lebesgue's Theorem

Suppose $X \subseteq \mathbb{R}^m$, $Y \subseteq \mathbb{R}^n$ and $f: X \times Y \rightarrow \mathbb{R}$ are such that Y is measurable and $\mathbf{y} \rightarrow f(\mathbf{x}, \mathbf{y})$ is integrable over Y for each $\mathbf{x} \in X$. Then

$$F(\mathbf{x}) = \int_Y f(\mathbf{x}, \mathbf{y}) d^n \mathbf{y}, \quad \mathbf{x} \in X,$$

defines a function $F: X \rightarrow \mathbb{R}$.

Example (LAPLACE Transform)

The *Laplace Transform* of an integrable function $f: [0, +\infty) \rightarrow \mathbb{R}$ is the function $F: [0, +\infty) \rightarrow \mathbb{R}$ defined by

$$F(s) = \int_0^{+\infty} f(t) e^{-st} dt.$$

Example (One-dimensional FOURIER Transform)

The *Fourier transform* of an integrable function $f: \mathbb{R} \rightarrow \mathbb{C}$ is the function $\hat{f}: \mathbb{R} \rightarrow \mathbb{C}$ defined by

$$\hat{f}(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(t) e^{-ixt} dt.$$

The Laplace transform is an important tool for solving Ordinary Differential Equations (ODE's).

The Fourier transform is a (if not the) key concept of advanced Real Analysis and has important applications in engineering as well (e.g., in signal processing).

Here we will not discuss these transforms further, but you'll meet them again in Math 285 next semester.

Question

What can be said about continuity, differentiability, and (partial) derivatives of F in terms of f ?

Answer

Recall that F is continuous at $\mathbf{x} \in X$ if for every sequence $\mathbf{x}_k \rightarrow \mathbf{x}$ the image sequence $F(\mathbf{x}_k)$ converges to $F(\mathbf{x})$; cf. Exercise H71 of Homework 12. (You may have learned this fact in Calculus I for one-variable functions.) In the case under consideration we have

$$F(\mathbf{x}_k) = \int_Y f(\mathbf{x}_k, \mathbf{y}) d^n \mathbf{y} = \int_Y f_k(\mathbf{y}) d^n \mathbf{y}, \quad \text{say,}$$

and $f_k(\mathbf{y}) \rightarrow f(\mathbf{x}, \mathbf{y})$ for each $\mathbf{y} \in Y$, provided that for each $\mathbf{y} \in Y$ the function $\mathbf{x} \rightarrow f(\mathbf{x}, \mathbf{y})$ is continuous.

\Rightarrow We can apply H. Lebesgue's Theorem to the sequence (f_k) and conclude that F is continuous if we have a “uniform” bound $|f(\mathbf{x}, \mathbf{y})| \leq \Phi(\mathbf{y})$ with Φ integrable over Y .

Theorem

Suppose $f: X \times Y \rightarrow \mathbb{R}$ ($X \subseteq \mathbb{R}^m$, $Y \subseteq \mathbb{R}^n$) is such that $\mathbf{y} \rightarrow f(\mathbf{x}, \mathbf{y})$ is integrable for each $\mathbf{x} \in X$ and $F: X \rightarrow \mathbb{R}$ is defined by $F(\mathbf{x}) = \int_Y f(\mathbf{x}, \mathbf{y}) d^n \mathbf{y}$.

- 1 If $\mathbf{x} \rightarrow f(\mathbf{x}, \mathbf{y})$ is continuous for each $\mathbf{y} \in Y$ and there exists an integrable function $\Phi: Y \rightarrow \mathbb{R}$ such that $|f(\mathbf{x}, \mathbf{y})| \leq \Phi(\mathbf{y})$ for all $(\mathbf{x}, \mathbf{y}) \in X \times Y$, then F is continuous.
- 2 If $\mathbf{x} \rightarrow f(\mathbf{x}, \mathbf{y})$ is a C^1 -function for each $\mathbf{y} \in Y$ and there exists an integrable function $\Phi: Y \rightarrow \mathbb{R}$ such that $\left| \frac{\partial f}{\partial x_j}(\mathbf{x}, \mathbf{y}) \right| \leq \Phi(\mathbf{y})$ for all $(\mathbf{x}, \mathbf{y}) \in X \times Y$ and $1 \leq j \leq m$, then F is itself a C^1 -function and satisfies

$$\frac{\partial F}{\partial x_j}(\mathbf{x}) = \int_Y \frac{\partial f}{\partial x_j}(\mathbf{x}, \mathbf{y}) d^n \mathbf{y} \quad \text{for } 1 \leq j \leq m.$$

Notes

- In the situation of Part (2) of the theorem we say that $F(\mathbf{x}) = \int_Y f(\mathbf{x}, \mathbf{y}) d^n \mathbf{y}$ can be differentiated under the integral sign.
- The conclusions of the theorem hold in particular if X is open (required anyway in (2)), Y is compact and, e.g., in (1) f is a continuous function of $(\mathbf{x}, \mathbf{y}) \in X \times Y$. Indeed, given $\mathbf{x} \in X$ we can replace X be a closed ball around \mathbf{x} and hence assume that X is compact as well. The product $X \times Y$ is then compact as well, and hence there exists $M \in \mathbb{R}$ such that $|f(\mathbf{x}, \mathbf{y})| \leq M$ for all $(\mathbf{x}, \mathbf{y}) \in X \times Y$. Thus we can take $\Phi(\mathbf{y}) = M$ for $\mathbf{y} \in Y$, which is integrable since Y is compact (and hence measurable).
- Since continuity and differentiability are local properties, the assumptions in the theorem can be weakened to $|f(\mathbf{x}', \mathbf{y})| \leq \Phi(\mathbf{y})$, resp., $\left| \frac{\partial f}{\partial x_j}(\mathbf{x}', \mathbf{y}) \right| \leq \Phi(\mathbf{y})$ for all $(\mathbf{x}', \mathbf{y}) \in B_\delta(\mathbf{x}) \times Y$, where $\delta > 0$ may depend on $\mathbf{x} \in X$. In other words, $\Phi(\mathbf{y}) = \Phi_{\mathbf{x}}(\mathbf{y})$ may depend on $\mathbf{x} \in X$, as long as it provides a uniform bound working for all \mathbf{x}' in some neighborhood of \mathbf{x} . In the subsequent examples the theorem is applied in this modified form.

Notes cont'd

- Part (2) of the theorem generalizes to *k*th-order partial derivatives $D^\alpha F = D_1^{\alpha_1} \cdots D_m^{\alpha_m} F$, $\alpha_1 + \cdots + \alpha_m = k \geq 2$, if $\mathbf{x} \rightarrow f(\mathbf{x}, \mathbf{y})$ is a C^k -function for each $\mathbf{y} \in Y$ and there exists an integrable bound $\Phi: Y \rightarrow \mathbb{R}$ for the partial derivatives of order k of $f(\mathbf{x}, \mathbf{y})$ with respect to x_1, \dots, x_m , which is independent of \mathbf{x} and α .

This can be proved by induction on k , using the fact that an integrable bound $\Phi(\mathbf{y})$ for all (!) partial derivatives of order k implies the existence of such a bound locally for partial derivatives of smaller order (just like a bound for $g'(x)$ implies one for $g(x)$ via $g(x) = g(a) + \int_a^x g'(t) dt$ on $[a - R, a + R]$ in the 1-variable case).

Since there are only finitely many partial derivatives of a given order k , individual bounds for each of them imply a uniform bound that works for all of them (just take the maximum of the individual bounds).

Proof of the theorem.

The proof of Part 1 has been outlined before stating the theorem.

The proof of Part 2 is similar but slightly more involved. We compute

$$\begin{aligned}\frac{\partial F}{\partial x_j}(\mathbf{x}) &= \lim_{t \rightarrow 0} \frac{F(\mathbf{x} + t \mathbf{e}_j) - F(\mathbf{x})}{t} \\&= \lim_{t \rightarrow 0} \frac{1}{t} \left(\int_Y f(\mathbf{x} + t \mathbf{e}_j, \mathbf{y}) d^n \mathbf{y} - \int_Y f(\mathbf{x}, \mathbf{y}) d^n \mathbf{y} \right) \\&= \lim_{t \rightarrow 0} \int_Y \frac{f(\mathbf{x} + t \mathbf{e}_j, \mathbf{y}) - f(\mathbf{x}, \mathbf{y})}{t} d^n \mathbf{y} \\&= \int_Y \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t \mathbf{e}_j, \mathbf{y}) - f(\mathbf{x}, \mathbf{y})}{t} d^n \mathbf{y} = \int_Y \frac{\partial f}{\partial x_j}(\mathbf{x}, \mathbf{y}) d^n \mathbf{y},\end{aligned}$$

provided we are allowed to interchange limit and integration.

The latter can be justified using Lebesgue's Theorem. The details are as follows.

Proof cont'd.

Firstly, it suffices to show that for every sequence (t_k) of real numbers satisfying $t_k \rightarrow 0$ for $k \rightarrow \infty$ we have

$$\lim_{k \rightarrow \infty} \int_Y \frac{f(\mathbf{x} + t_k \mathbf{e}_j, \mathbf{y}) - f(\mathbf{x}, \mathbf{y})}{t_k} d^n \mathbf{y} = \int_Y \frac{\partial f}{\partial x_j}(\mathbf{x}, \mathbf{y}) d^n \mathbf{y}.$$

Considering \mathbf{x} as fixed, we set

$$f_k(\mathbf{y}) = \frac{f(\mathbf{x} + t_k \mathbf{e}_j, \mathbf{y}) - f(\mathbf{x}, \mathbf{y})}{t_k}, \quad k = 1, 2, \dots$$

Since $\lim_{k \rightarrow \infty} f_k(\mathbf{y}) = \frac{\partial f}{\partial x_j}(\mathbf{x}, \mathbf{y})$, the assertion follows from Lebesgue's Theorem, provided it can be applied. For this we need a bound $|f_k(\mathbf{y})| \leq \phi(\mathbf{y})$, which is independent of k and integrable over Y .

The Mean Value Theorem of Calculus I gives $\tau_k \in (0, t_k)$ such that $f_k(\mathbf{y}) = \frac{\partial f}{\partial x_j}(\mathbf{x} + \tau_k \mathbf{e}_j, \mathbf{y})$.

$$\Rightarrow |f_k(\mathbf{y})| = \left| \frac{\partial f}{\partial x_j}(\mathbf{x} + \tau_k \mathbf{e}_j, \mathbf{y}) \right| \leq \Phi(\mathbf{y}), \quad \text{since } \mathbf{x} + \tau_k \mathbf{e}_j \in X.$$

\Rightarrow We can take $\phi := \Phi$.

Proof cont'd.

Finally, Part 1 gives that the partial derivatives $\partial F/\partial x_j$ are continuous (because $\partial f/\partial x_j$ are continuous), so that F is a C^1 -function. □

The proof shows that the conclusions of the theorem hold under the weaker assumption that every point $\mathbf{x} \in X$ possesses a neighborhood U (which can be taken as a ball $B_\delta(\mathbf{x})$ of small radius $\delta > 0$), such that integrable bounds $|f(\mathbf{x}', \mathbf{y})| \leq \Phi(\mathbf{y})$, respectively, $\left| \frac{\partial f}{\partial x_j}(\mathbf{x}', \mathbf{y}) \right| \leq \Phi(\mathbf{y})$ can be found which are independent of $\mathbf{x}' \in U$. (We should rather write $\Phi_{\mathbf{x}}$ or Φ_U in place of Φ , because these functions and their domains may vary with \mathbf{x} .)

To adapt the proof to this situation, simply replace near the end “since $\mathbf{x} + \tau_k \mathbf{e}_j \in X$ ” by “since $\mathbf{x} + \tau_k \mathbf{e}_j \in U$ if k is sufficiently large”.

As noted after the theorem, in the subsequent examples the theorem will be applied in this modified form.

Example

The Gamma function $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, $x \in (0, \infty)$, is differentiable with derivative

$$\Gamma'(x) = \int_0^\infty (\ln t) t^{x-1} e^{-t} dt. \quad (*)$$

This follows from Part (2) of the theorem, since

$$\frac{\partial}{\partial x} (t^{x-1} e^{-t}) = \frac{\partial}{\partial x} \left(e^{(\ln t)(x-1)} e^{-t} \right) = (\ln t) t^{x-1} e^{-t},$$

and similar reasoning as in the existence proof of $\int_0^\infty t^{x-1} e^{-t} dt$. The key fact used here is that $t \mapsto |\ln t|$ grows very slowly both for $t \rightarrow +\infty$ and $t \downarrow 0$.

More precisely, by inspecting the proof we get an integrable bound

$$|(\ln t) t^{x-1} e^{-t}| \leq \Phi_N(t) \quad \text{for } 1/N < x < N.$$

Letting $X = (\frac{1}{N}, N)$ in the theorem, this proves $(*)$ for $x \in (\frac{1}{N}, N)$. Finally, letting $N \rightarrow \infty$ then proves it for all $x \in (0, \infty)$.

Example (cont'd)

Alternatively, given $x > 0$ choose N with $1/N < x$ and use the ball (interval) $B_\delta(x)$, $\delta := x - 1/N$, in the modified form of the theorem.

Repeating the argument shows further that

$$\Gamma^{(k)}(x) = \int_0^\infty (\ln t)^k t^{x-1} e^{-t} dt \quad \text{for } k = 1, 2, 3, \dots$$

In particular, $\Gamma''(x) > 0$, so that Γ is strictly convex ("cup-shaped"); cf. our earlier plot of the Gamma function.

Feynman's Technique

American physicist RICHARD P. FEYNMAN (1918–1988), who was awarded the Nobel Prize in Physics in 1965, popularized the evaluation of difficult 1-dimensional integrals using differentiation under the integral sign. Nowadays the method is known as “Feynman’s Technique” or “Feynman’s Trick”.

Example

Evaluate $\int_0^\infty \frac{\ln(x^2 + 1)}{x^2 + 1} dx.$

Though it seems paradoxical at the first glance, it is often easier to solve a more general integral with a parameter. Here we set

$$I(t) = \int_0^\infty \frac{\ln(x^2 + t)}{x^2 + 1} dx \quad \text{for } t \geq 0.$$

If we are able to determine the whole function $I(t)$, we can just insert $t = 1$ afterwards to solve the original problem.

Example (cont'd)

Differentiation under the integral sign (which requires justification!) yields

$$I'(t) = \int_0^\infty \frac{d}{dt} \frac{\ln(x^2 + t)}{x^2 + 1} dx = \int_0^\infty \frac{dx}{(x^2 + 1)(x^2 + t)}.$$

Since the integrand is a rational function of x , this integral is comparatively easy to solve, thereby explaining the choice of the specific parameter. This determines $I(t)$ up to a real constant, which can be found, e.g., if we are able to evaluate $I(t)$ for at least one instance of t .

Here we use

$$I(0) = \int_0^\infty \frac{\ln(x^2)}{x^2 + 1} dx = 2 \int_0^\infty \frac{\ln(x)}{x^2 + 1} dx,$$

which turns out to be zero.

First observe that this integral, which is improper at both ends, converges (even absolutely), so that we can rewrite it as follows:

Example (cont'd)

$$\begin{aligned}\int_0^\infty \frac{\ln(x)}{x^2 + 1} dx &= \int_0^1 \frac{\ln(x)}{x^2 + 1} dx + \int_1^\infty \frac{\ln(x)}{x^2 + 1} dx \\ &= \int_0^1 \frac{\ln(x)}{x^2 + 1} dx + \int_1^0 \frac{\ln(1/y)}{(1/y)^2 + 1} \frac{-1}{y^2} dy \\ &\quad (\text{Subst. } x = 1/y, dx = (-1/y^2) dy) \\ &= \int_0^1 \frac{\ln(x)}{x^2 + 1} dx - \int_0^1 \frac{\ln(y)}{y^2 + 1} dy = 0,\end{aligned}$$

since $\ln(1/y) = -\ln y$ and $\int_1^0 \dots = -\int_0^1 \dots$

$I'(t)$ can be evaluated using partial fractions, where we assume $t > 0$ (for $t = 0$ the integral isn't finite), and $t \neq 1$ for the moment:

$$\begin{aligned}I'(t) &= \frac{1}{t-1} \left(\int_0^\infty \frac{dx}{x^2 + 1} - \int_0^\infty \frac{dx}{x^2 + t} \right) \\ &= \frac{1}{t-1} \left(\int_0^\infty \frac{dx}{x^2 + 1} - \frac{1}{\sqrt{t}} \int_0^\infty \frac{dy}{y^2 + 1} \right) = \frac{\pi}{2} \frac{1}{t-1} \left(1 - \frac{1}{\sqrt{t}} \right) \\ &\quad (\text{Subst. } x = \sqrt{t} y, dx = \sqrt{t} dy)\end{aligned}$$

Example (cont'd)

Using $t - 1 = (\sqrt{t} + 1)(\sqrt{t} - 1)$ this can be rewritten as

$$\begin{aligned} I'(t) &= \frac{\pi}{2} \frac{1}{(\sqrt{t} + 1)\sqrt{t}}. \\ \implies I(t) &= \pi \log(\sqrt{t} + 1) + C, \quad C \in \mathbb{R}. \end{aligned}$$

$I(0) = 0$ gives $C = 0$, so that finally $I(t) = \pi \log(\sqrt{t} + 1)$ and

$$\int_0^\infty \frac{\ln(x^2 + 1)}{x^2 + 1} dx = I(1) = \pi \ln(2).$$

Question: Does this complete the proof?

Answer: Not at all! Several things need to be justified:

- ① Why is $I(t)$ continuous at $t = 0$?
- ② Why is $I(t)$ continuous at $t = 1$?
- ③ Why are we allowed to differentiate under the integral sign?

It is obvious that our computation would be wrong if $I(t)$ had a jump discontinuity at $t = 0$ or $t = 1$, because we inferred the constant C and the value $I(1)$ by letting $t \rightarrow 0$, resp., $t \rightarrow 1$.

Example (cont'd)

(1) and (2) are remedied by showing that $I(t)$ is continuous on $[0, \infty)$ with the aid of Lebesgue Dominated Convergence

Theorem. For this we need to find an integrable bound $\Phi(x)$ for the integrand of $I(t)$, which doesn't depend on t . Such a bound cannot be found on the whole interval $[0, \infty)$, but it suffices to find it in a neighborhood of any “fixed” $t_0 \in [0, \infty)$.

Given t_0 , let $R > \max(t_0, 1)$. Then $[0, R)$ is a neighborhood of t_0 (relative to the domain $[0, \infty)$ of $I(t)$), and for $t \in [0, R)$ we have

$$\frac{\ln(x^2 + t)}{x^2 + 1} \begin{cases} \leq \frac{\ln(x^2 + R)}{x^2 + 1}, \\ \geq \frac{\ln(x^2)}{x^2 + 1} = \frac{2 \ln x}{x^2 + 1}. \end{cases}$$

It then follows that

$$\Phi(x) := \begin{cases} \frac{\ln(x^2 + R)}{x^2 + 1} & \text{for } x \geq 1, \\ -\frac{2 \ln x}{x^2 + 1} & \text{for } 0 \leq x < 1 \end{cases}$$

(with the convention $\Phi(0) := +\infty$), is a suitable integrable bound. (Integrability of Φ follows from the slow growth of the logarithm, both for $x \rightarrow \infty$ and $x \downarrow 0$.)

Example (cont'd)

For validating (3) in the same way, we need to find an integrable bound $\Phi(x)$ for the integrand of $I'(t)$, independent of t in some neighborhood of any “fixed” $t_0 > 0$. (At $t = 0$ the function $I(t)$ is not differentiable and we can do nothing.)

Given t_0 , let $0 < \delta < t_0$. Then (δ, ∞) is a neighborhood of t_0 , and for $t \in (\delta, \infty)$ we have

$$\frac{1}{(x^2 + 1)(x^2 + t)} \leq \frac{1}{(x^2 + 1)(x^2 + \delta)} =: \Phi(x).$$

Clearly $\Phi(x)$ is integrable over $[0, \infty)$, and hence Lebesgue’s Theorem yields that $I'(t)$ is differentiable for $t > 0$ with

$$I'(t) = \int_0^\infty \frac{dx}{(x^2+1)(x^2+t)}.$$

Of course this also proves the continuity of $I(t)$ at $t = 1$, but for $t = 0$ the proof on the previous slide is still necessary.

Finally, note that the formula $I(t) = I(0) + \int_0^t I'(s) ds$, which was also used during the proof, doesn’t require differentiability of $I(t)$ at $t = 0$ (only continuity).

Example (cont'd)

An example of what can happen with such parameter integrals is *Dirichlet's discontinuous integral*

$$I(a) = \int_0^\infty \frac{\sin(ax)}{x} dx, \quad a \in \mathbb{R}.$$

This integral is conditionally convergent as an improper Riemann integral for every $a \in \mathbb{R}$, but cannot be evaluated by differentiating under the integral sign:

$$I'(a) = \int_0^\infty \frac{d}{da} \frac{\sin(ax)}{x} dx = \int_0^\infty \cos(ax) dx$$

doesn't exist for $a \neq 0$.

However, from the known value $I(1) = \int_0^\infty \frac{\sin(x)}{x} dx = \pi/2$ we can easily determine

$$I(a) = \int_0^\infty \frac{\sin y}{y/a} \frac{1}{a} dy = \frac{\pi}{2} \quad (\text{Subst. } y = ax, dy = a dx)$$

for $a > 0$, and similarly $I(a) = -\pi/2$ for $a < 0$. Since $I(0) = 0$, the values of $I(a)$ for $a > 0$, say, won't give you any clue about $I(0)$.

There exist similar counterexamples with Lebesgue integrable functions.

Example (Newton Potential)

The Newton potential of a closed and bounded solid $S \subset \mathbb{R}^3$ with integrable density function $\rho \geq 0$ is, up to a normalizing factor, equal to

$$u(\mathbf{x}) = \int_S \frac{\rho(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|} d^3\mathbf{y}, \quad \mathbf{x} \in \mathbb{R}^3 \setminus S.$$

We know from an earlier exercise that the Newton potential of a point mass located in $\mathbf{y} \in \mathbb{R}^3$, i.e., the function $\mathbb{R}^3 \setminus \{\mathbf{y}\} \rightarrow \mathbb{R}$, $\mathbf{x} \mapsto \frac{1}{|\mathbf{x}-\mathbf{y}|}$ (up to a normalizing factor), is a solution of Laplace's Equation (i.e., a harmonic function).

Question

Does $\Delta u = 0$ hold as well?

Answer

Yes, it does. If we are allowed to differentiate under the integral sign, it would follow immediately:

$$\begin{aligned} \Delta u &= \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_3^2} = \int_S \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \right) \frac{\rho(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|} d^3\mathbf{y} \\ &= \int_S 0 = 0. \end{aligned}$$

Example (cont'd)

In order to justify this computation, we need to bound the (1st- and) 2nd-order partial derivatives of $f(\mathbf{x}, \mathbf{y}) = \frac{\rho(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|}$ with respect to \mathbf{x} .

Given $\mathbf{x}^{(0)} \in \mathbb{R}^3 \setminus S$, we choose $d > 0$ such that $|\mathbf{x} - \mathbf{y}| \geq d$ for all $\mathbf{x} \in B_d(\mathbf{x}^{(0)})$ and $\mathbf{y} \in S$.

The necessary partial derivatives of $1/r = 1/|\mathbf{x} - \mathbf{y}|$ are

$$\frac{\partial}{\partial x_i} \left(\frac{1}{r} \right) = \frac{-1}{r^3} (x_i - y_i),$$

$$\frac{\partial^2}{\partial x_i^2} \left(\frac{1}{r} \right) = \frac{-1}{r^3} + \frac{3}{r^5} (x_i - y_i)^2,$$

$$\frac{\partial^2}{\partial x_i \partial x_j} \left(\frac{1}{r} \right) = \frac{3}{r^5} (x_i - y_i)(x_j - y_j) \quad \text{for } i \neq j.$$

This gives the bounds

$$\left| \frac{\partial f}{\partial x_i} \right| \leq \frac{1}{d^2} \rho(\mathbf{y}), \quad \left| \frac{\partial^2 f}{\partial x_i \partial x_j} \right| \leq \frac{4}{d^3} \rho(\mathbf{y})$$

for $\mathbf{x} \in B_d(\mathbf{x}^{(0)})$ and $i, j \in \{1, 2, 3\}$.

Example (cont'd)

Since ρ is integrable, these are sufficient for applying Part (2) of the theorem. (As remarked earlier, the bounds for $\left| \frac{\partial f}{\partial x_i} \right|$ are redundant.)

As an application of Part (1) of the theorem, we will now show that for any unit vector \mathbf{x}

$$u(r\mathbf{x}) = \frac{M}{r} + o\left(\frac{1}{r}\right) \quad \text{for } r \rightarrow +\infty, \quad \text{where } M = \int_S \rho(\mathbf{y}) d^3\mathbf{y}.$$

This says that for points at sufficiently large distance from S (or from the origin of \mathbb{R}^3) the potential is well-approximated by that of a point with the same mass as S .

$$r \cdot u(r\mathbf{x}) = \int_S \frac{r \cdot \rho(\mathbf{y})}{|r\mathbf{x} - \mathbf{y}|} d^3\mathbf{y} = \int_S \frac{\rho(\mathbf{y})}{|\mathbf{x} - \xi\mathbf{y}|} d^3\mathbf{y}, \quad \text{where } \xi := 1/r$$

Provided that we are allowed to take the limit under the integral sign, we get

$$\lim_{r \rightarrow +\infty} (r \cdot u(r\mathbf{x})) = \lim_{\xi \downarrow 0} \left(\int_S \frac{\rho(\mathbf{y})}{|\mathbf{x} - \xi\mathbf{y}|} d^3\mathbf{y} \right) = \int_S \frac{\rho(\mathbf{y})}{|\mathbf{x}|} d^3\mathbf{y} = M,$$

since $|\mathbf{x}| = 1$. This proves our assertion.

Example (cont'd)

In order to justify this operation, we have to bound the integrand

$$g(\xi, \mathbf{y}) = \frac{\rho(\mathbf{y})}{|\mathbf{x} - \xi\mathbf{y}|}, \quad (\xi, \mathbf{y}) \in [0, \delta) \times S,$$

for some $\delta > 0$ by an integrable function $\Phi(\mathbf{y})$.

This is done in a similar way as before. Since S is bounded, there exists $R > 0$ such that $S \subset B_R(\mathbf{0})$. For $\xi \leq 1/2R$ we then have

$$|\xi\mathbf{y}| \leq 1/2.$$

$$\implies |\mathbf{x} - \xi\mathbf{y}| \geq 1 - 1/2 = 1/2 \implies |g(\xi, \mathbf{y})| \leq 2\rho(y).$$

Hence we can take $\delta = 1/2R$ and $\Phi(\mathbf{y}) = 2\rho(y)$.

Theorem (FUBINI's General Theorem)

Suppose that $f: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ is integrable. Then the integral $F(\mathbf{y}) = \int_{\mathbb{R}^m} f(\mathbf{x}, \mathbf{y}) d^m \mathbf{x}$ exists for almost all $\mathbf{y} \in \mathbb{R}^n$. Moreover, extending F to \mathbb{R}^n in any way, we have that F is integrable and

$$\int_{\mathbb{R}^n} F(\mathbf{y}) d^n \mathbf{y} = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f(\mathbf{x}, \mathbf{y}) d^m \mathbf{x} \right) d^n \mathbf{y} = \int_{\mathbb{R}^m \times \mathbb{R}^n} f(\mathbf{x}, \mathbf{y}) d^{m+n}(\mathbf{x}, \mathbf{y}).$$

Notes

- An analogous theorem holds for $G(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{x}, \mathbf{y}) d^n \mathbf{y}$.
- $\mathbb{R}^m \times \mathbb{R}^n$ can be identified with \mathbb{R}^{m+n} , of course.
- Applying Fubini's Theorem repeatedly reduces the computation of an n -dimensional integral $\int_{\mathbb{R}^n} f(\mathbf{x}) d^n \mathbf{x}$ to the computation of 1-dimensional integrals.
- The order of variables in the iterated integral is arbitrary, so that there are $n!$ ways to evaluate an n -dimensional integral as an iteration of 1-dimensional integrals.

Example

We prove the identity

$$\int_{\mathbb{R}^2} e^{-x^2-y^2} d^2(x, y) = \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2$$

in a direct way:

Assuming that $(x, y) \mapsto e^{-x^2-y^2}$ is integrable over \mathbb{R}^2 , we can apply Fubini's Theorem and obtain

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-x^2-y^2} d^2(x, y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-x^2-y^2} dx dy = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-x^2} e^{-y^2} dx dy \\ &= \int_{\mathbb{R}} e^{-y^2} \int_{\mathbb{R}} e^{-x^2} dx dy = \int_{\mathbb{R}} e^{-x^2} dx \int_{\mathbb{R}} e^{-y^2} dy \\ &= \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2, \end{aligned}$$

using the formula $\int(cf) = c \int f$, $c \in \mathbb{R}$, both for the inner and outer integral. The same reasoning proves that in general $\int_{\mathbb{R}^2} f(x)g(y) d^2(x, y) = \int_{\mathbb{R}} f(x) dx \int_{\mathbb{R}} g(y) dy$.

Example (cont'd)

Of course integrability of $(x, y) \mapsto e^{-x^2-y^2}$ must be shown first. For this we can look back to our earlier proof involving the sets $A_k = [-k, k]^2$ (rendering the direct proof rather useless), or argue in the same way with the disks $B_k = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq k^2\}$. For (x, y) in the annulus $B_{k+1} \setminus B_k$ we have $e^{-x^2-y^2} \leq e^{-k^2}$ and hence

$$\int_{B_{k+1} \setminus B_k} e^{-x^2-y^2} d^2(x, y) \leq e^{-k^2} \text{vol}(B_{k+1} \setminus B_k) = e^{-k^2} (2k+1)\pi.$$

It follows that

$$\int_{B_k} e^{-x^2-y^2} d^2(x, y) \leq \pi \sum_{i=0}^{k-1} (2i+1)e^{-i^2} < \pi \sum_{i=0}^{\infty} (2i+1)e^{-i^2}.$$

This series clearly converges and provides a bound for the integrals over B_k that is independent of k , as needed for applying the corollary to the Monotone Convergence Theorem. Hence $(x, y) \mapsto e^{-x^2-y^2}$ is integrable over \mathbb{R}^2 .

Example

We compute the volume β_n of the n -dimensional unit ball

$$\overline{B_1(\mathbf{0})} = \{\mathbf{x} \in \mathbb{R}^n; x_1^2 + \cdots + x_n^2 \leq 1\}.$$

Since $\overline{B_1(\mathbf{0})}$ is closed and bounded, it is measurable and so the volume exists. Moreover, the open unit ball $B_1(\mathbf{0})$ has the same volume as $\overline{B_1(\mathbf{0})}$, since the boundary

$S^{n-1} = S_1(\mathbf{0}) = \{\mathbf{x} \in \mathbb{R}^n; x_1^2 + \cdots + x_n^2 = 1\}$ is a smooth hypersurface in \mathbb{R}^n and hence $\text{vol}(S^{n-1}) = 0$.

Applying Fubini's Theorem, we obtain

$$\begin{aligned}\beta_n &= \text{vol}(B_1(\mathbf{0})) = \int_{x_1^2 + \cdots + x_n^2 \leq 1} 1 d^n \mathbf{x} \\ &= \int_{-1}^1 \left(\int_{x_1^2 + \cdots + x_{n-1}^2 \leq 1 - x_n^2} 1 d^{n-1}(x_1, \dots, x_{n-1}) \right) dx_n \\ &= \int_{-1}^1 \text{vol}\left(B'_{\sqrt{1-x_n^2}}(\mathbf{0})\right) dx_n,\end{aligned}$$

Example (cont'd)

where $B'_{\sqrt{1-x_n^2}}(\mathbf{0})$ denotes the open ball of radius $\sqrt{1-x_n^2}$ around $\mathbf{0} \in \mathbb{R}^{n-1}$.

Lemma

For a measurable set $A \subset \mathbb{R}^n$ and $r \in \mathbb{R}^+$ the set $rA = \{r\mathbf{x}; \mathbf{x} \in A\}$ is also measurable and satisfies $\text{vol}(rA) = r^n \text{vol}(A)$.

Proof.

This is a special case of the general change-of-variables theorem for the Lebesgue integral (to be discussed later). It obviously holds for n -dimensional (closed, say) intervals $Q = \prod_{i=1}^n [a_i, b_i]$, in which case $\text{vol}(rQ) = \text{vol}(\prod_{i=1}^n [ra_i, rb_i]) = \prod_{i=1}^n (ra_i - rb_i) = r^n \prod_{i=1}^n (a_i - b_i) = r^n \text{vol}(Q)$, and extends to arbitrary measurable sets by step function approximation. □

The lemma gives

$$\text{vol}\left(B'_{\sqrt{1-x_n^2}}(\mathbf{0})\right) = (1 - x_n^2)^{\frac{n-1}{2}} \text{vol}(B'_1(\mathbf{0})) = (1 - x_n^2)^{\frac{n-1}{2}} \beta_{n-1}$$

and hence

Example (cont'd)

$$\begin{aligned}\beta_n &= \int_{-1}^1 (1 - x_n^2)^{\frac{n-1}{2}} \beta_{n-1} dx_n = \beta_{n-1} \int_{-1}^1 (1 - x^2)^{\frac{n-1}{2}} dx \\ &= \beta_{n-1} \int_0^\pi \sin^n t dt && (\text{substitution } x = \cos t) \\ &= 2\beta_{n-1} \int_0^{\pi/2} \sin^n t dt\end{aligned}$$

The integrals $S_n = \int_0^{\pi/2} \sin^n t dt$, $n = 0, 1, 2, \dots$, can be evaluated using integration by parts,

$$\begin{aligned}S_n &= \int_0^\pi (1 - \cos^2 t) \sin^{n-2} t dt = S_{n-2} - \int_0^{\pi/2} \sin^{n-2} t \cos^2 t dt \\ &= S_{n-2} - \frac{1}{n-1} \left[\sin^{n-1} t \cos t \right]_0^{\pi/2} - \frac{1}{n-1} \int_0^{\pi/2} \sin^n t dt \\ &\quad u' = \sin^{n-2} t \cos t, v = \cos t \\ &= S_{n-2} - \frac{1}{n-1} S_n \quad \Rightarrow \quad S_n = \frac{n-1}{n} S_{n-2}\end{aligned}$$

Example (cont'd)

Together with $S_0 = \pi/2$, $S_1 = 1$ it follows that

$$S_{2k} = \frac{(2k-1)(2k-3)\cdots 3 \cdot 1}{2k(2k-2)\cdots 4 \cdot 2} \frac{\pi}{2},$$

$$S_{2k+1} = \frac{2k(2k-2)\cdots 4 \cdot 2}{(2k+1)(2k-1)\cdots 5 \cdot 3},$$

and

$$\beta_n = 2\beta_{n-1} S_n = 4\beta_{n-2} S_{n-1} S_n = \frac{2\pi}{n} \beta_{n-2}.$$

Finally, using $\beta_1 = 2$, $\beta_2 = \pi$ we get

$$\beta_{2k} = \frac{1}{k!} \pi^k, \quad \beta_{2k+1} = \frac{2^{k+1}}{(2k+1)(2k-1)\cdots 3 \cdot 1} \pi^k.$$

n	1	2	3	4	5	6	7	8	9	10
β_n	2	π	$\frac{4}{3}\pi$	$\frac{1}{2}\pi^2$	$\frac{8}{15}\pi^2$	$\frac{1}{6}\pi^3$	$\frac{16}{105}\pi^3$	$\frac{1}{24}\pi^4$	$\frac{32}{945}\pi^4$	$\frac{1}{120}\pi^5$
\approx	2	3.14	4.19	4.93	5.26	5.17	4.72	4.06	3.30	2.55

Remark (Cavalieri's Principle)

In the preceding computation we have used that the volume of an n -dimensional body B can be obtained by first computing the volume of all $(n - 1)$ -dimensional “cross-sections”

$$B_y = \{\mathbf{x} \in \mathbb{R}^{n-1}; (x_1, \dots, x_{n-1}, y) \in B\}$$

and then integrating the resulting function $y \mapsto \text{vol}(B_y)$ (which must be defined for almost all $y \in \mathbb{R}$) over \mathbb{R} :

$$\text{vol}(B) = \int_{\mathbb{R}} \text{vol}(B_y) \, dy.$$

In particular this implies that two 3-dimensional bodies B and B' , whose horizontal cross-sections B_z, B'_z for all $z \in \mathbb{R}$ have the same area, must have the same volume. This fact is named *Cavalieri's Principle* after B. CAVALIERI (1598–1647), but was used already by ARCHIMEDES (287–212 BC) in his computation of the volume of a sphere.

Locally Integrable Functions

Definition

$f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ is said to be *locally integrable* if every $\mathbf{x} \in \mathbb{R}^n$ has a neighborhood $U_{\mathbf{x}}$ (e.g., a ball $B = B_r(\mathbf{x})$ of radius $r > 0$) over which f is integrable.

One can show that the following properties are equivalent:

- 1 f is locally integrable;
- 2 f is integrable over every bounded open subset of \mathbb{R}^n ;
- 3 f is integrable over every compact (i.e., bounded and closed) subset of \mathbb{R}^n ;
- 4 f is integrable over $B_k(\mathbf{0})$ for every positive integer k .

The proof uses the Heine-Borel covering property of compact sets K (finitely many of the neighborhoods $U_{\mathbf{x}}$, $\mathbf{x} \in K$, cover K) and some rather obvious facts regarding integration over subsets of \mathbb{R}^n (such as “if f is integrable over A and $B \subseteq A$ is a measurable subset then f is integrable over B ” and “if f is integrable over sets A and B then f is integrable over $A \cup B$ ”).

Exercise

Suppose A, B are measurable subsets of \mathbb{R}^n and $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ is a function. Show:

- ① If f is integrable over A and $B \subseteq A$ then f is integrable over B .
- ② If f is integrable over A and B then f is integrable over $A \cup B$.

Hint: $\chi_{A \cap B} = \chi_A \chi_B$ and $\chi_{A \cup B} = \chi_A + \chi_B - \chi_A \chi_B$.

Exercise

Prove the equivalence of the four properties stated on the previous slide.

Example

Every continuous function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is locally integrable, as is every function $g: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ which arises from a continuous function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ by changing the values on a null set (set of measure zero).

For this consider the functions $f_k = f\chi_{B_k}$, $B_k = B_k(\mathbf{0})$. Choose an n -dimensional interval $Q_k \supseteq B_k$. Since f_k is continuous on Q_k except possibly for points on the boundary of B_k , which is a null set, we can easily show that f_k is Riemann integrable over Q_k , and hence Lebesgue integrable. Thus f satisfies Property 4.

Proposition

For $f: \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ the following are equivalent:

- ① f is integrable;
- ② f is locally integrable and $\|f\|_1 < \infty$.

Proof.

If f is integrable then f is integrable over every measurable subset of \mathbb{R}^n , in particular over every ball. Conversely, if f is locally integrable, it is integrable over $B_k = B_k(\mathbf{0})$, $k \in \mathbb{N}$, and $\int_{B_k} |f| = \|f\chi_{B_k}\|_1 \leq \|f\|_1$ independently of k .

Proof cont'd.

\Rightarrow We can apply the corollary to the Monotone Convergence Theorem to conclude that f is integrable over \mathbb{R}^n . (For this note that $B_k \subset B_{k+1}$ and $\bigcup_{k=1}^{\infty} B_k = \mathbb{R}^n$.) □

The following theorem, known as *Tonelli's Theorem*, provides sort of a converse to Fubini's Theorem:

Theorem (TONELLI)

Suppose $f: \mathbb{R}^m \times \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ is locally integrable and at least one of the iterated integrals

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} |f(\mathbf{x}, \mathbf{y})| d^m \mathbf{x} d^n \mathbf{y} \quad \text{or} \quad \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} |f(\mathbf{x}, \mathbf{y})| d^n \mathbf{y} d^m \mathbf{x}$$

exists. Then f is integrable (and, consequently, both iterated integrals exist and have the same value).

Existence, e.g., of the first iterated integral means the following:
 For all $\mathbf{y} \in \mathbb{R}^n$ except those in a null set N the function
 $\mathbf{x} \rightarrow |f(\mathbf{x}, \mathbf{y})|$ is integrable over \mathbb{R}^m , and the so-defined function
 $F(\mathbf{y}) = \int_{\mathbb{R}^m} |f(\mathbf{x}, \mathbf{y})| d^m \mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n \setminus N$, is integrable over \mathbb{R}^n .

Example

Inspecting our direct proof of $\int_{\mathbb{R}^2} e^{-x^2-y^2} d^2(x, y) = \left(\int_{\mathbb{R}} e^{-x^2} dx \right)^2$, we see that $(x, y) \mapsto e^{-x^2-y^2}$ satisfies the assumptions in Tonelli's Theorem, and hence is integrable over \mathbb{R}^2 , provided $\int_{\mathbb{R}} e^{-x^2} dx$ exists. Thus our subsequent justification by means of the Monotone Convergence Theorem can be spared.

Change of Variables

First recall the substitution formula for 1-dimensional integrals:

$$\int_a^b f(g(x))g'(x) \, dx = \int_{g(a)}^{g(b)} f(y) \, dy. \quad (\text{S})$$

The usual proof uses the Fundamental Theorem of Calculus and requires f to be continuous and g to be a C^1 -function.

If we require in addition that $g'(x) \neq 0$ for $x \in [a, b]$, then g induces a bijection from $[a, b]$ to $[g(a), g(b)]$ (if g is increasing) or $[g(b), g(a)]$ (if g is decreasing), and (S) can be rewritten as

$$\int_{[a,b]} f(g(x)) |g'(x)| \, dx = \int_{g([a,b])} f(y) \, dy.$$

The substitution rule extends to other forms of intervals (non-closed, unbounded) and to the case where g' doesn't exist at the endpoints (such as $g(x) = \sqrt{x}$ for $a = 0$). For the improper Riemann integrals involved, individual proofs need to be given in each case. In Lebesgue integration theory all cases are covered by the general change-of-variables formula.

Definition

Suppose $U, V \subseteq \mathbb{R}^n$ are open sets (i.e., $U^\circ = U, V^\circ = V$). A map $T: U \rightarrow V$ is called a **diffeomorphism** if

- 1 T is a bijection;
- 2 both T and its inverse map $T^{-1}: V \rightarrow U$ are C^1 -maps (i.e., continuously differentiable).

Remark

The condition on the differentiability of T^{-1} can be replaced by $\text{rk } \mathbf{J}_T(\mathbf{x}) = n$ for $\mathbf{x} \in U$, which implies local C^1 -invertibility of T by the Implicit Function Theorem.

Theorem (Change of Variables)

Let $U, V \subseteq \mathbb{R}^n$ be open sets and $T: U \rightarrow V$ a diffeomorphism.

Then a function $f: V \rightarrow \mathbb{R}$ is integrable over V if and only if $\mathbf{x} \mapsto f(T(\mathbf{x})) |\det \mathbf{J}_T(\mathbf{x})|$ is integrable over U . If this is the case, we have

$$\int_U f(T(\mathbf{x})) |\det \mathbf{J}_T(\mathbf{x})| d^n \mathbf{x} = \int_V f(\mathbf{y}) d^n \mathbf{y}.$$

Notes

- You can memorize the change-of-variables formula as “setting $\mathbf{y} = T(\mathbf{x})$, $d\mathbf{y} = |\det \mathbf{J}_T(\mathbf{x})| d\mathbf{x}$ ” (omitting n).
- The function $U \rightarrow \mathbb{R}$, $\mathbf{x} \mapsto \det \mathbf{J}_T(\mathbf{x})$ is called the *functional determinant* or *Jacobian* of T . This definition makes, more generally, sense for any map $T: D \rightarrow \mathbb{R}^n$, $D \subseteq \mathbb{R}^n$, and every point $\mathbf{x} \in D$ at which T is differentiable.
- The 1-dimensional cases of the change-of-variables formula discussed before are covered by the theorem (writing $T = g$), because for any function $f: I \rightarrow \mathbb{R}$ on an interval $I \subseteq \mathbb{R}$ we have $\int_I f = \int_{I^\circ} f$ (the one or two endpoints of a non-open interval have Lebesgue measure zero!).
- The special case $f = \chi_A$, $A \subseteq U$ measurable, implies that $T(A) \subseteq V$ is measurable as well and

$$\text{vol}(T(A)) = \int_{T(A)} 1 d^n \mathbf{y} = \int_A |\det \mathbf{J}_T(\mathbf{x})| d^n \mathbf{x}.$$

In particular, if $T(\mathbf{x}) = \mathbf{T}\mathbf{x}$ is linear then $\mathbf{x} \mapsto \mathbf{J}_T(\mathbf{x}) = \mathbf{T}$ is constant on U and we get $\text{vol}(T(A)) = |\det \mathbf{T}| \text{vol}(A)$.

Notes cont'd

- If $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a Euclidean motion, i.e., $T(\mathbf{x}) = \mathbf{T}\mathbf{x} + \mathbf{b}$ for some orthogonal matrix $\mathbf{T} \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$, then $\det \mathbf{J}_T(\mathbf{x}) = \det \mathbf{T} = \pm 1$ and the change-of-variables formula takes the form

$$\int_{\mathbb{R}^n} f(\mathbf{T}\mathbf{x} + \mathbf{b}) d^n \mathbf{x} = \int_{\mathbb{R}^n} f(\mathbf{x}) d^n \mathbf{x}.$$

For a measurable set $A \subseteq \mathbb{R}^n$ this implies $\text{vol}(T(A)) = \text{vol}(A)$.

Thus the Lebesgue measure and Lebesgue integral are invariant under Euclidean motions. In particular the Lebesgue measure is translation-invariant—a property, which together with the normalization requirement $\text{vol}([0, 1]^n) = 1$ characterizes the Lebesgue measure on \mathbb{R}^n .

A Very Rough Sketch of the Proof

In Linear Algebra we have seen that the volume of the parallelepiped P spanned by $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$ (with one vertex at the origin) is given by $\text{vol}(P) = |\det \mathbf{A}|$, where $\mathbf{A} = (\mathbf{a}_1 | \dots | \mathbf{a}_n)$. If T is linear, $T(\mathbf{x}) = \mathbf{T}\mathbf{x}$, the image $T(P)$ is a parallelepiped as well and is spanned by the columns of $\mathbf{T}\mathbf{A}$.

$$\implies \text{vol } T(P) = |\det(\mathbf{T}\mathbf{A})| = |\det \mathbf{T}| |\det \mathbf{A}| = |\det \mathbf{T}| \text{vol}(P)$$

The same formula holds for a general parallelepiped $\mathbf{b} + \mathbf{P}$.

Now suppose T is a not necessarily linear diffeomorphism. In this case the idea is to decompose P into many small parallelepipeds $\mathbf{x} + Q_i$, $1 \leq i \leq M$. For a small parallelepiped $\mathbf{x} + Q$ we have

$$\begin{aligned} T(\mathbf{x} + Q) &= \{T(\mathbf{x} + \mathbf{h}); \mathbf{h} \in Q\} \\ &= \{T(\mathbf{x}) + dT(\mathbf{x})(\mathbf{h}) + o(\mathbf{h}); \mathbf{h} \in Q\} \\ &\approx T(\mathbf{x}) + dT(\mathbf{x})(Q), \end{aligned}$$

which is again a parallelepiped and has volume

$$|\det \mathbf{J}_T(\mathbf{x})| \text{vol}(Q).$$

Sketch of Proof Cont'd

Adding the various volumes gives

$$\text{vol } T(P) \approx \sum_{i=1}^M |\det \mathbf{J}_T(\mathbf{x}_i)| \text{ vol}(Q_i),$$

and in the limit $M \rightarrow \infty$

$$\text{vol } T(P) = \int_P |\det \mathbf{J}_T(\mathbf{x})| d^n \mathbf{x}.$$

Using characteristic functions this can be written as

$$\int_{T(P)} 1 d^n \mathbf{y} = \int_P 1 \cdot |\det \mathbf{J}_T(\mathbf{x})| d^n \mathbf{x}.$$

The general change-of-variables formula follows from this using the approximation of integrable functions by step functions.

One-dimensional Examples

Example

Recall that $\int_1^\infty \frac{dx}{x^s} = \frac{1}{s-1}$ for $s > 1$ and $\int_0^1 \frac{dx}{x^t} = \frac{1}{1-t}$ for $t < 1$. (For other values of s, t these integrals don't exist.)

Using the change of variables $y = 1/x$, $x \in (1, \infty)$, we can transform the first integral into the second:

$$\begin{aligned}\int_1^\infty \frac{dx}{x^s} &= \int_0^1 \frac{y^s}{y^2} dy && (\text{Subst. } x = 1/y, dx = |-1/y^2| dy) \\ &= \int_0^1 y^{s-2} dy.\end{aligned}$$

This is of the above type with $t = 2 - s < 1$, and we have indeed

$$\frac{1}{1-t} = \frac{1}{1-(2-s)} = \frac{1}{s-1}.$$

With the Riemann integral, the substitution would be $x = 1/y$, $dx = -\frac{1}{y^2} dy$, giving the intermediate form $\int_1^0 -y^{s-2} dy$, which then happily reduces to the same result as above. For $s \geq 2$ the new Riemann integral becomes even a proper one.

With the Lebesgue integral we just integrate over $(0, 1)$.

One-dimensional Examples

Cont'd

Example

We show that $\int_{-\infty}^{+\infty} e^{-x^2} dx = 2 \int_0^{\infty} e^{-x^2} dx = \Gamma(1/2)$.

To this end, we apply the change-of-variables formula with $U = V = (0, +\infty)$ and $x = T(t) = \sqrt{t}$, for which

$$|T'(t)| = T'(t) = \frac{1}{2\sqrt{t}}.$$

$$\begin{aligned}\int_{-\infty}^{+\infty} e^{-x^2} dx &= 2 \int_0^{\infty} \frac{e^{-t}}{2\sqrt{t}} dt = \int_0^{\infty} \frac{e^{-t}}{\sqrt{t}} dt \\ &= \int_0^{\infty} t^{1/2-1} e^{-t} dt = \Gamma(1/2).\end{aligned}$$

Of course you can continue using the mnemonic $x = \sqrt{t}$, $dx = \frac{1}{2\sqrt{t}} dt$ in the 1-dimensional case.

Polar Coordinates

The change-of-variables theorem of Lebesgue integration theory is an extremely powerful tool for evaluating integrals, as you will see. We start with the following

Corollary

$f: \mathbb{R}^2 \rightarrow \mathbb{R}$ is integrable if and only if $(r, \phi) \mapsto r \cdot f(r \cos \phi, r \sin \phi)$ is integrable over $(0, +\infty) \times (0, 2\pi)$. If this is the case, we have

$$\int_{\mathbb{R}^2} f(x, y) d^2(x, y) = \int_{(0, +\infty) \times (0, 2\pi)} r \cdot f(r \cos \phi, r \sin \phi) d^2(r, \phi).$$

Proof.

$T(r, \phi) = \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix}$ defines a diffeomorphism from $(0, +\infty) \times (0, 2\pi)$ to $\mathbb{R}^2 \setminus \{(x, y); x \geq 0\}$, since it is bijective and its jacobian $\det \mathbf{J}_T(r, \phi) = \det \begin{pmatrix} \cos \phi & -r \sin \phi \\ \sin \phi & r \cos \phi \end{pmatrix} = r$ is nonzero everywhere. Since the non-negative real axis $\mathbb{R}_0^+ = \{(x, 0); x \geq 0\}$ has measure zero, the corollary follows directly from the change-of-variables theorem. □

Notes

- The function f doesn't need to be defined everywhere; "almost everywhere" is enough. Similarly, we can extend the domain of integration on the right-hand side to the closed "horizontal strip" $[0, +\infty) \times [0, 2\pi]$, since the boundary of the strip has measure zero.
- The corollary extends to integration over measurable subsets A of \mathbb{R}^2 via the replacement $f \rightarrow f\chi_A$:

$$\begin{aligned} \int_A f(x, y) d^2(x, y) &= \int_{\mathbb{R}^2} f(x, y) \chi_A(x, y) d^2(x, y) \\ &= \int_{(0,+\infty) \times (0,2\pi)} r \cdot f(r \cos \phi, r \sin \phi) \chi_A(r \cos \phi, r \sin \phi) d^2(r, \phi) \\ &= \int_{T^{-1}(A)} r \cdot f(r \cos \phi, r \sin \phi) d^2(r, \phi). \end{aligned}$$

That $T^{-1}(A)$ doesn't include preimages for the elements of A on \mathbb{R}_0^+ (if any) poses no problem, since $\text{vol}(\mathbb{R}_0^+) = 0$.

Example

We compute the area of the (open) unit disk

$B = B_1(0, 0) \subset \mathbb{R}^2$ using polar coordinates.

Via $T(r, \phi) = \begin{pmatrix} r \cos \phi \\ r \sin \phi \end{pmatrix}$ the unit disk corresponds to

$$T^{-1}(B) = \{(r, \phi) \in \mathbb{R}^2; 0 < r \leq 1, 0 < \phi < 2\pi\}.$$

The corollary gives

$$\begin{aligned} \text{vol}(B) &= \int_B 1 \, d^2(x, y) = \int_{[0,1] \times [0,2\pi]} r \, d^2(r, \phi) \\ &= \int_0^1 \int_0^{2\pi} r \, dr \, d\phi = \int_0^1 2\pi r \, dr = 2\pi \left[\frac{r^2}{2} \right]_0^1 = \pi. \end{aligned}$$

The area of the closed unit disk \bar{B} is the same, since the boundary $\partial B = S^1$ (the unit circle) is a smooth 1-dimensional surface and hence has area zero.

We compute the Gauss integral $\int_{-\infty}^{+\infty} e^{-x^2} dx$.

$$\begin{aligned}
 \left(\int_{-\infty}^{+\infty} e^{-x^2} dx \right)^2 &= \left(\int_{\mathbb{R}} e^{-x^2} dx \right) \left(\int_{\mathbb{R}} e^{-y^2} dy \right) \\
 &= \int_{\mathbb{R}^2} e^{-(x^2+y^2)} d^2(x, y) \quad (\text{Fubini}) \\
 &= \int_{(0,\infty) \times (0,2\pi)} r e^{-r^2} d^2(r, \phi) \\
 &= 2\pi \int_0^{\infty} r e^{-r^2} dr = 2\pi \left[-\frac{1}{2} e^{-r^2} \right]_0^{+\infty} = \pi
 \end{aligned}$$

$$\implies \int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$$

Since the Gauss integral is equal to $\Gamma(1/2)$, we also have $\Gamma(1/2) = \sqrt{\pi}$. From this and the functional equation $\Gamma(x+1) = x \Gamma(x)$ you can recursively compute $\Gamma(n/2)$ for all odd positive integers n .

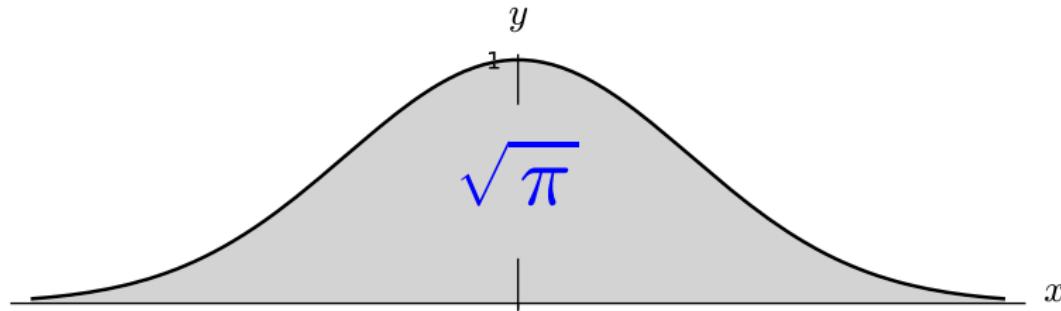


Figure: The Gauss integral has value $\sqrt{\pi}$.

From this one can derive by an easy change of variables that the probability density function of the normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$, viz.,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{has} \quad \int_{-\infty}^{\infty} f(x) dx = 1.$$

This explains the normalization factor $1/(\sigma\sqrt{2\pi})$ in f .

Spherical Coordinates

Spherical coordinates (also called 3-dimensional polar coordinates) are defined as follows:

Definition

The spherical coordinates of a point $(x, y, z) \in \mathbb{R}^3 \setminus \{\mathbf{0}\}$ are (r, θ, ϕ) , where $r = \sqrt{x^2 + y^2 + z^2} > 0$ (distance from the origin), $\theta \in [0, 2\pi]$ is the angle assigned to (x, y) in plane polar coordinates, and $\phi \in [0, \pi]$ is the angle between (x, y, z) and $\mathbf{e}_3 = (0, 0, 1)$ (i.e., $z = r \cos \phi$ and $\sqrt{x^2 + y^2} = r \sin \phi$). In terms of (r, θ, ϕ) the cartesian coordinates (x, y, z) are expressed as

$$x = r \cos \theta \sin \phi, \quad y = r \sin \theta \sin \phi, \quad z = r \cos \phi.$$

According to the definition, the origin $(0, 0, 0) \in \mathbb{R}^3$ has no spherical coordinates, but we can of course amend it by saying that the origin is given by $r = 0$.

Corollary

$f: \mathbb{R}^3 \rightarrow \mathbb{R}$ is integrable (over \mathbb{R}^3) if and only if

$$(r, \theta, \phi) \mapsto f(r \cos \theta \sin \phi, r \sin \theta \sin \phi, r \cos \phi) r^2 \sin \phi$$

is integrable over $U = (0, +\infty) \times (0, 2\pi) \times (0, \pi)$. If this is the case, we have

$$\int_{\mathbb{R}^3} f(x, y, z) d^3(x, y, z) = \int_U f(r \cos \theta \sin \phi, r \sin \theta \sin \phi, r \cos \phi) r^2 \sin \phi d^3(r, \theta, \phi).$$

Proof.

The transformation $T(r, \theta, \phi) = \begin{pmatrix} r \cos \theta \sin \phi \\ r \sin \theta \sin \phi \\ r \cos \phi \end{pmatrix}$ defines a diffeomorphism from U to $\mathbb{R}^3 \setminus \{(x, 0, z); x \geq 0\}$ (i.e., half of the (x, z) -plane is excluded), since it is bijective and its Jacobian

$$\det \mathbf{J}_T(r, \theta, \phi) = \det \begin{pmatrix} \cos \theta \sin \phi & -r \sin \theta \sin \phi & r \cos \theta \cos \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \phi & 0 & -r \sin \phi \end{pmatrix} = -r^2 \sin \phi$$

is nonzero everywhere on U . Since the excluded set is contained a plane, it has measure zero, and the corollary follows again directly from the change-of-variables theorem. □

Notes

- The notes on the first corollary (integration using 2-dimensional polar coordinates) carry over mutatis mutandis to the present situation.
- Compare the statement of our second corollary with the much less general (yet more involved) Formula 3 in [Ste21], Ch. 15.8, p. 1104).

Example

We recompute the volume $\beta_3 = \text{vol}(B_1(\mathbf{0}))$ of the (closed) unit ball in \mathbb{R}^3 using spherical coordinates.

Since $B_1(\mathbf{0})$ is given by $r \leq 1$ in spherical coordinates, T maps $Q = (0, 1) \times (0, 2\pi) \times (0, \pi)$ to $B_1(\mathbf{0})$, up to some set of measure zero.

$$\begin{aligned}\beta_3 &= \int_{B_1(\mathbf{0})} 1 \, d^3(x, y, z) = \int_Q r^2 \sin \phi \, d^3(r, \theta, \phi) \\ &= \left(\int_0^1 r^2 \, dr \right) \left(\int_0^{2\pi} 1 \, d\theta \right) \left(\int_0^\pi \sin \phi \, d\phi \right) = \frac{4\pi}{3}.\end{aligned}$$

Rotation-Invariant Functions

Definition

$f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, is said to be *rotation-invariant* if there exists a function $g: I \rightarrow \mathbb{R}$ defined on a subset $I \subseteq [0, +\infty)$ and such that

$$f(\mathbf{x}) = g(|\mathbf{x}|) = g\left(\sqrt{x_1^2 + \cdots + x_n^2}\right).$$

W.l.o.g. the domain of f may be taken as the maximal one, which is the spherical shell $S(I) = \{\mathbf{x} \in \mathbb{R}^n; |\mathbf{x}| \in I\}$.

For the cases $n = 2, 3$ the change-of-variables theorem gives

$$\int_{S(I)} f(\mathbf{x}) d^2 \mathbf{x} = \int_{I \times (0, 2\pi)} g(r) r d^2(r, \phi) = 2\pi \int_I g(r) r dr,$$

$$\int_{S(I)} f(\mathbf{x}) d^3 \mathbf{x} = \int_{I \times (0, 2\pi) \times (0, \pi)} g(r) r^2 \sin \phi d^3(r, \theta, \phi) = 4\pi \int_I g(r) r^2 dr,$$

whereby the integrability of f over $S(I)$ is equivalent to the integrability of $r \mapsto g(r)r$, resp., $r \mapsto g(r)r^2$ over I .

Generalization

Suppose $I \subseteq [0, +\infty)$ is measurable (for example, an interval), $g: I \rightarrow \mathbb{R}$ is any function, and $f: S(I) \rightarrow \mathbb{R}$, $S(I) \subseteq \mathbb{R}^n$, is defined by $f(\mathbf{x}) = g(|\mathbf{x}|)$. Then f is integrable over the (measurable) set $S(I)$ iff $r \mapsto g(r)r^{n-1}$ is integrable over I . If this is the case, we have

$$\int_{S(I)} f(\mathbf{x}) d^n \mathbf{x} = n\beta_n \int_I g(r)r^{n-1} dr.$$

Example

Find the volume of the cone C in \mathbb{R}^3 with base $x^2 + y^2 \leq R^2$ and apex (tip) in $(0, 0, h)$.

The cone is bounded by the (x, y) -plane and the graph G_f of the rotation-invariant function $f(x, y) = h(1 - r/R)$. (You can think of G_f as being obtained by rotating the 1-dimensional graph $z = h(1 - x/R)$, $0 \leq x \leq R$ (a line segment) around the z -axis. The above formula for $n = 2$ gives

$$\begin{aligned} \text{vol}(C) &= \int_{x^2+y^2 \leq R^2} f(x, y) d^2(x, y) = 2\pi \int_0^R h(1 - r/R)r dr \\ &= 2\pi h \left[\frac{r^2}{2} - \frac{r^3}{3R} \right]_0^R = \frac{1}{3} R^2 \pi h. \end{aligned}$$

Example (cont'd)

Since the horizontal sections of a cone are circles, $\text{vol}(C)$ can be also be determined by the method of Cavalieri:

$$\begin{aligned}\text{vol}(C) &= \int_{\mathbb{R}} \text{vol}(C_z) dz = \int_0^h (R(1 - z/h))^2 \pi dz \\ &= R^2 \pi \int_0^h (1 - z/h)^2 dz \\ &= R^2 \pi \int_0^1 w^2 |-h| dw \quad (\text{substitution } w = 1 - z/h) \\ &= \frac{1}{3} R^2 \pi h,\end{aligned}$$

where the substitution rule for 1-dimensional integrals has been applied in the form formulated in the change-of-variables theorem ($w = T(z) = 1 - z/h$, $dw = |T'(z)| dz = \frac{1}{h} dz$, $T([0, h]) = [0, 1]$).

Example

We compute the volume of

$$B = \{(x, y, z) \in \mathbb{R}^3; 1 \leq z \leq 1/\sqrt{x^2 + y^2}\}$$

B is the region between the plane $z = 1$ and the surface that is generated by rotating the curve $z = 1/x \wedge 0 < x < 1$ around the z -axis. Note that the area between the curve and the z -axis, viz. $\int_1^\infty \frac{dz}{z}$, is infinite.

Writing $f(x, y) = \frac{1}{\sqrt{x^2 + y^2}}$ and $g(r) = 1/r$, we have

$f(x, y) = g(\sqrt{x^2 + y^2})$ and hence

$$\begin{aligned} \text{vol}(B) &= \int_{x^2+y^2 \leq 1} f(x, y) d^2(x, y) - \text{vol}(\{(x, y, z); x^2 + y^2 \leq 1, 0 \leq z \leq 1\}) \\ &= 2\pi \int_0^1 g(r)r dr - \pi = 2\pi - \pi = \pi, \quad \text{since } g(r)r = 1. \end{aligned}$$

Thus, quite surprisingly, the volume of B is finite and equal to that of its “socle”, i.e., the cylinder with base the unit disk and height 1. (Cavalieri’s method gives $\text{vol}(B) = \int_1^\infty (1/z)^2 \pi dz = \pi[-1/z]_1^\infty = \pi$.)

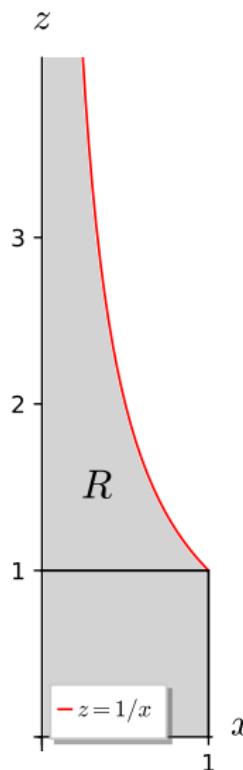


Figure: The curve (or plane region R) generating B by rotating around the z -axis; in the same way, the bottom square generates the “socle” of B

Math 241
Calculus III

Thomas
Honold

Vector Fields
and
Differential
Forms

Line Integrals

The
Fundamental
Theorem for
Line Integrals

Locally Exact
1-Forms

Line Integrals
in Complex
Analysis
(optional)

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Vector Fields and Differential Forms

2 Line Integrals

3 The Fundamental Theorem for Line Integrals

4 Locally Exact 1-Forms

5 Line Integrals in Complex Analysis (optional)

Math 241
Calculus III

Thomas
Honold

Vector Fields
and
Differential
Forms

Line Integrals

The
Fundamental
Theorem for
Line Integrals

Locally Exact
1-Forms

Line Integrals
in Complex
Analysis
(optional)

Today's Lecture: Vector Analysis

Vector Fields

Definition

A *vector field* is a mapping $F: D \rightarrow \mathbb{R}^n$ with domain $D \subseteq \mathbb{R}^n$.

Example

We define $F_1, F_2: \mathbb{R}^2 \setminus \{(0,0)\} \rightarrow \mathbb{R}^2$ by

$$F_1(x, y) = \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} x \\ y \end{pmatrix}, \quad F_2(x, y) = \frac{1}{\sqrt{x^2 + y^2}} \begin{pmatrix} -y \\ x \end{pmatrix}$$

The field F_1 attaches to each point $(x, y) \in \mathbb{R}^2 \setminus \{(0,0)\}$ a unit vector with the same direction as (x, y) , and F_2 attaches to (x, y) the vector $F_1(x, y)$ rotated by 90° counterclock-wise.

Example (cont'd)

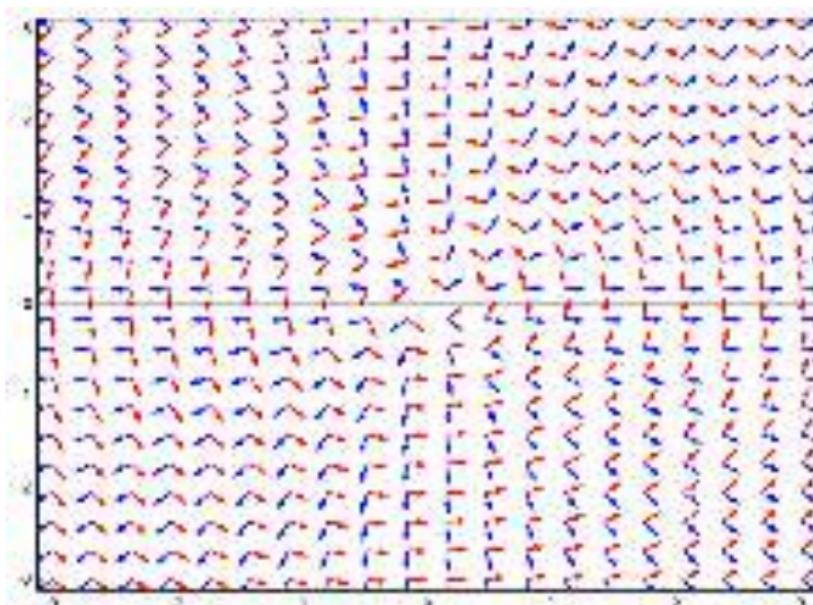


Figure: The vector fields F_1 (blue) and F_2 (red)
Lengths are scaled by a positive constant

Example (Gravitational field/electric field)

Newton's Law of Gravitation $\mathbf{F} = -\frac{GMm}{r^3} \mathbf{r} = -\frac{GMm}{r^2} \frac{\mathbf{r}}{r}$ says that up to a positive normalizing constant the gravitational force exerted by a mass located at the origin on another mass has the form

$$G(x, y, z) = -\frac{1}{(x^2 + y^2 + z^2)^{3/2}} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (x, y, z) \in \mathbb{R}^3 \setminus \{\mathbf{0}\},$$

where (x, y, z) denotes the position vector of the second mass. Accordingly, G is called *gravitational field*.

Similarly, by Coulomb's Law the electric force exerted by a charge located at the origin on another charge has the form $\pm G(x, y, z)$, depending on the signs of the two charges.

Example (Winding field)

The 2-dimensional vector field $W(x, y) = \frac{1}{x^2 + y^2} \begin{pmatrix} -y \\ x \end{pmatrix}$,

$(x, y) \in \mathbb{R}^2$, is called *winding field*. It is similar to example F_2 , except that the lengths $|W(x, y)| = \frac{1}{\sqrt{x^2 + y^2}}$ of its vectors are inversely proportional to the distance from (x, y) to the origin.

Example (Gradient fields)

A differentiable function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, gives rise to the vector field $D \rightarrow \mathbb{R}^n$, $\mathbf{x} \mapsto \nabla f(\mathbf{x})$. Such vector fields are called *gradient fields* or *conservative* vector fields with associated *potential function* f .

The vector fields

$$F_1(x, y) = \nabla f(x, y), \quad f(x, y) = \sqrt{x^2 + y^2} \quad \text{for} \quad \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \setminus \{\mathbf{0}\},$$

$$G(x, y, z) = \nabla g(x, y, z), \quad g(x, y, z) = \frac{1}{\sqrt{x^2 + y^2 + z^2}} \quad \text{for} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \setminus \{\mathbf{0}\},$$

are gradient fields.

The vector fields F_2 and W are not gradient fields, as we will see.

Differential 1-Forms

Recall that a (real) *linear form* on \mathbb{R}^n is a linear map from \mathbb{R}^n to $\mathbb{R} = \mathbb{R}^1$; “linear” means that the map λ should satisfy

$$\lambda(\mathbf{x} + \mathbf{y}) = \lambda(\mathbf{x}) + \lambda(\mathbf{y}), \quad \lambda(c\mathbf{x}) = c\lambda(\mathbf{x}) \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \text{ and } c \in \mathbb{R}.$$

Linear forms are in 1-1 correspondence $\lambda \leftrightarrow \mathbf{a}$ with vectors in \mathbb{R}^n via $\lambda(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} = \mathbf{a}^\top \mathbf{x} = a_1 x_1 + \dots + a_n x_n$, $\mathbf{x} \in \mathbb{R}^n$.

The set of all linear forms on \mathbb{R}^n , equipped with addition

$$(\lambda_1 + \lambda_2)(\mathbf{x}) = \lambda_1(\mathbf{x}) + \lambda_2(\mathbf{x}) \text{ and scalar multiplication}$$

$(c\lambda)(\mathbf{x}) = c\lambda(\mathbf{x})$, $(\mathbf{x} \in \mathbb{R}^n, c \in \mathbb{R})$, is called the *dual space* of \mathbb{R}^n and denoted by $(\mathbb{R}^n)^*$.

Definition

A map with domain $D \subseteq \mathbb{R}^n$ and codomain $(\mathbb{R}^n)^*$, is called a *differential 1-form* on D .

Thus a differential 1-form $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, assigns to each $\mathbf{x} \in D$ a linear form $\omega(\mathbf{x}): \mathbb{R}^n \rightarrow \mathbb{R}$.

Example

The standard example is provided by the differentials

$$df: D \rightarrow (\mathbb{R}^n)^* \text{ of differentiable functions } f: D \rightarrow \mathbb{R}, D \subseteq \mathbb{R}^n.$$

Notes

- Differential 1-forms are vector fields “in disguise”: For $\mathbf{x} \in D$ there exists a unique vector $F(\mathbf{x}) \in \mathbb{R}^n$ such that

$$\omega(\mathbf{x})(\mathbf{h}) = F(\mathbf{x}) \cdot \mathbf{h} = F(\mathbf{x})^\top \mathbf{h}, \quad \mathbf{h} \in \mathbb{R}^n.$$

In this way, every differential 1-form $\omega: D \rightarrow (\mathbb{R}^n)^*$ gives rise to a vector field $F: D \rightarrow \mathbb{R}^n$, and conversely every vector field determines a differential 1-form with the same domain. Moreover, these correspondences are mutually inverse.

- More generally, a *differential k-form* on $D \subseteq \mathbb{R}^n$ is a map from D to the space of so-called alternating k -fold linear (k -multilinear) forms on \mathbb{R}^n . A differential 0-form on D is just a function $f: D \rightarrow \mathbb{R}$, and a differential n -form has the form $\omega(\mathbf{x})(\mathbf{h}_1, \dots, \mathbf{h}_n) = f(\mathbf{x}) \det(\mathbf{h}_1, \dots, \mathbf{h}_n)$, where $\det(\mathbf{h}_1, \dots, \mathbf{h}_n)$ denotes the determinant of the matrix in $\mathbb{R}^{n \times n}$ with columns $\mathbf{h}_1, \dots, \mathbf{h}_n$. Differential 2-forms on $D \subseteq \mathbb{R}^3$ have the form

$$\omega(\mathbf{x})(\mathbf{u}, \mathbf{v}) = \det \begin{pmatrix} f_1(\mathbf{x}) & u_1 & v_1 \\ f_2(\mathbf{x}) & u_2 & v_2 \\ f_3(\mathbf{x}) & u_3 & v_3 \end{pmatrix} \text{ with } f_i: D \rightarrow \mathbb{R}. \text{ One writes this as } \omega = f_1 \, dy \wedge dz + f_2 \, dz \wedge dx + f_3 \, dx \wedge dy.$$

Standard Representation

Recall that the differential df of a differentiable function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, admits the representation

$$df = \frac{\partial f}{\partial x_1} dx_1 + \cdots + \frac{\partial f}{\partial x_n} dx_n$$

This generalizes to the following

Observation

For every differential 1-form ω on $D \subseteq \mathbb{R}^n$ there exist uniquely determined functions $f_i: D \rightarrow \mathbb{R}$, $1 \leq i \leq n$, such that for $\mathbf{x} \in D$ we have

$$\begin{aligned}\omega(\mathbf{x})(\mathbf{h}) &= f_1(\mathbf{x})h_1 + \cdots + f_n(\mathbf{x})h_n \\ &= f_1(\mathbf{x}) dx_1(\mathbf{h}) + \cdots + f_n(\mathbf{x}) dx_n(\mathbf{h}), \quad \mathbf{h} \in \mathbb{R}^n.\end{aligned}$$

Conversely, every n -tuple of functions $f_i: D \rightarrow \mathbb{R}$ determines a differential 1-form on D in this way.

As above we write this as $\omega = f_1 dx_1 + \cdots + f_n dx_n$. The functions f_i are just the coordinate functions of the vector field F associated with ω , i.e., $F = (f_1, \dots, f_n)$.

Example

The differential form representations of our vector field examples are

$$\omega_1 = \omega_{F_1} = \frac{x}{\sqrt{x^2 + y^2}} dx + \frac{y}{\sqrt{x^2 + y^2}} dy = \frac{x dx + y dy}{\sqrt{x^2 + y^2}},$$

$$\omega_2 = \omega_{F_2} = \frac{-y}{\sqrt{x^2 + y^2}} dx + \frac{x}{\sqrt{x^2 + y^2}} dy = \frac{x dy - y dx}{\sqrt{x^2 + y^2}},$$

$$\omega_3 = \omega_G = -\frac{x dx + y dy + z dz}{(x^2 + y^2 + z^2)^{3/2}},$$

$$\omega_4 = \omega_W = \frac{x dy - y dx}{x^2 + y^2}.$$

Complex Differential 1-Forms on $D \subseteq \mathbb{C}$

Recall that the differential of a holomorphic function $f: D \rightarrow \mathbb{C}$, $D \subseteq \mathbb{C}$, admits the representation $df(z) = f'(z) dz$.

Definition

A *complex differential 1-form* on $D \subseteq \mathbb{C}$ is a mapping $\omega: D \rightarrow \mathbb{C}^*$, where \mathbb{C}^* is the dual space of \mathbb{C} .

Since $\lambda \in \mathbb{C}^*$ must satisfy $\lambda(h) = \lambda(h1) = h\lambda(1)$, \mathbb{C}^* consists of all multiplication maps $\mathbb{C} \rightarrow \mathbb{C}$, $h \mapsto ch$ with $c \in \mathbb{C}$.

⇒ For every complex differential 1-form on D there exists a unique function $f: D \rightarrow \mathbb{C}$ such that

$$\omega(z)(h) = f(z)h = f(z) dz(h), \quad h \in \mathbb{C}.$$

As usual we write this as $\omega = f dz$. (But note that the “product” $f dz$ now involves the multiplication in \mathbb{C} .)

Example

Consider $\omega = \frac{1}{z} dz$, defined on $\mathbb{C} \setminus \{0\}$.

Using the identification $\mathbb{C} \triangleq \mathbb{R}^2$, we can write ω as

$$\begin{aligned}\omega &= \frac{\bar{z}}{|z|^2} (dx + i dy) = \frac{x - iy}{x^2 + y^2} (dx + i dy) \\ &= \frac{x dx + y dy}{x^2 + y^2} + i \frac{x dy - y dx}{x^2 + y^2}.\end{aligned}$$

The says that the real and imaginary parts of ω are ordinary (i.e., real) differential 1-forms on $\mathbb{C} \setminus \{0\} = \mathbb{R}^2 \setminus \{(0, 0)\}$.

In fact, the imaginary part of $\omega = \frac{1}{z} dz$ is just the “winding form” corresponding to the winding field defined earlier, and the real part corresponds to a gradient field:

$$F(x, y) = \frac{1}{x^2 + y^2} \begin{pmatrix} x \\ y \end{pmatrix} = \nabla f, \quad f(x, y) = \ln \sqrt{x^2 + y^2} \text{ for } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \setminus \{\mathbf{0}\}.$$

Line Integrals

Motivation

Line integrals were invented to model the work done by a variable force in moving an object from a point \mathbf{p} in 3-space to another point \mathbf{q} along a certain curve C . If C is the straight line segment $[\mathbf{p}, \mathbf{q}] = \{(1 - t)\mathbf{p} + t\mathbf{q}; 0 \leq t \leq 1\}$ and the force is a constant vector \mathbf{F} , the familiar $\text{work} = \text{force} \times \text{distance}$ translates into

$$W = |\mathbf{F}| \cos \phi \times |\mathbf{q} - \mathbf{p}| = \mathbf{F} \cdot (\mathbf{q} - \mathbf{p})$$

If the force is variable and the curve $\gamma: [a, b] \rightarrow \mathbb{R}^3$ is an arbitrary parametric curve, we can interpolate the curve by straight line segments connecting successive curve points $P_i = \gamma(t_i)$, $0 \leq i \leq N$ (as we did when defining arc length), evaluate the force on an “intermediate” point of the line segment, apply the preceding formula to each line segment $[P_{i-1}, P_i]$, and take the sum. It is then reasonable to define the work as the limit of this quantity as the “mesh size” tends to zero (provided the limit exists). In the following this idea of a line integral will be made precise and generalized to curves in \mathbb{R}^n .

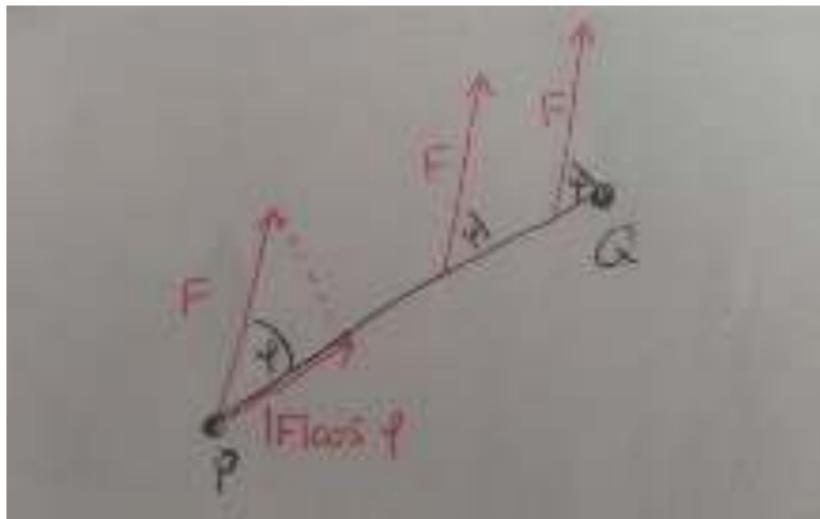


Figure: The work of a constant force field \mathbf{F} along a line segment $[P, Q]$ is $W = |\mathbf{F}| \cos \phi |Q - P| = \mathbf{F} \cdot (Q - P)$

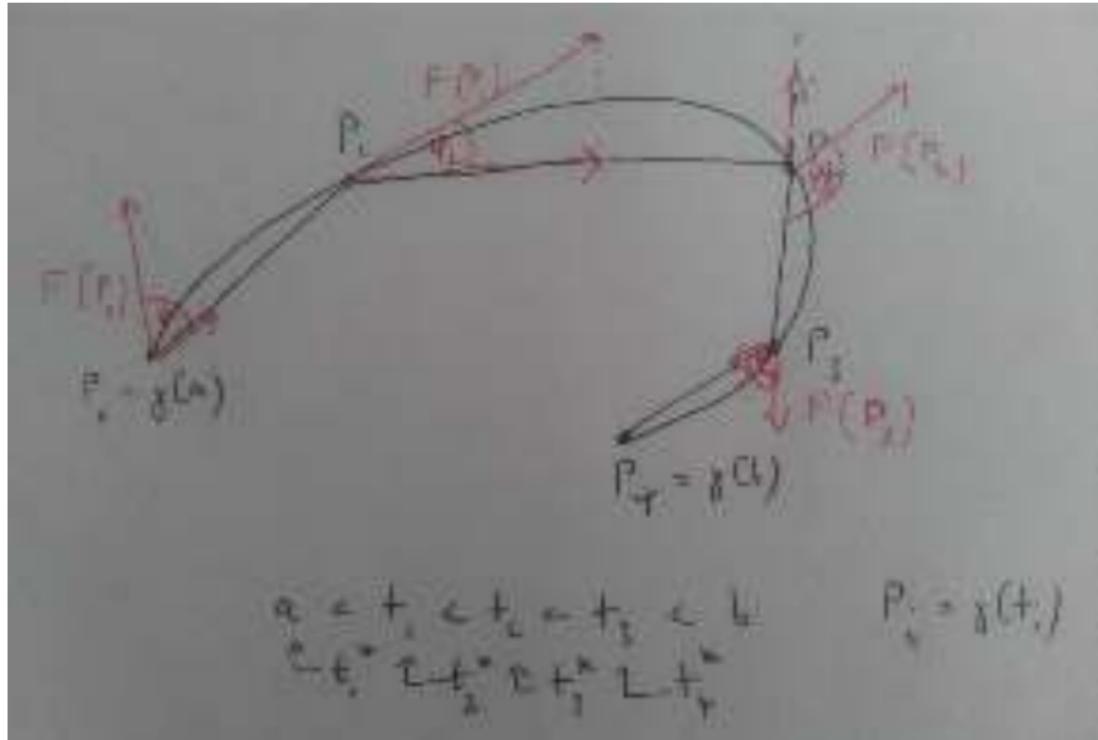


Figure: The general case with tags $t_i^* = t_{i-1}$ (i.e., the left endpoint of $[t_{i-1}, t_i]$ is used), which gives the approximation

$$W \approx \sum_{i=1}^4 |\mathbf{F}(P_{i-1})| \cos \phi_i |P_i - P_{i-1}| = \sum_{i=1}^4 \mathbf{F}(P_{i-1}) \cdot (P_i - P_{i-1})$$

Suppose $\gamma: [a, b] \rightarrow \mathbb{R}^n$ is a curve whose range $\gamma([a, b])$ is contained in the domain $D \subseteq \mathbb{R}^n$ of a vector field F or differential 1-form ω .

Definition

A *tagged partition* (or *tagged subdivision*) of $[a, b]$ is a pair (T, T^*) consisting of a partition $T = \{t_0, t_1, \dots, t_N\}$ of $[a, b]$ (as usual this entails $a = t_0 < t_1 < \dots < t_N = b$) and a set (or sequence) $T^* = \{t_1^*, \dots, t_N^*\}$ of tags satisfying $t_k^* \in [t_{k-1}, t_k]$. The *mesh size* of (T, T^*) is defined as the length of the largest subinterval among $[t_{k-1}, t_k]$, $1 \leq k \leq N$.

Furthermore we define

$$S_F(T, T^*) = \sum_{k=1}^N F(\gamma(t_k^*)) \cdot (\gamma(t_k) - \gamma(t_{k-1})),$$

$$S_\omega(T, T^*) = \sum_{k=1}^N \omega(\gamma(t_k^*)) (\gamma(t_k) - \gamma(t_{k-1})).$$

Note

If ω corresponds to F via $\omega(\mathbf{x})(\mathbf{h}) = F(\mathbf{x}) \cdot \mathbf{h}$ then

$S_\omega(T, T^*) = S_F(T, T^*)$ for any tagged partition (T, T^*) of $[a, b]$.

Definition (Line integral)

The vector field F (respectively, the differential 1-form ω) is said to be *integrable* along the curve γ with *line integral* $\int_{\gamma} F \cdot d\mathbf{r} = I \in \mathbb{R}$ (resp., $\int_{\gamma} \omega = I \in \mathbb{R}$), if for every $\epsilon > 0$ there exists $\delta > 0$ such that for all tagged partitions (T, T^*) of $[a, b]$ of mesh size $< \delta$ we have $|S_F(T, T^*) - I| < \epsilon$ (resp., $|S_{\omega}(T, T^*) - I| < \epsilon$).

Notes

- If ω corresponds to F as explained above, the equalities $S_{\omega}(T, T^*) = S_F(T, T^*)$ imply that $\int_{\gamma} F \cdot d\mathbf{r} = \int_{\gamma} \omega$, provided that one integral (and hence both integrals) exist.
- In [Ste21], Ch. 16.2, also line integrals $\int_{\gamma} f ds$ of functions $f: D \rightarrow \mathbb{R}$ with respect to arc length are defined. The above definition can be adapted to cover this case: Just use the sums

$$S_f(T, T^*) = \sum_{k=1}^N f(\gamma(t_k^*)) |\gamma(t_k) - \gamma(t_{k-1})|$$

in place of $S_F(T, T^*)$, $S_{\omega}(T, T^*)$. The arc length $L(\gamma) = \int_{\gamma} 1 ds$ is a special case of this.

Properties of Line Integrals

We discuss in detail only properties of line integrals of vector fields/differential 1-forms. As discussed we may restrict ourselves to the more convenient differential forms view.

The most important property is a formula involving the derivative $\gamma'(t)$ analogous to that for the arc length.

Theorem

Suppose $\omega = f_1 dx_1 + \cdots + f_n dx_n$ is continuous (i.e., the functions f_1, \dots, f_n are continuous on D) and $\gamma = (\gamma_1, \dots, \gamma_n) : [a, b] \rightarrow D$ is a piecewise C^1 -curve. Then ω is integrable along γ with line integral

$$\int_{\gamma} \omega = \int_a^b \omega(\gamma(t)) (\gamma'(t)) dt = \int_a^b \sum_{i=1}^n f_i(\gamma(t)) \gamma'_i(t) dt.$$

Proof.

Let $\epsilon > 0$ be given. Since the n functions $t \mapsto f_i(\gamma(t))$ are uniformly continuous on $[a, b]$ and $L = L(\gamma)$ is finite, there exists $\delta > 0$ such that for $1 \leq i \leq n$

$$|f_i(\gamma(t)) - f_i(\gamma(t^*))| < \frac{\epsilon}{nL} \quad \text{whenever } t, t^* \in [a, b] \text{ and } |t - t^*| < \delta.$$

Proof cont'd.

Now we use that for any subinterval $[a', b'] \subseteq [a, b]$ we have

$$\int_{a'}^{b'} \gamma'_i(t) dt = \gamma_i(b') - \gamma_i(a').$$

This is clear when γ'_i is continuous on $[a', b']$ and follows by subdivision in the general, piecewise continuous case. For example, if $c \in [a', b']$ is such that γ'_i is continuous on $[a', c]$ and $[c, b']$, then

$$\begin{aligned} \int_{a'}^{b'} \gamma'_i(t) dt &= \int_{a'}^c \gamma'_i(t) dt + \int_c^{b'} \gamma'_i(t) dt \\ &= \gamma_i(c) - \gamma_i(a') + \gamma_i(b') - \gamma_i(c) = \gamma_i(b') - \gamma_i(a') \end{aligned}$$

Writing $I = \int_a^b \omega(\gamma(t)) (\gamma'(t)) dt$, it follows that for any tagged partition (Z, Z^*) of $[a, b]$ of mesh size $< \delta$ we have

$$\begin{aligned} |S(Z, Z^*) - I| &= \left| \sum_{i=1}^n \sum_{k=1}^N \int_{t_{k-1}}^{t_k} (f_i(\gamma(t_k^*)) - f_i(\gamma(t))) \gamma'_i(t) dt \right| \\ &\leq \frac{\epsilon}{nL} \sum_{i=1}^n \int_a^b |\gamma'_i(t)| dt \leq \frac{\epsilon}{L} \int_a^b |\gamma(t)| dt = \epsilon. \quad \square \end{aligned}$$

Note

In terms of the vector field $\mathbf{F} = (f_1, \dots, f_n)$ associated to $\omega = f_1 dx_1 + \dots + f_n dx_n$, the formula in the theorem becomes

$$\int_{\gamma} \omega = \int_a^b \sum_{i=1}^n f_i(\gamma(t)) \gamma'_i(t) dt = \int_a^b \mathbf{F}(\gamma(t)) \cdot \gamma'(t) dt = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt,$$

motivating the notation $\int_{\gamma} \mathbf{F} \cdot d\mathbf{r}$.

Example

Let $\omega = y^2 dx + dy$ and $\gamma_{\alpha}(t) = (t, t^{\alpha})$ for $t \in [0, 1]$, where $\alpha > 0$ is a parameter.

$$\begin{aligned}\int_{\gamma_{\alpha}} \omega &= \int_0^1 \begin{pmatrix} t^{2\alpha} \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ \alpha t^{\alpha-1} \end{pmatrix} dt = \int_0^1 t^{2\alpha} + \alpha t^{\alpha-1} dt \\ &= \left[\frac{t^{2\alpha+1}}{2\alpha+1} + t^{\alpha} \right]_0^1 = \frac{1}{2\alpha+1} + 1.\end{aligned}$$

Note that the curves γ_{α} have the same starting point $\gamma_{\alpha}(0) = (0, 0)$ and the same end point $\gamma_{\alpha}(1) = (1, 1)$, but the integrals $\int_{\gamma_{\alpha}} \omega$ depend on the particular curve.

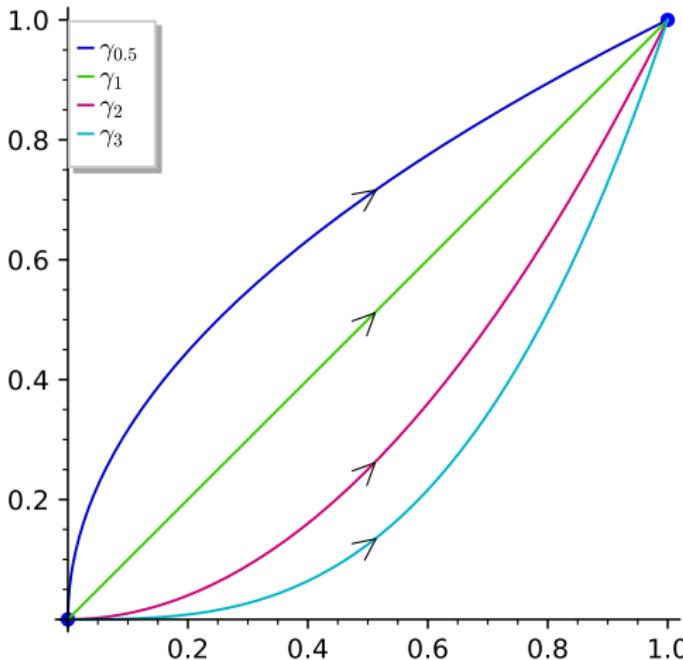


Figure: The curves γ_{α} from the preceding example

Some further terminology

- A piecewise C^1 -curve $\gamma: [a, b] \rightarrow \mathbb{R}^n$ is also referred to as a *path* (of integration).
- A path (or a general curve) $\gamma: [a, b] \rightarrow \mathbb{R}^n$ is *closed* if $\gamma(a) = \gamma(b)$, i.e., the starting point and end point of γ coincide.

Caution: This has nothing (well, almost nothing) to do with the concept of a closed subset of \mathbb{R}^n in the sense of topology (i.e., a subset containing its boundary or, equivalently, a subset for which the complementary set in \mathbb{R}^n is open).

Example

Consider the winding form $\omega = \frac{x \, dy - y \, dx}{x^2 + y^2}$ and a closed path $\gamma: [a, b] \rightarrow \mathbb{R}^2$ that does not contain the origin, i.e., $\gamma(t) \neq (0, 0)$ for $t \in [a, b]$. Writing $\gamma(t) = (x(t), y(t))$, we have

$$\int_{\gamma} \omega = \int_a^b \frac{x(t)y'(t) - y(t)x'(t)}{x(t)^2 + y(t)^2} \, dt = 2A,$$

where A is the oriented area enclosed by the normalized curve $\gamma / |\gamma|$ (normalized to unit length); cf. exercises.

Example (cont'd)

Intuitively it is clear that $A = m\pi$ (m times the area of the unit disk), where m counts how often the curve winds around the origin.

The number

$$n(\gamma; \mathbf{0}) = \frac{1}{2\pi} \int_{\gamma} \frac{x \, dy - y \, dx}{x^2 + y^2} = \int_a^b \frac{x(t)y'(t) - y(t)x'(t)}{x(t)^2 + y(t)^2} \, dt$$

is called *winding number* of γ relative to the origin $\mathbf{0} = (0, 0)$ of \mathbb{R}^2 . For an arbitrary point $\mathbf{p} = (p_1, p_2) \in \mathbb{R}^2$ one defines

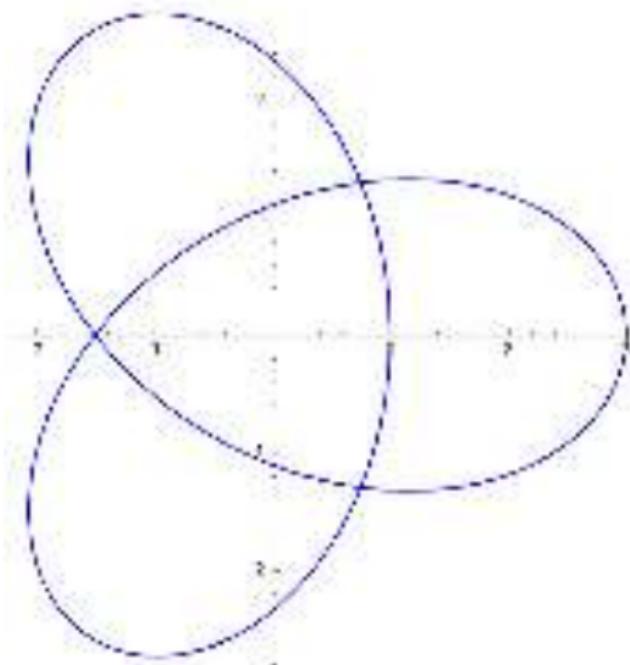
$n(\gamma; \mathbf{p}) = n(\gamma - \mathbf{p}; \mathbf{0})$, where $\gamma - \mathbf{p}$ denotes the translated curve $[a, b] \rightarrow \mathbb{R}^2$, $t \mapsto (x(t) - p_1, y(t) - p_2)$.

Example

The counterclock-wise traversed "trefoil" curve

$$\gamma(t) = (\cos t + 2 \cos(2t), -\sin t + 2 \sin(2t)), \quad t \in [0, 2\pi],$$

has winding number $n(\gamma; \mathbf{0}) = 2$.



Properties of Line Integrals Cont'd

Linearity

Suppose $\omega_1, \omega_2: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, are differential 1-forms, $\gamma: [a, b] \rightarrow D$ is a curve and $c_1, c_2 \in \mathbb{R}$. If $\int_{\gamma} \omega_1$ and $\int_{\gamma} \omega_2$ exist then so does $\int_{\gamma} (c_1 \omega_1 + c_2 \omega_2)$ and

$$\int_{\gamma} (c_1 \omega_1 + c_2 \omega_2) = c_1 \int_{\gamma} \omega_1 + c_2 \int_{\gamma} \omega_2.$$

Composition of paths

Suppose $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, is a differential 1-form, $\gamma: [a, b] \rightarrow D$ a curve and $c \in [a, b]$. Let $\gamma_1 = \gamma|_{[a, c]}$ and $\gamma_2 = \gamma|_{[c, b]}$. If $\int_{\gamma_1} \omega$ and $\int_{\gamma_2} \omega$ exist then so does $\int_{\gamma} \omega$ and

$$\int_{\gamma} \omega = \int_{\gamma_1} \omega + \int_{\gamma_2} \omega.$$

Properties of Line Integrals Cont'd

Reparametrization

Suppose $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, is a differential 1-form, $\gamma: [a, b] \rightarrow D$ a curve and $t: [\alpha, \beta] \rightarrow [a, b]$, $s \mapsto t(s)$ a continuous bijection. Then we have

$$\int_{\gamma \circ t} \omega = \pm \int_{\gamma} \omega;$$

“+” holds if t is increasing, and “−” holds if t is decreasing.

We can change the direction of γ (in particular interchange the starting and end point) by defining $\gamma^-(t) = \gamma(ta + (1 - t)b)$ for $t \in [0, 1]$. This is a decreasing reparametrization, and hence we have

$$\int_{\gamma^-} \omega = - \int_{\gamma} \omega.$$

The corresponding curve $\gamma^+(t) = \gamma((1 - t)a + tb)$, $t \in [0, 1]$ (“reparametrization of γ to the unit interval”), which satisfies $\gamma^+(0) = \gamma(a)$, $\gamma^+(1) = \gamma(b)$ arises from an increasing reparametrization and hence satisfies $\int_{\gamma^+} \omega = \int_{\gamma} \omega$.

Exercise

Prove the assertion about increasing/decreasing reparametrizations of line integrals.

Hint: Use the substitution rule for ordinary integrals from Calculus I/II.

The Fundamental Theorem for Line Integrals

Definition

A differential 1-form $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, is said to be *exact* if there exists a function $f: D \rightarrow \mathbb{R}^n$ satisfying $\omega = df$ (a so-called *antiderivative*). Equivalently, the vector field F corresponding to ω is a gradient field, viz. $F = \nabla f$.

Theorem

Suppose $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, is continuous and exact, $\omega = df$, and $\gamma: [a, b] \rightarrow D$ is any path. Then we have

$$\int_{\gamma} \omega = f(\gamma(b)) - f(\gamma(a)).$$

Proof.

Since $\omega(\mathbf{x}) = df(\mathbf{x})$ for all $\mathbf{x} \in D$ have

$$\int_{\gamma} \omega = \int_a^b df(\gamma(t))(\gamma'(t)) dt \equiv \int_a^b \nabla f(\gamma(t)) \cdot \gamma'(t) dt.$$

The integrand is continuous and is the derivative of $t \mapsto f(\gamma(t))$ (by the chain rule of multivariable calculus). Hence the Fundamental Theorem of Calculus I yields the result. □

Corollary

Suppose $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, is continuous and exact.

- 1 The integrals $\int_{\gamma} \omega$ are independent of path, i.e., for any two paths γ_1, γ_2 in D whose starting points and end points coincide we have $\int_{\gamma_1} \omega = \int_{\gamma_2} \omega$.
- 2 For every closed path γ in D we have $\int_{\gamma} \omega = 0$.

Remark

If $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, is continuous but not necessarily exact, then Conditions (1), (2) in the corollary are equivalent:

If (1) holds and $\gamma: [a, b] \rightarrow D$ is a closed path with $\gamma(a) = \gamma(b) = \mathbf{p}$, say, we can use the constant curve $\delta: [0, 1] \rightarrow D$, $t \mapsto \mathbf{p}$ to evaluate $\int_{\gamma} \omega = \int_{\delta} \omega = 0$.

Conversely, suppose that (2) holds and γ_1, γ_2 are paths in D with the same starting point and end point. After reparametrization we can assume that γ_1, γ_2 both have parameter interval $[0, 1]$.

Remark (cont'd)

Now define $\gamma = \gamma_1 \gamma_2^- : [0, 1] \rightarrow D$ (" γ_1 followed by the reverse of γ_2 ") as follows:

$$\gamma(t) = \begin{cases} \gamma_1(2t) & \text{if } 0 \leq t \leq \frac{1}{2}, \\ \gamma_2(2 - 2t) & \text{if } \frac{1}{2} \leq t \leq 1. \end{cases}$$

Since $\gamma(0) = \gamma_1(0) = \gamma_2(0) = \gamma(1)$, γ is a closed path and hence $\int_{\gamma} \omega = 0$.

On the other hand we have $\int_{\gamma} \omega = \int_{\gamma_1 \gamma_2^-} \omega = \int_{\gamma_1} \omega - \int_{\gamma_2} \omega$.

Both properties together imply $\int_{\gamma_1} \omega = \int_{\gamma_2} \omega$.

Example

We compute the line integrals $\int_{\gamma_1} \frac{x \, dx + y \, dy}{\sqrt{x^2+y^2}}$ and $\int_{\gamma_2} \frac{x \, dy - y \, dx}{x^2+y^2}$ for the two curves

$$\gamma_1(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}, \quad t \in [0, \pi/2], \quad \gamma_2(t) = \begin{pmatrix} 1-t \\ t \end{pmatrix}, \quad t \in [0, 1].$$

Both curves have initial point $(1, 0)$ and terminal point $(0, 1)$.

For this we will use the derivatives

$$\gamma'_1(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}, \quad \gamma'_2(t) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Since the first 1-form is exact, $\frac{x \, dx + y \, dy}{\sqrt{x^2+y^2}} = df$ for

$f(x, y) = \sqrt{x^2 + y^2} = |(\begin{pmatrix} x \\ y \end{pmatrix})|$, the first two integrals are easy.
By the Fundamental Theorem for Line Integrals,

$$\int_{\gamma_1} \frac{x \, dx + y \, dy}{\sqrt{x^2 + y^2}} = \int_{\gamma_2} \frac{x \, dx + y \, dy}{\sqrt{x^2 + y^2}} = \left| \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right| - \left| \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right| = 0.$$

Example (con't)

Now we verify this result by direct computation:

$$\int_{\gamma_1} \frac{x \, dx + y \, dy}{\sqrt{x^2 + y^2}} = \int_0^{\pi/2} \frac{\cos t(-\sin t) + \sin t \cos t}{\sqrt{\cos^2 t + \sin^2 t}} dt = 0,$$

$$\int_{\gamma_2} \frac{x \, dx + y \, dy}{\sqrt{x^2 + y^2}} = \int_0^1 \frac{(1-t)(-1) + t \cdot 1}{\sqrt{(1-t)^2 + t^2}} dt = \int_0^1 \frac{2t - 1}{\sqrt{2t^2 - 2t + 1}} dt$$

$$= \frac{1}{2} \int_{-1}^1 \frac{s}{\sqrt{\frac{1}{2}s^2 + \frac{1}{2}}} ds \quad (\text{substitution } s = 2t - 1)$$

$$= 0. \quad (\text{by symmetry})$$

For the integrals of the second 1-form, which is the by now familiar winding form, we obtain

$$\int_{\gamma_1} \frac{x \, dy - y \, dx}{x^2 + y^2} = \int_0^{\pi/2} \frac{\cos^2 t - \sin t(-\sin t)}{\cos^2 t + \sin^2 t} dt = \int_0^{\pi/2} 1 \, dt = \frac{\pi}{2},$$

$$\int_{\gamma_2} \frac{x \, dy - y \, dx}{x^2 + y^2} = \int_0^1 \frac{(1-t)1 - t(-1)}{(1-t)^2 + t^2} dt = \int_0^1 \frac{1}{2t^2 - 2t + 1} dt$$

Example (cont'd)

$$\begin{aligned}
 &= \frac{1}{2} \int_{-1}^1 \frac{1}{\frac{1}{2}s^2 + \frac{1}{2}} ds && (\text{substitution } s = 2t - 1) \\
 &= \int_{-1}^1 \frac{ds}{s^2 + 1} = [\arctan(s)]_{-1}^1 = \frac{\pi}{4} - \left(-\frac{\pi}{4}\right) = \frac{\pi}{2}.
 \end{aligned}$$

Hence in the second case

$$\int_{\gamma_1} \frac{x dy - y dx}{x^2 + y^2} = \int_{\gamma_2} \frac{x dy - y dx}{x^2 + y^2}$$

as well, although this is not a consequence of the Fundamental Theorem for Line Integrals.

Moreover, we see from the computation of the first integral that for any curve γ of the form $\gamma(t) = (\cos t, \sin t)$, $t \in [a, b]$, we have

$$\int_{\gamma} \frac{x dy - y dx}{x^2 + y^2} = b - a = L(\gamma).$$

If γ is closed (i.e. $b \in a + 2\pi\mathbb{Z}$), this integral is equal to $2\pi n(\gamma; \mathbf{0})$.

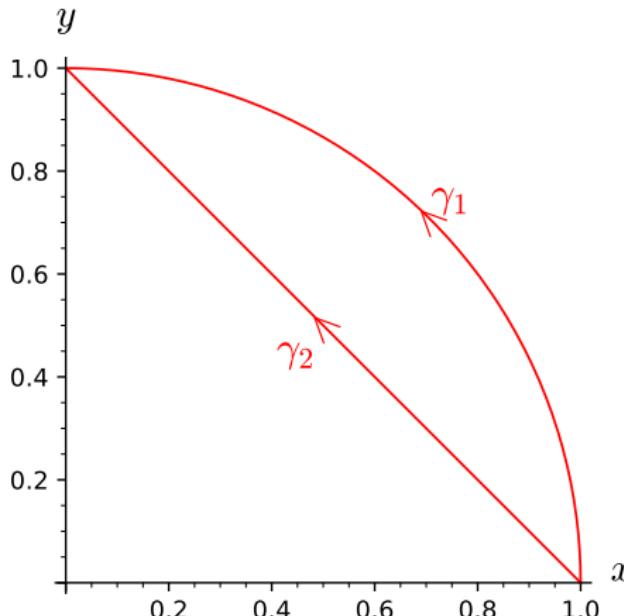


Figure: The two paths γ_1, γ_2 from the example

Example (taken from [Ste21])

Find the work W done when moving a particle in the gravitational field from the point $A = (2, 2, 0)$ along some path C to the point $B = (3, 4, 12)$.

Solution: The gravitational form $\omega_G = -\frac{x \, dx + y \, dy + z \, dz}{(x^2 + y^2 + z^2)^{3/2}}$ is exact with antiderivative $g(x, y, z) = (x^2 + y^2 + z^2)^{-1/2} = 1/r$.

According to Newton's 3rd Law of Motion the force exerted on the particle is the opposite of the gravitational force, viz., $-G(x, y, z)$, which has potential/antiderivative $-g(x, y, z) = -1/r$.

$$\begin{aligned} \implies W &= \int_C (-G) \cdot d\mathbf{r} = \int_C (-\omega_G) = -g(B) - (-g(A)) = \frac{1}{r_A} - \frac{1}{r_B} \\ &= \frac{1}{\sqrt{2^2 + 2^2}} - \frac{1}{\sqrt{3^2 + 4^2 + 12^2}} = \frac{1}{2\sqrt{2}} - \frac{1}{13} \end{aligned}$$

In the physical world this quantity has to be multiplied by mMG , where M denotes the mass generating the gravitational field, m the mass of the particle, and G the gravitational constant.

The Converse of the Corollary

Independence of path implies exactness

First observe/recall the following:

- If $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, is exact then D must be *open*, because $\omega(\mathbf{x}) = df(\mathbf{x})$ for $\mathbf{x} \in D$ implies in particular $D = D^\circ$.
- If in addition D is path-connected then any two antiderivatives f_1, f_2 of ω differ by a constant, since $\omega = df_1 = df_2$ implies $d(f_1 - f_2) = 0$ on D , and we have seen that in this case $f_1 - f_2$ must be constant.

Connected subsets of \mathbb{R}^n

In Topology a subset $D \subseteq \mathbb{R}^n$ is said to be *connected* if it cannot be partitioned into two nonempty relatively open subsets, i.e., there do not exist open sets $U, V \subseteq \mathbb{R}^n$ such that $U \cap D \neq \emptyset$, $V \cap D \neq \emptyset$, but $U \cap V \cap D = \emptyset$.

If D is open, this reduces to the non-existence of a partition into nonempty open sets and is equivalent to D being path-connected. From now on we shall use the more common term “connected”.

Theorem (cf. [Ste21], Ch. 16.3, Th. 4)

A continuous differential 1-form $\omega: D \rightarrow (\mathbb{R}^n)^*$ on a connected open set $D \subseteq \mathbb{R}^n$ is exact if and only if the integrals $\int_{\gamma} \omega$ are independent of path.

Proof.

Necessity was shown in the corollary to the Fundamental Theorem of Line Integrals. It remains to prove sufficiency.

Choose a fixed point $\mathbf{a} \in D$ and define for $\mathbf{x} \in D$

$$f(\mathbf{x}) := \int_{\mathbf{a}}^{\mathbf{x}} \omega := \int_{\gamma} \omega, \quad \text{where } \gamma(0) = \mathbf{a}, \gamma(1) = \mathbf{x},$$

and of course $\gamma: [0, 1] \rightarrow D$ is a path. The function $f: D \rightarrow \mathbb{R}$ is well-defined, since D is connected and $\int_{\gamma} \omega$ is independent of path.

It remains to show that the partial derivatives $\partial f / \partial x_i$ exist and are equal to the component functions f_i of $\omega = \sum_{i=1}^n f_i dx_i$. (Since ω is continuous, this implies that f is a C^1 -function, hence in particular differentiable.)

Proof cont'd.

In what follows we assume w.l.o.g. $n = 2$ (cf. the proof of Clairaut's Theorem).

Suppose $\mathbf{x} = (x, y) \in D$ and $r > 0$ is such that $B_r(x, y) \subseteq D$
 $\Rightarrow f(\mathbf{x} + h\mathbf{e}_1)$ is defined for all $h < r$ and can be computed using any path from \mathbf{a} to \mathbf{x} followed by the horizontal path $\beta: [0, h] \rightarrow D$, $t \mapsto \mathbf{x} + t\mathbf{e}_1$. Using $\int_{\mathbf{a}}^{\mathbf{x}+h\mathbf{e}_1} \omega = \int_{\mathbf{a}}^{\mathbf{x}} \omega + \int_{\beta} \omega$, we obtain

$$\begin{aligned}\frac{f(\mathbf{x} + h\mathbf{e}_1) - f(\mathbf{x})}{h} &= \frac{1}{h} \int_{\beta} \omega = \frac{1}{h} \int_{\beta} f_1 \, dx + f_2 \, dy \\ &= \frac{1}{h} \int_0^h f_1(\mathbf{x} + t\mathbf{e}_1) \, dt \\ &= f_1(\mathbf{x}) + \frac{1}{h} \int_0^h (f_1(\mathbf{x} + t\mathbf{e}_1) - f_1(\mathbf{x})) \, dt \\ &\rightarrow f_1(\mathbf{x}) \quad \text{for } h \rightarrow 0,\end{aligned}$$

where we have used continuity of f_1 at \mathbf{x} . (Note the similarity to the proof of the Fundamental Theorem of Calculus.)

This shows that $\partial f / \partial x = f_1$.

In the same way one proves that $\partial f / \partial y = f_2$. □

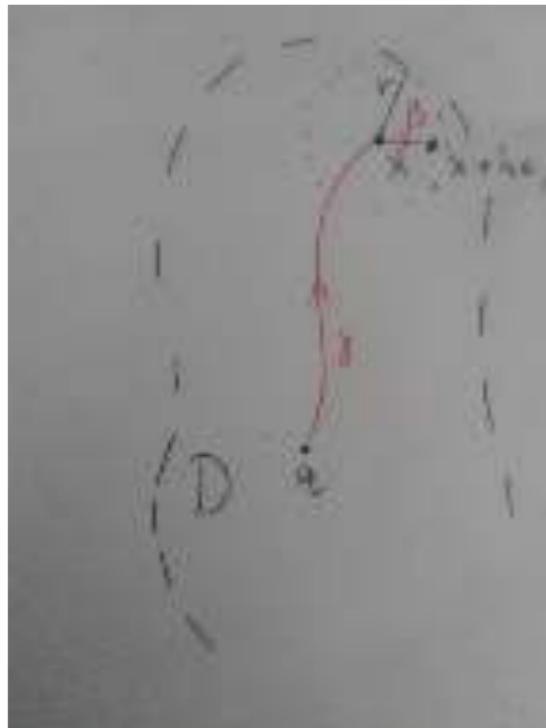


Figure: The path $\gamma\beta$ used in the proof

Example

We determine an antiderivative f of $\omega = (x^2 + y) dx + (x - y) dy$.

We must have $f_x = x^2 + y$, which can be integrated with respect to x (while y is kept fixed):

$$f(x, y) = \frac{x^3}{3} + xy + C(y)$$

for some function $y \mapsto C(y)$, which does not depend on x .

$$\implies f_y = x + C'(y),$$

which should be equal to $x - y$.

This implies $C'(y) = -y$ and hence $C(y) = -\frac{y^2}{2} + C$ with a constant $C \in \mathbb{R}$.

$$\implies f(x, y) = \frac{x^3}{3} + xy - \frac{y^2}{2}, \quad (x, y) \in \mathbb{R}^2$$

is the desired antiderivative. (If you have doubts, compute the partial derivatives of f and verify the claim directly!)

Any further antiderivative of ω on \mathbb{R}^2 (or a connected open subset thereof) has the form $f_C(x, y) = \frac{x^3}{3} + xy - \frac{y^2}{2} + C$ for some $C \in \mathbb{R}$.

Example (conservation of energy)

Suppose an object is moved from A to B in \mathbb{R}^3 along a path $\mathbf{r}(t) = \gamma(t)$, $t \in [a, b]$, under the sole influence of a continuous force field F .

$$\implies F(\gamma(t)) = m\gamma''(t) \quad (\text{Newton's 2nd Law})$$

$$\begin{aligned} \implies W &= \int_a^b F(\gamma(t)) \cdot \gamma'(t) dt = m \int_a^b \gamma''(t) \cdot \gamma'(t) dt, \\ &= (m/2) \int_a^b \frac{d}{dt} |\gamma'(t)|^2 dt = \frac{m|\gamma'(b)|^2}{2} - \frac{m|\gamma'(a)|^2}{2} \\ &= \frac{mv(b)^2}{2} - \frac{mv(a)^2}{2} = K(B) - K(A), \end{aligned}$$

where $K(X)$ denotes the kinetic energy of the object when it is at the point X . If $F = \nabla f$ is conservative and $P(X) = -f(X)$ is the corresponding potential function, we have $\nabla P = -F$ and $W = f(\gamma(b)) - f(\gamma(a)) = f(B) - f(A) = P(A) - P(B)$.

$$\implies P(A) + K(A) = P(B) + K(B)$$

This expresses the so-called *Law of Conservation of Energy*.

Locally Exact 1-Forms

Definition

A differential 1-form $\omega: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, is *locally exact* if every point $\mathbf{x} \in D$ has a neighborhood, on which ω is exact.

Of course we can assume that this neighborhood is a ball $B_r(\mathbf{x})$ for some $r > 0$. If ω is exact on D then it is locally exact with D serving as the desired neighborhood for all $\mathbf{x} \in D$.

Proposition

If $\omega = \sum_{i=1}^n f_i dx_i$ is locally exact (in particular, if ω is exact) and continuously differentiable (on D) then

$$(f_i)_{x_j} = (f_j)_{x_i} \quad \text{for all } 1 \leq i < j \leq n \quad (\text{on } D).$$

Proof.

Writing $\omega = df$, we have $f_i = f_{x_i}$ for $1 \leq i \leq n$ and hence $(f_i)_{x_j} = f_{x_i x_j} = f_{x_j x_i} = (f_j)_{x_i}$ by Clairaut's Theorem. Clairaut's Theorem can be applied, since the 2nd-order partial derivatives $f_{x_i x_j} = (f_i)_{x_j}$ are assumed to be continuous. □

Question

Does the converse of the proposition hold, i.e., are the conditions $(f_i)_{x_j} = (f_j)_{x_i}$ for $1 \leq i < j \leq n$ sufficient for local exactness?

Poincaré's Lemma (after HENRI POINCARÉ (1854–1912)), to be discussed later, gives an affirmative answer to this question.

Notes

- For $n = 2$ there is only one condition in the proposition, viz.
 $\omega = f_1 dx + f_2 dy$ should satisfy

$$(f_1)_y = (f_2)_x, \quad \text{or} \quad \partial_2 f_1 = \partial_1 f_2.$$

- For $n = 3$ there are three conditions, viz.
 $\omega = f_1 dx + f_2 dy + f_3 dz$ should satisfy

$$\partial_2 f_1 = \partial_1 f_2, \quad \partial_3 f_1 = \partial_1 f_3, \quad \partial_3 f_2 = \partial_2 f_3.$$

Formally, this can also be written as

$$\begin{pmatrix} \partial_1 \\ \partial_2 \\ \partial_3 \end{pmatrix} \times \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = 0.$$

Curl and Divergence

Definition

The *curl* and *divergence* of a 3-dimensional vector field $\mathbf{F} = (f_1, f_2, f_3)$ are defined as

$$\operatorname{curl} \mathbf{F} = \begin{pmatrix} \partial_2 f_3 - \partial_3 f_2 \\ \partial_3 f_1 - \partial_1 f_3 \\ \partial_1 f_2 - \partial_2 f_1 \end{pmatrix},$$

$$\operatorname{div} \mathbf{F} = \partial_1 f_1 + \partial_2 f_2 + \partial_3 f_3.$$

Writing the *Nabla* (or *gradient*) differential operator $f \mapsto \nabla f$ in the form $\nabla = (\partial_1, \partial_2, \partial_3)$, the definitions of curl and divergence can be formally written as

$$\operatorname{curl} = \nabla \times \mathbf{F}, \quad \operatorname{div} = \nabla \cdot \mathbf{F}.$$

In this form they are easy to remember.

Notes

- If F has domain D (which must be a subset of \mathbb{R}^3) and is partially differentiable then $\operatorname{curl} F$ and $\operatorname{div} F$ have domain D as well; $\operatorname{curl} F$ is another 3-dimensional vector field, while $\operatorname{div} F$ is a real valued function.
- Curl and divergence have physical interpretations in terms of fluid flow. If $F(x, y, z)$ represents the vectorial velocity of a fluid (or gas) at the point (x, y, z) then $\operatorname{curl} F(x, y, z) \cdot \mathbf{u}$, $|\mathbf{u}| = 1$, represents the circulation ("rotation") per unit area of the fluid at (x, y, z) around the axis $\mathbb{R}\mathbf{u}$; F is *irrotational* (i.e., whirl-free) if $\operatorname{curl} F = 0$. Similarly, $\operatorname{div} F(x, y, z)$ represents the net flow per unit volume from (x, y, z) ; F is *incompressible* (i.e., has no sources or sinks) if $\operatorname{div} F = 0$.
- Within the wider scope of differential k -forms, curl and divergence are essentially just special cases of the derivative of a differential k -form. For

$$\omega_0 = f,$$

$$\omega_1 = g_1 dx_1 + g_2 dx_2 + g_3 dx_3,$$

$$\omega_2 = h_1 dx_2 \wedge dx_3 + h_2 dx_1 \wedge dx_3 + h_3 dx_1 \wedge dx_2,$$

$$\omega_3 = k dx_1 \wedge dx_2 \wedge dx_3$$

Notes cont'd

- (cont'd) we have

$$d\omega_0 = df = (\partial_1 f) dx_1 + (\partial_2 f) dx_2 + (\partial_3 f) dx_3,$$

$$d\omega_1 = dg_1 \wedge dx_1 + dg_2 \wedge dx_2 + dg_3 \wedge dx_3$$

$$= (\partial_1 g_2 - \partial_2 g_1) dx_1 \wedge dx_2 + (\partial_1 g_3 - \partial_3 g_1) dx_1 \wedge dx_3 + (\partial_2 g_3 - \partial_3 g_2) dx_2 \wedge dx_3$$

$$d\omega_2 = dh_1 \wedge dx_2 \wedge dx_3 + dh_2 \wedge dx_1 \wedge dx_3 + dh_3 \wedge dx_1 \wedge dx_2$$

$$= (\partial_1 h_1 - \partial_2 h_2 + \partial_3 h_3) dx_1 \wedge dx_2 \wedge dx_3,$$

$$d\omega_3 = 0,$$

using the identities $dx_i \wedge dx_i = 0$ and $dx_j \wedge dx_i = -dx_i \wedge dx_j$ for $i \neq j$.

The derivative for differential k -forms satisfies the fundamental identity

$$d(d\omega) = 0$$

for all $0 \leq k \leq n-2$ and all differential k -forms whose component functions are C^2 -functions. In the two special cases $n=3$, $k=0$ and $k=1$ these are equivalent to

$$\operatorname{curl}(\nabla f) = 0 \quad \text{and} \quad \operatorname{div}(\operatorname{curl} G) = 0.$$

We now give direct proofs of these two identities and formulate them in the language of vector fields.

Theorem

- ① For any C^2 -function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^3$, we have $\operatorname{curl}(\nabla f) = 0$.
- ② For any C^2 -vector-field $G = (g_1, g_2, g_3): D \rightarrow \mathbb{R}^3$, $D \subseteq \mathbb{R}^3$, we have $\operatorname{div}(\operatorname{curl} G) = 0$.

Proof.

$$\operatorname{curl}(\nabla f) = \begin{pmatrix} \partial_1 \\ \partial_2 \\ \partial_3 \end{pmatrix} \times \begin{pmatrix} \partial_1 f \\ \partial_2 f \\ \partial_3 f \end{pmatrix} = \begin{pmatrix} \partial_2(\partial_3 f) - \partial_3(\partial_2 f) \\ \partial_3(\partial_1 f) - \partial_1(\partial_3 f) \\ \partial_1(\partial_2 f) - \partial_2(\partial_1 f) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\begin{aligned} \operatorname{div}(\operatorname{curl} G) &= \begin{pmatrix} \partial_1 \\ \partial_2 \\ \partial_3 \end{pmatrix} \cdot \begin{pmatrix} \partial_2 g_3 - \partial_3 g_2 \\ \partial_3 g_1 - \partial_1 g_3 \\ \partial_1 g_2 - \partial_2 g_1 \end{pmatrix} \\ &= \partial_1(\partial_2 g_3 - \partial_3 g_2) + \partial_2(\partial_3 g_1 - \partial_1 g_3) + \partial_3(\partial_1 g_2 - \partial_2 g_1) \\ &= (\partial_2 \partial_3 - \partial_3 \partial_2)g_1 + (\partial_3 \partial_1 - \partial_1 \partial_3)g_2 + (\partial_1 \partial_2 - \partial_2 \partial_1)g_3 \\ &= 0 \end{aligned}$$

Poincaré's Lemma

Definition

A subset D of \mathbb{R}^n is said to be *star-shaped* if there exists a “central” point $\mathbf{c} \in D$ such that for any other point $\mathbf{x} \in D$ the line segment $[\mathbf{c}, \mathbf{x}] = \{(1 - t)\mathbf{c} + t\mathbf{x}; 0 \leq t \leq 1\}$ is contained in D .

Examples

- 1 Convex sets are star-shaped. In a convex set every point can be taken as center in the definition of “star-shaped”.
- 2 The “slotted plane” $\mathbb{R}^2 \setminus \{(x, 0); x \leq 0\}$ is star-shaped (but not convex). Any point $(x, 0)$ with $x > 0$ can serve as center.
- 3 The punctured plane $\mathbb{R}^2 \setminus \{(0, 0)\}$ is not star-shaped.

Theorem (Poincaré's Lemma)

Let $\omega = \sum_{i=1}^n f_i dx_i$ be a continuously differentiable differential 1-form on $D \subseteq \mathbb{R}^n$. If ω satisfies $(f_i)_{x_j} = (f_j)_{x_i}$ for $1 \leq i < j \leq n$ and D is star-shaped then ω is exact.

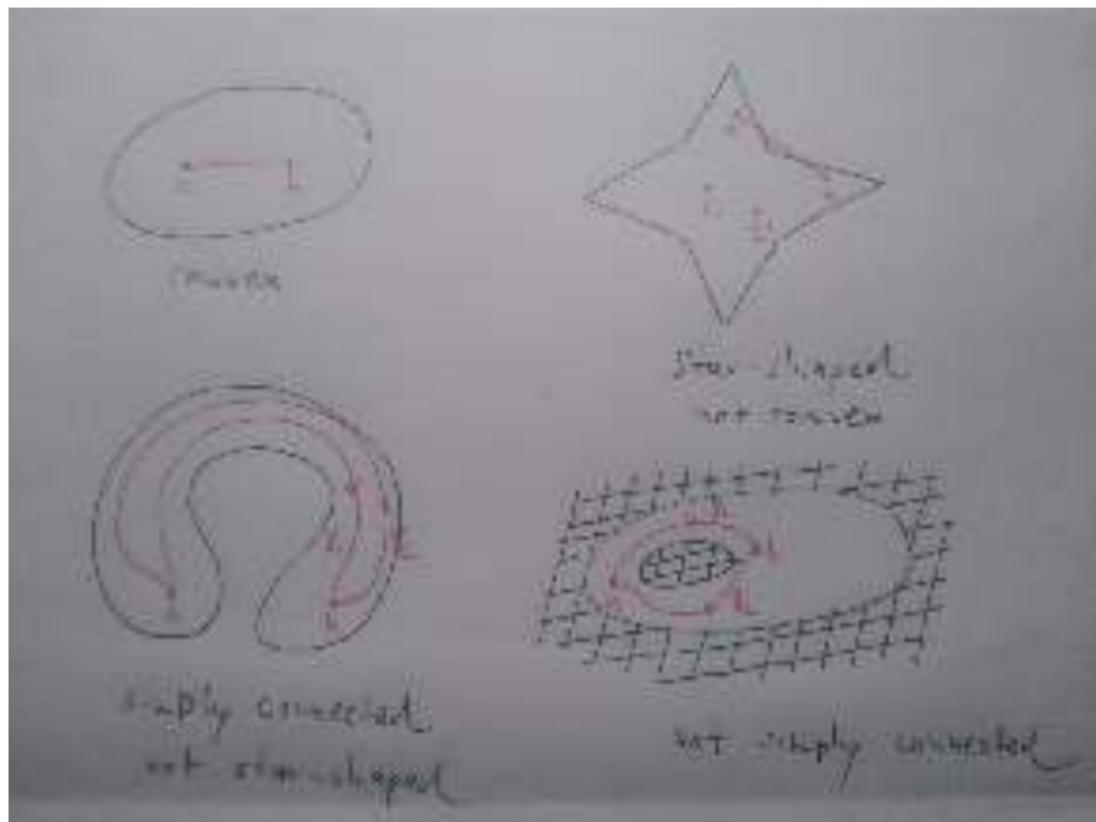
Vector Fields
and
Differential
Forms

Line Integrals

The
Fundamental
Theorem for
Line Integrals

Locally Exact
1-Forms

Line Integrals
in Complex
Analysis
(optional)



Note

In the calculus of differential forms one shows

$$d\omega = \sum_{i=1} df_i \wedge dx_i = \sum_{1 \leq i < j \leq n} (\partial_i f_j - \partial_j f_i) dx_i \wedge dx_j.$$

Hence the conditions in Poincaré's Lemma can be succinctly stated as " $d\omega = 0$ ".

Corollary

A continuously differentiable differential 1-form ω on an (open) set $D \subseteq \mathbb{R}^n$ is locally exact if and only if it satisfies the conditions in Poincaré's Lemma.

Proof.

" \implies " was shown earlier. For the proof of " \impliedby " let $\mathbf{x} \in D$. There exists a star-shaped neighborhood $D_1 \subseteq D$ of \mathbf{x} , for example $D_1 = B_r(\mathbf{x})$ with $r > 0$ sufficiently small. Poincaré's Lemma then gives that ω is exact on D_1 .



Proof of Poincaré's Lemma.

W.l.o.g. assume that D has center $\mathbf{c} = \mathbf{0}$. The Fundamental Theorem for Line Integrals gives that a suitable candidate for an antiderivative of ω is

$$f(\mathbf{x}) = \int_{\gamma_{\mathbf{x}}} \omega = \int_0^1 \sum_{i=1}^n f_i(t\mathbf{x}) x_i dt,$$

where $\gamma_{\mathbf{x}}(t) = t\mathbf{x}$ for $t \in [0, 1]$ (the parametrized line segment $[\mathbf{0}, \mathbf{x}]$). It remains to show that $f_{x_j} = f_j$ for $1 \leq j \leq n$.

Differentiating f under the integral sign (justification will follow!), we obtain

$$\begin{aligned} \frac{\partial f}{\partial x_j}(\mathbf{x}) &= \int_0^1 \frac{\partial f_j}{\partial x_j}(t\mathbf{x}) t x_j + f_j(t\mathbf{x}) + \sum_{\substack{i=1 \\ i \neq j}}^n \frac{\partial f_i}{\partial x_j}(t\mathbf{x}) t x_i dt \\ &= \int_0^1 f_j(t\mathbf{x}) + t \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_j}(t\mathbf{x}) x_i \right) dt \\ &= \int_0^1 f_j(t\mathbf{x}) + t \left(\sum_{i=1}^n \frac{\partial f_j}{\partial x_i}(t\mathbf{x}) x_i \right) dt = \int_0^1 \frac{d}{dt} (t f_j(t\mathbf{x})) dt = f_j(\mathbf{x}). \end{aligned}$$

Proof cont'd.

We have yet to justify the differentiation under the integral sign.

An inspection of the first line of the computation shows that the integrand

$$(\mathbf{x}, t) \mapsto \frac{\partial f_j}{\partial x_j}(t\mathbf{x})tx_j + f_j(t\mathbf{x}) + \sum_{\substack{i=1 \\ i \neq j}}^n \frac{\partial f_i}{\partial x_j}(t\mathbf{x})tx_i$$

is continuous on $D \times [0, 1]$ (as a function of $n + 1$ variables), and hence that differentiation under the integral sign was justified.

(The integrable bound $\Phi(t)$ required in the differentiation theorem for parameter integrals can be taken as the maximum absolute value of the integrand on a compact set of the form $\overline{B_r(\mathbf{x})} \times [0, 1]$ (and thus as a suitable constant). □

Example

The winding form $\omega = \frac{x \, dy - y \, dx}{x^2 + y^2}$ on $\mathbb{R}^2 \setminus \{(0, 0)\}$ satisfies the condition $(f_1)_y = (f_2)_x$ in Poincaré's Lemma:

$$(f_1)_y = \frac{\partial}{\partial y} \left(\frac{-y}{x^2 + y^2} \right) = \frac{-(x^2 + y^2) - (-y)(2y)}{(x^2 + y^2)^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2},$$

$$(f_2)_x = \frac{\partial}{\partial x} \left(\frac{x}{x^2 + y^2} \right) = \frac{x^2 + y^2 - x(2x)}{(x^2 + y^2)^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2}.$$

$\implies \omega$ has an antiderivative $f: D \rightarrow \mathbb{R}$ on the star-shaped region $D = \mathbb{R}^2 \setminus \{(x, 0); x \leq 0\}$.

Since $\int_{\gamma} \omega$ is independent of path, we can choose the most convenient path of integration to determine $f(x, y)$. Starting at $(1, 0)$ (thereby fixing the additive constant) and integrating along the x -axis to $[r, 0]$ with $r = \sqrt{x^2 + y^2}$, followed by the circular arc from $(r, 0)$ to (x, y) . Writing $x = r \cos \theta$, $y = r \sin \theta$ and using the curves $\gamma_1(t) = \begin{pmatrix} t \\ 0 \end{pmatrix}$ for $t \in [1, r]$, $\gamma_2(t) = \begin{pmatrix} r \cos t \\ r \sin t \end{pmatrix}$ for $t \in [0, \theta]$, we obtain

Example (cont'd)

$$\begin{aligned}
 f(x, y) &= \int_{\gamma_1} \frac{x \, dy - y \, dx}{x^2 + y^2} + \int_{\gamma_2} \frac{x \, dy - y \, dx}{x^2 + y^2} \\
 &= \int_1^r \frac{t \cdot 0 - 0 \cdot 1}{t^2 + 0} \, dt + \int_0^\theta \frac{(r \cos t)^2 - r \sin t(-r \sin t)}{(r \cos t)^2 + (r \sin t)^2} \, dt \\
 &= 0 + \int_0^\theta 1 \, dt = \theta.
 \end{aligned}$$

This also holds (mutatis mutandis) in the cases $0 < r < 1$, $-\pi < \theta < 0$.

Note that for points $(x_0, 0)$ with $x_0 < 0$ we have

$$\lim_{\substack{(x,y) \rightarrow (x_0,0) \\ y>0}} f(x, y) = \pi, \quad \lim_{\substack{(x,y) \rightarrow (x_0,0) \\ y<0}} f(x, y) = -\pi,$$

and hence $f(x, y)$ cannot be extended to the punctured plane $\mathbb{R}^2 \setminus \{(0, 0)\}$.

In fact, ω does not have an anti-derivative on $\mathbb{R}^2 \setminus \{(0, 0)\}$, since $\int_\gamma \omega$ is not independent of path there (cf. winding number).

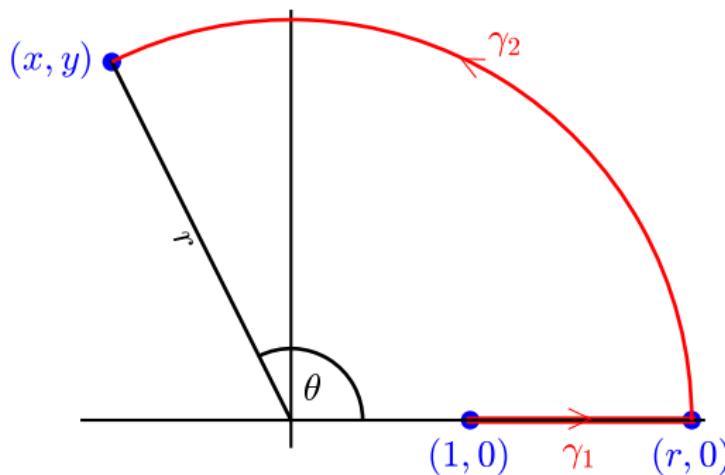


Figure: The path $\gamma_1\gamma_2$ used to compute an anti-derivative of the winding form on $\mathbb{R}^2 \setminus \{(x, 0); x \leq 0\}$

Afternote

Several students asked questions about this example and about exact vs locally exact differential 1-forms/vector fields. Hopefully the following additional remarks clarify the situation.

The region $E = \mathbb{R}^2 \setminus \{(x, 0); x \geq 0\}$ (which is just the mirror image of the region D in the previous example) is also star-shaped, and hence the winding form ω is also exact on E . Applying the same method as in the example but with starting point $(-1, 0) \in E$ (since $(1, 0)$ is not in E , it cannot be used) gives $\omega = dg$, where $g(x, y)$ measures the angle of (x, y) from the negative x -axis. We have $D \cup E = \mathbb{R}^2 \setminus \{(0, 0)\}$ (the domain of ω) but, since ω is not exact on $\mathbb{R}^2 \setminus \{(0, 0)\}$, the anti-derivatives of ω on D and E cannot be “glued” together.

This may seem paradoxical at the first glance: For example, if a differential 1-form is exact on each of two overlapping disks D_1 , D_2 then on $D_1 \cap D_2$ the anti-derivatives must differ by a constant, which can be adjusted on one disk to yield an anti-derivative on $D_1 \cup D_2$. However, the situation in the example is different: $D \cap E = H^+ \cup H^-$ is disconnected, and we have $g = f + \pi$ on H^+ , $g = f - \pi$ on H^- . Hence there is no way to adjust one of f, g to yield an anti-derivative on the whole punctured plane $\mathbb{R}^2 \setminus \{(0, 0)\}$.

Simply Connected regions

Roughly speaking, a connected open set $D \subseteq \mathbb{R}^n$ is “simply connected” if every closed path in D can be continuously contracted to a point entirely within D . Subsets of \mathbb{R}^2 with this property must not contain holes.

It will be shown that locally exact continuous differential 1-forms on a simply connected set $D \subseteq \mathbb{R}^n$ are exact. This vastly generalizes Poincaré’s Lemma.

Definition

Let $D \subseteq \mathbb{R}^n$ be open and connected.

- 1 Two paths $\gamma_0, \gamma_1: [0, 1] \rightarrow D$ with the same initial and terminal point \mathbf{a} , resp., \mathbf{b} are said to be *homotopic* in D if there exists a continuous map $H: [0, 1] \times [0, 1] \rightarrow D$ such that $\gamma_0(t) = H(t, 0)$, $\gamma_1(t) = H(t, 1)$ for all $t \in [0, 1]$ and $H(0, s) = \mathbf{a}$, $H(1, s) = \mathbf{b}$ for all $s \in [0, 1]$.

In other words, the family of paths $\gamma_s: [0, 1] \rightarrow D$, $t \mapsto H(t, s)$, $0 \leq s \leq 1$, starts with γ_0 , ends with γ_1 , deforms γ_0 continuously into γ_1 , and all intermediate paths have the same initial and terminal points as γ_0, γ_1 . (The crucial condition is that all paths γ_s are entirely contained in D .)

Definition (cont'd)

- ② Two closed paths are said to be *freely homotopic* in D if there exists a continuous map $H: [0, 1] \times [0, 1] \rightarrow D$ satisfying $\gamma_0(t) = H(t, 0)$, $\gamma_1(t) = H(t, 1)$ for all $t \in [0, 1]$ and $H(0, s) = H(1, s)$ for all $s \in [0, 1]$. In other words, the family of paths $\gamma_s: [0, 1] \rightarrow D$, $t \mapsto H(t, s)$, $0 \leq s \leq 1$, transforms γ_0 continuously into γ_1 as in (1); all intermediate paths must be closed as well but may have different start/end points.
- ③ D is said to be *simply connected* if every closed path in D is (freely) homotopic to a path of the form $[0, 1] \rightarrow D$, $t \mapsto \mathbf{a} \in D$ (*point path*)

Notes

- In the definition we have assumed that all parameter intervals are equal to $[0, 1]$. For curves with arbitrary parameter intervals one can define homotopy either in terms of reparametrizations to $[0, 1]$ or use maps H with a more general (non-rectangular) domain. The results are the same.
- Any two paths with the same initial and terminal points in a simply connected region are homotopic, and this property being equivalent to (3) also characterizes simply connected regions.

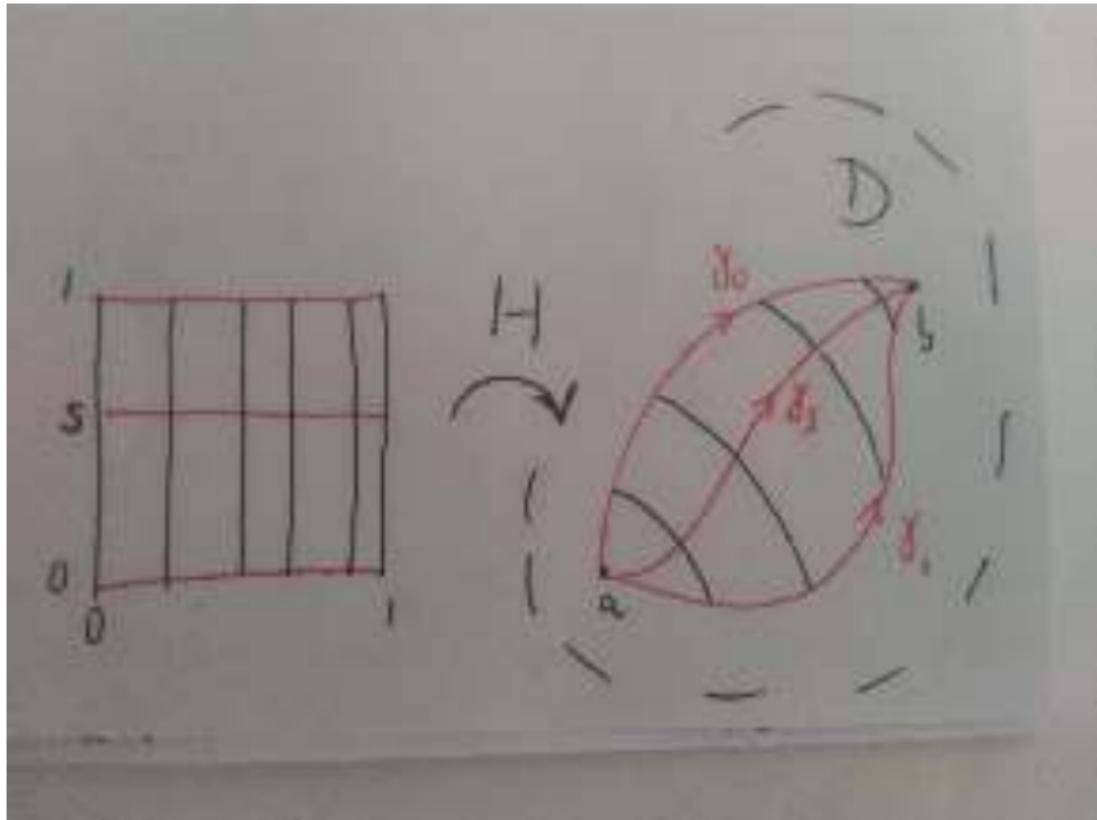


Figure: A homotopy between γ_0 and γ_1

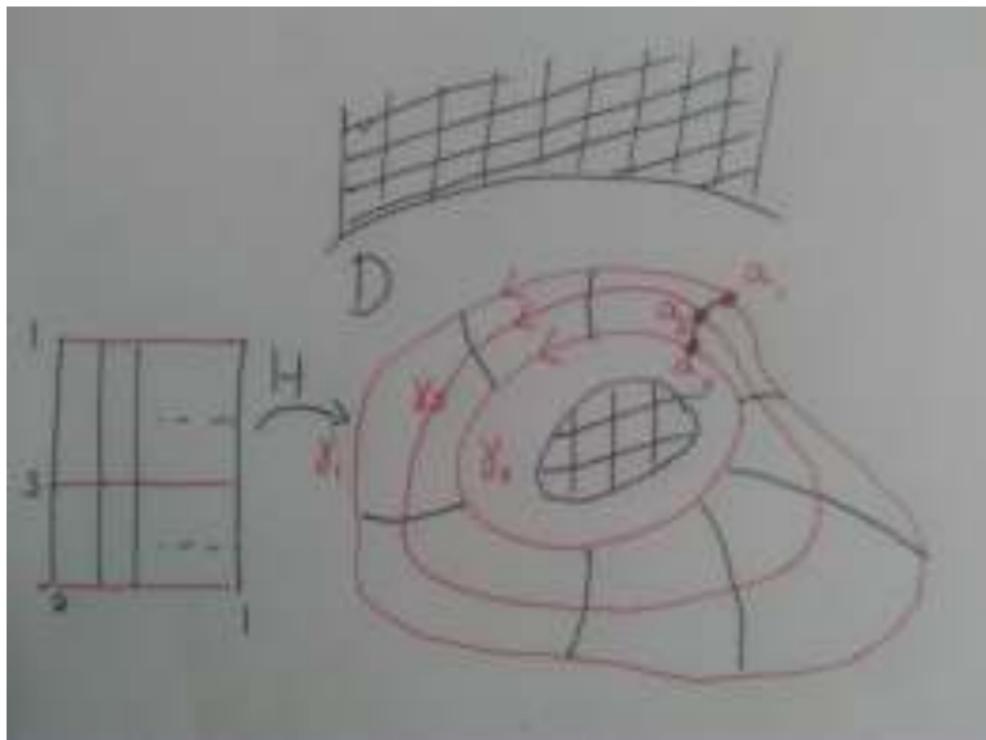


Figure: A free homotopy between closed curves γ_0 and γ_1

All intermediate curves γ_s must be closed as well, i.e.,
 $\gamma_s(0) = \gamma_s(1) = \mathbf{a}_s$ for $s \in [0, 1]$.

Theorem

Suppose ω is a locally exact continuous differential 1-form on $D \subseteq \mathbb{R}^n$. (For continuously differentiable 1-forms $\omega = \sum_{i=1}^n f_i dx_i$, this is equivalent to the conditions $(f_i)_{x_j} = (f_j)_{x_i}$, $1 \leq i < j \leq n$, in Poincaré's Lemma.) If the paths γ_0 and γ_1 are homotopic or freely homotopic in D then

$$\int_{\gamma_0} \omega = \int_{\gamma_1} \omega.$$

Corollary

A locally exact continuous differential 1-form on a simply connected region $D \subseteq \mathbb{R}^n$ is exact.

Proof of the theorem.

We consider only the case of homotopic paths. Suppose $\gamma_0(0) = \gamma_1(0) = \mathbf{a}$, $\gamma_0(1) = \gamma_1(1) = \mathbf{b}$, and $H: [0, 1] \times [0, 1] \rightarrow D$ is a homotopy between γ_0 and γ_1 .

Claim: There exist a positive integer m such that the obvious subdivision of $[0, 1]^2$ into m^2 smaller squares Q_{ij} ($0 \leq i, j \leq m - 1$) of side length $1/m$ has the property that ω is exact on each set $H(Q_{ij})$.

Proof cont'd.

Proof: Otherwise there exists a sequence $(Q_m)_{m \geq 1}$ of squares $Q_m \subseteq [0, 1]^2$ with side lengths $1/m \rightarrow 0$ on which ω is not exact. By the Bolzano-Weierstrass Theorem, the bottom-left corners of Q_m , say, must have a convergent subsequence with limit $(t_0, s_0) \in [0, 1]^2$. Since ω is locally exact, there exists a ball around $H(t_0, s_0) \in D$ on which ω is exact, and continuity of H then in turn implies the existence of a disk B around (t_0, s_0) such that ω is exact on the image $H(B)$. But B contains a square Q_m (in fact almost all such squares), and hence ω must be exact on $H(Q_m) \subseteq H(B)$. Contradiction!

With the claim at hand, the proof is now easy to finish.

The sides of the m^2 squares Q_{ij} define $\binom{2m}{m}$ paths in $[0, 1]^2$ from $(0, 0)$ to $(1, 1)$ and hence their images under H the same number of paths in D from $H(0, 0) = \mathbf{a}$ to $H(1, 1) = \mathbf{b}$: At each lower-left corner $(i/m, j/m)$ either follow the “horizontal” path $t \mapsto H(t, j/m)$, $i/m \leq t \leq (i+1)m$, or the “vertical” path $s \mapsto H(i/m, s)$, $j/m \leq s \leq (j+1)/m$.

Such a path is uniquely described by a word of length $2m$ containing m letters ‘h’ (for “horizontal”) and m letters ‘v’ (for “vertical”).

Proof cont'd.

Since $H(0, s) = \mathbf{a}$, $H(1, s) = \mathbf{b}$, we have

$$\int_{\gamma_0} \omega = \int_{\sigma} \omega \quad \text{for } \sigma \triangleq hh\dots hvv\dots v,$$

$$\int_{\gamma_1} \omega = \int_{\tau} \omega \quad \text{for } \tau \triangleq vv\dots vhvv\dots h,$$

Since ω is exact and continuous on each set $H(Q_{ij})$, we can exchange in such a path any subword 'hv' for 'vh' without affecting the integral, because the endpoints of the corresponding subpaths are the same (hence the Fundamental Theorem for Line Integrals applies).

Using such exchanges repeatedly, we can transform σ into τ (this is obvious!).

$$\Rightarrow \int_{\gamma_0} \omega = \int_{\sigma} \omega = \int_{\tau} \omega = \int_{\gamma_1} \omega,$$

and the proof is complete. □

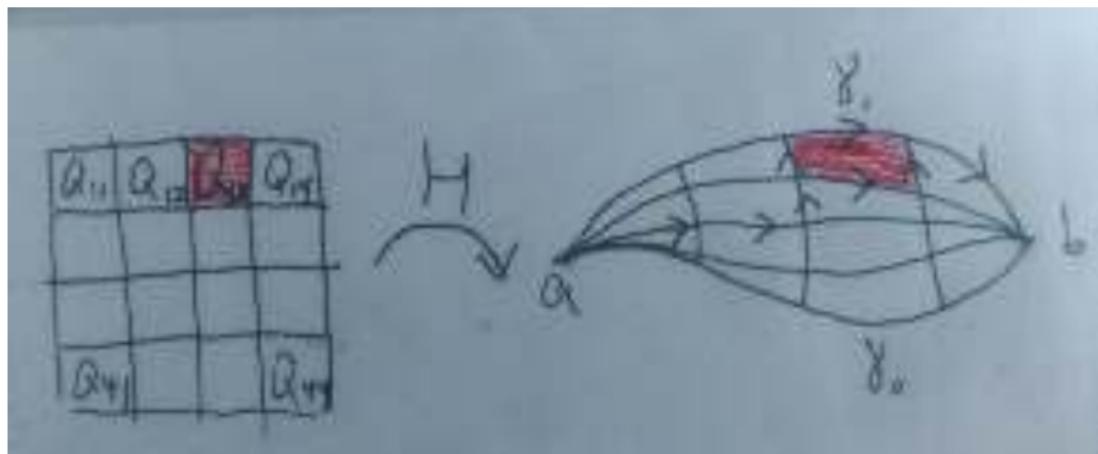


Figure: Two paths from **a** to **b** which differ by a switch $hv \rightarrow vh$

Example

We compute the line integral of the winding form $\omega = \frac{x \, dy - y \, dx}{x^2 + y^2}$

along the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ parametrized by
 $\epsilon(t) = (a \cos t, b \sin t)$, $t \in [0, 2\pi]$.

It is obvious that ϵ is homotopic in $\mathbb{R}^2 \setminus \{(0, 0)\}$ to the circle
 $\kappa(t) = (a \cos t, a \sin t)$, $t \in [0, 2\pi]$, an explicit homotopy being provided by

$$H(t, s) = (a \cos t, ((1-s)a + sb) \sin t), \quad (t, s) \in [0, 2\pi] \times [0, 1].$$

For the circle we have $\int_{\kappa} \omega = 2\pi n(\kappa; \mathbf{0}) = 2\pi$.

$$\Rightarrow \int_{\epsilon} \frac{x \, dy - y \, dx}{x^2 + y^2} = \int_0^{2\pi} \frac{ab}{a^2 \cos^2 t + b^2 \sin^2 t} dt = 2\pi$$

or, alternatively,

$$\int_0^{2\pi} \frac{dt}{a^2 \cos^2 t + b^2 \sin^2 t} = \frac{2\pi}{ab}.$$

This integral can also be evaluated in other ways, but the present evaluation is particularly short and elegant.

Remark

In case you are wondering if there are examples of non-exact but locally exact differential 1-forms on $\mathbb{R}^2 \setminus \{(0, 0)\}$ other than the winding form ω_W , the answer is “No” in the following sense: For each differential 1-form $\omega = f dx + g dy$ on $\mathbb{R}^2 \setminus \{(0, 0)\}$ satisfying $f_y = g_x$ there exists a constant $c \in \mathbb{R}$ such that $\epsilon = \omega - c\omega_W$ is exact. The constant is determined by

$$c = \frac{1}{2\pi} \int_{\partial B} \omega,$$

where ∂B denotes the unit circle in its positive (i.e., counterclock-wise) orientation.

The proof uses that $\int_{\partial B} \epsilon = 0$ (obvious from $\int_{\partial B} \omega_W = 2\pi$) and that every closed path in $\mathbb{R}^2 \setminus \{(0, 0)\}$ is freely homotopic to the m -times traversed unit circle in either positive or negative orientation for some integer $m \geq 0$.

There is an elaborate theory investigating the analogous problem—How far can a differential k -form ω on a given domain $D \subseteq \mathbb{R}^n$ with $d\omega = 0$ (such forms are said to be *closed*) be from being exact?—in general (\rightarrow *de Rham cohomology*).

Complex Line Integrals

Recall that integrals of complex-valued functions $f = \operatorname{Re} f + i \operatorname{Im} f$ are defined component-wise:

$$\int f = \int \operatorname{Re} f + i \int \operatorname{Im} f.$$

The complex view is sometimes convenient even if one is interested only in ordinary integrals of real-valued functions.

Example

Show that $\int_0^{2\pi} \cos(mt) \cos(nt) dt = 0$ for all $m, n \in \mathbb{N}$ with $m \neq n$.

Solution: Using $\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$ we obtain

$$\begin{aligned}\cos(mt) \cos(nt) &= \frac{1}{4} (e^{imt} + e^{-imt}) (e^{int} + e^{-int}) \\ &= \frac{1}{4} \left(e^{i(m+n)t} + e^{i(m-n)t} + e^{i(-m+n)t} + e^{i(-m-n)t} \right),\end{aligned}$$

$$\int_0^{2\pi} \cos(mt) \cos(nt) dt = \frac{1}{4} \left[\frac{e^{i(m+n)t}}{i(m+n)} + \frac{e^{i(m-n)t}}{i(m-n)} + \frac{e^{i(-m+n)t}}{i(-m+n)} + \frac{e^{i(-m-n)t}}{i(-m-n)} \right]_0^{2\pi} = 0,$$

since $t \mapsto e^{ikt}$, $k \in \mathbb{Z}$, is 2π -periodic.

Definition

The line integral of a complex-valued differential 1-form $\omega = \omega_1 + i\omega_2$, $\omega_1, \omega_2: D \rightarrow (\mathbb{R}^n)^*$, $D \subseteq \mathbb{R}^n$, along a curve $\gamma: [a, b] \rightarrow D$ is defined as

$$\int_{\gamma} \omega = \int_{\gamma} \omega_1 + i \int_{\gamma} \omega_2.$$

Of course, ω is said to be integrable along γ if both ω_1 and ω_2 are integrable along γ .

Observation

If ω is \mathbb{C} -linear, i.e., $\omega = f dz$ for some function $f: D \rightarrow \mathbb{C}$, ω is continuous (i.e., f is continuous) and γ is a path then

$$\int_{\gamma} \omega = \int_{\gamma} f(z) dz = \int_a^b f(\gamma(t)) \gamma'(t) dt.$$

Proof.

Writing $f = u + iv$, $\gamma = x + iy$, we have

$$\omega = (u + iv)(dx + idy) = u dx - v dy + i(v dx + u dy)$$

and hence

Proof cont'd.

$$\begin{aligned}\int_{\gamma} \omega &= \int_a^b u((\gamma(t))x'(t) - v(\gamma(t))y'(t)) dt + i \int_a^b v((\gamma(t))x'(t) + u(\gamma(t))y'(t)) dt \\&= \int_a^b u((\gamma(t))x'(t) - v(\gamma(t))y'(t) + i(v((\gamma(t))x'(t) + u(\gamma(t))y'(t))) dt \\&= \int_a^b [u(\gamma(t)) + i v(\gamma(t))] \cdot [x'(t) + i y'(t)] dt \\&= \int_a^b f(\gamma(t))\gamma'(t) dt.\end{aligned}$$



A Glimpse of Complex Analysis

Recall that $f: D \rightarrow \mathbb{C}$, $D \subseteq \mathbb{C}$ open, is said to be holomorphic if $f'(z) = \lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{C}}} \frac{f(z+h) - f(z)}{h}$ exists for all $z \in D$.

Theorem (CAUCHY's Integral Theorem)

Suppose $f: D \rightarrow \mathbb{C}$, $D \subseteq \mathbb{C}$ open, is holomorphic.

- ① $\omega = f \, dz$ is locally exact on D .
- ② $\int_{\gamma} f(z) \, dz = 0$ for every closed path in D that is (freely) homotopic to a point path.
- ③ $\int_{\gamma_1} f(z) \, dz = \int_{\gamma_2} f(z) \, dz$ for any two homotopic paths (or freely homotopic closed paths) in D .
- ④ If D is simply connected then there exists a holomorphic function $F: D \rightarrow \mathbb{C}$ such that $F' = f$.

Note that in (1) $\omega = \omega_1 + i\omega_2$ is locally exact iff both ω_1 and ω_2 are. (This follows from $dF = d(U + iV) = dU + i \, dV$, i.e., $dF = \omega$ is equivalent to $dU = \omega_1 \wedge dV = \omega_2$.)

Proof.

(1) We have seen that in terms of $u = \operatorname{Re} f$, $v = \operatorname{Im} f$ the 1-form ω is expressed as

$$\omega = f dz = u dx - v dy + i(v dx + u dy).$$

Since f is holomorphic, we have $u_x = v_y$, $u_y = -v_x$, which are the conditions required in Poincaré's Lemma for $\omega_1 = u dx - v dy$ and $\omega_2 = v dx + u dy$.

However, Poincaré's Lemma also requires u_x, v_x (equivalently $f' = u_x + iv_x$) to be continuous, which we cannot prove at this point. Instead we simply remark that this gap can be closed and holomorphic functions are indeed C^1 -functions.

Subject to this claim, Poincaré's Lemma then finishes the proof of Part (1).

(2) and (3) reduce to the corresponding properties of $\int_{\gamma} \omega_1$ and $\int_{\gamma} \omega_2$ (resp., $\int_{\gamma_i} \omega_1$ and $\int_{\gamma_i} \omega_2$), which we have proved before.

(4) Here the preceding theory similarly gives that $\omega_1 = dU$ and $\omega_2 = dV$ are exact on D . For $F = U + iV$ we then have $dF = \omega$ (as noted before the proof) and

$$U_x = u, \quad U_y = -v, \quad V_x = v, \quad V_y = u.$$

Proof cont'd.

$\implies U_x = V_y \wedge U_y = -V_x \implies F$ is holomorphic.

Finally, comparing $dF = F' dz$ with $\omega = f dz$ yields $F' = f$, as claimed.



Example (Principal branch of the complex logarithm)

The function $f(z) = 1/z$ is holomorphic on $\mathbb{R}^2 \setminus \{(0, 0)\}$ with $f'(z) = -1/z^2$. (The proof of $\frac{d}{dx}(1/x) = -1/x^2$ from Calculus I generalizes to the complex case.)

$\Rightarrow f$ has a complex antiderivative F on the simply connected region $D = \mathbb{R}^2 \setminus \{(x, y); x \leq 0\}$, and up to an additive constant $F(z)$ can be found by integrating $1/z$ along a path from $1 = (1, 0)$ to $z = x + iy = (x, y)$.

We choose the same path as for integrating the winding form, but this time in the more convenient complex notation: $\gamma_1(t) = t$ for $t \in [1, r]$, where $r = \sqrt{x^2 + y^2} = |z|$, followed by $\gamma_2(t) = (r \cos t, r \sin t) = re^{it}$ for $t \in [0, \theta]$.

$$\begin{aligned}\Rightarrow F(z) &= \int_1^z \frac{dw}{w} = \int_1^r \frac{1}{t} dt + \int_0^\theta \frac{ire^{it}}{re^{it}} dt \\ &= \ln r + i\theta = \ln |z| + i \arg(z),\end{aligned}$$

where $\theta = \arg(z)$ is defined by $x = |z| \cos \theta$, $y = |z| \sin \theta$.

The function F satisfies $e^{F(z)} = z$ for all $z \in D$ and is referred to as *principal branch of the complex (natural) logarithm*.

Math 241
Calculus III

Thomas
Honold

Surface
Integrals

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Outline I

1 Surface Integrals

Today's Lecture:

Surface Integration on Curves

Also called integration with respect to arc length

Question

How to integrate a function f over a non-parametric curve such as the boundary $C = \partial D$ of the half disk
 $D = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 \leq 1, y \geq 0\}$?

Answer

We want the surface integral $\int_C f \, ds$ to satisfy the following properties:

- 1 For the constant function $f = 1$ the integral should coincide with the arc length.
- 2 The integral should be linear in f , i.e.,
$$\int_C (c_1 f_1 + c_2 f_2) \, ds = c_1 \int_C f_1 \, ds + c_2 \int_C f_2 \, ds \quad (c_1, c_2 \in \mathbb{R}).$$

This leads to the following definition of $\int_C f \, ds$.

- 1 Remove the singularities of f . (In our example these are the points $(\pm 1, 0)$.)
- 2 If C is “well-behaved”, this leaves mutually disjoint pieces C_i which can be parametrized in the form $\gamma_i: I_i \rightarrow C_i$ with $I_i = (a_i, b_i) \subseteq \mathbb{R}$ and γ_i smooth and one-to-one.
- 3 Define

$$\int_{C_i} f \, ds = \sum_{a_i}^{b_i} f(\gamma(t)) |\gamma'(t)| dt$$

and $\int_C f \, ds = \sum_i \int_{C_i} f \, ds$.

(3) is motivated by our two postulates and the arc length formula. More precisely, sampling f at the curve points $P_j^* = \gamma_i(t_j^*)$ corresponding to a tagged partition $a_i = t_0 < t_1 < \dots < t_N = b_i$ with tags $t_j^* \in [t_{j-1}, t_j]$ yields the two approximations

$$\begin{aligned} \int_{C_i} f \, ds &\approx \sum_{j=1}^N f(P_j^*) |P_j - P_{j-1}| = \sum_{j=1}^N f(\gamma(t_j^*)) |P_j - P_{j-1}| \\ &\approx \sum_{j=1}^N f(\gamma(t_j^*)) |\gamma'(t_{j-1})| (t_j - t_{j-1}), \end{aligned}$$

where we have used the abbreviations $\gamma = \gamma_i$ and $P_j = \gamma(t_j) = \gamma_i(t_j)$.

Question: Is the integral in (3), which is defined in terms of one particular parametrization of C_i , well-defined (i.e., the same for all parametrizations of C_i)?

Answer: Yes.

We verify this for our example curve and two particular parametrizations of the piece C_2 , which is the half circle.

(The other piece is the line segment $C_1 = [-1, 1] \subseteq \mathbb{C} \triangleq \mathbb{R}^2$, for which $\int_{C_1} f \, ds = \int_{-1}^1 f(x, 0) \, dx$.)

$$\gamma(t) = (\cos t, \sin t), \quad t \in (0, \pi),$$

$$\beta(x) = (x, \sqrt{1 - x^2}), \quad x \in (-1, 1).$$

Using $\gamma'(t) = (-\sin t, \cos t)$, $\beta'(x) = (1, -\frac{x}{\sqrt{1-x^2}})$, this gives the two values

$$\int_{C_2} f \, ds = \int_0^\pi f(\cos t, \sin t) \, dt = \int_{-1}^1 \frac{f(x, \sqrt{1 - x^2})}{\sqrt{1 - x^2}} \, dx$$

The change-of-variables $x = \cos t$, $dx = -\sin t \, dt$ shows that the two integrals have the same value.

Example

As a concrete example, consider $f(x, y) = y$.

Since $y = 0$ on C_1 , we have $\int_{C_1} y \, ds = 0$.

$$\int_{C_2} y \, ds = \int_0^\pi \sin t \, dt = \int_{-1}^1 \frac{\sqrt{1-x^2}}{\sqrt{1-x^2}} = 2,$$

$$\Rightarrow \int_C y \, ds = \int_{C_1} y \, ds + \int_{C_2} y \, ds = 0 + 2 = 2.$$

Parametric Surfaces

Definition

A C^1 -map $\gamma: \Omega \rightarrow \mathbb{R}^n$, $\Omega \subseteq \mathbb{R}^d$ open, is called an *immersion* (or a *regular map*) if the Jacobi matrix $\mathbf{J}_\gamma(\mathbf{u}) \in \mathbb{R}^{n \times d}$ has full column rank d for every $\mathbf{u} \in \Omega$. The range (image) $S = \gamma(\Omega) \subseteq \mathbb{R}^n$ of an immersion γ is called a d -dimensional *parametric surface*.

Notes

- A parametric surface need not be smooth at every point. As an example consider the parametric curve (1-dim. parametric surface)

$$\gamma(t) = (t^2 - 1, t^3 - t), \quad t \in \mathbb{R}.$$

$\gamma'(t) = (2t, 3t^2 - 1) \neq (0, 0)$ for all $t \in \mathbb{R} \implies \gamma$ is an immersion.
 $S = \gamma(\mathbb{R})$ has the double point $\gamma(\pm 1) = (0, 0)$ (the origin of \mathbb{R}^2), at which $\gamma'(\pm 1) = (\pm 2, 2)$. Thus S has no unique tangent at $(0, 0)$.

In this example the problem is related to the fact that γ is not injective, but there are even injective immersions with a non smooth range, e.g., $\gamma(t) = \sin(2t) \left(\begin{smallmatrix} \cos t \\ \sin t \end{smallmatrix} \right)$ for $t \in (-\pi/2, \pi/2)$.

Notes cont'd

- For $\mathbf{u} \in \Omega$ there exists a ball $B_r(\mathbf{u}) \subseteq \Omega$, and we can consider the d parametric curves $\epsilon_j: (-r, r) \rightarrow S$, $t \mapsto \gamma(\mathbf{u} + t\mathbf{e}_j)$, $1 \leq j \leq d$. The chain rule gives

$$\epsilon'_j(t) = \mathbf{J}_\gamma(\mathbf{u})\mathbf{e}_j = j\text{-th column of } \mathbf{J}_\gamma(\mathbf{u}),$$

so that the set of tangent vectors at S in $\mathbf{x} = \gamma(\mathbf{u})$ contains a d -dimensional subspace of \mathbb{R}^n (viz., the column space of $\mathbf{J}_\gamma(\mathbf{u})$).

- It can be shown that each parameter $\mathbf{u} \in \Omega$ has an open neighborhood $\Omega_1 \subseteq \Omega$ such that the restriction $S_1 = \gamma(\Omega_1)$ is a smooth d -dimensional surface. This means that the set of tangent vectors at S_1 in $\mathbf{x} = \gamma(\mathbf{u})$ is precisely the column space of $\mathbf{J}_\gamma(\mathbf{u})$ and that S_1 is the level set of a C^1 -function $f: \mathbb{R}^n \rightarrow \mathbb{R}^{n-d}$ with $\mathbf{J}_f(\mathbf{x}) \in \mathbb{R}^{(n-d) \times n}$ of full rank $n - d$.

Example (The unit sphere S^2)

An immersion parametrizing the 2-dimensional unit sphere $S^2 = S_1(\mathbf{0}) \subseteq \mathbb{R}^3$ can be obtained by restricting spherical coordinates to $r = 1$.

$$\gamma(\theta, \phi) = T(1, \theta, \phi) = \begin{pmatrix} \cos \theta \sin \phi \\ \sin \theta \sin \phi \\ \cos \phi \end{pmatrix}, \quad (\theta, \phi) \in (0, 2\pi) \times (0, \pi).$$

The range of γ is not the whole of S^2 but omits the 0-meridian $\{(x, 0, z); x^2 + z^2 = 1, x \geq 0\}$.

$$\mathbf{J}_\gamma(\theta, \phi) = \begin{pmatrix} -\sin \theta \sin \phi & \cos \theta \cos \phi \\ \cos \theta \sin \phi & \sin \theta \cos \phi \\ 0 & -\sin \phi \end{pmatrix}$$

It is clear that γ is regular (because $\phi = 0, \pi$ is not allowed and $\sin \theta, \cos \theta$ have no common zero).

Scaling γ by $r > 0$ and translating by $\mathbf{a} \in \mathbb{R}^3$, i.e., using the same domain and $(\theta, \phi) \mapsto r \gamma(\theta, \phi) + \mathbf{a}$, gives a parametrization of the general sphere $S_r(\mathbf{a}) \subset \mathbb{R}^3$.

Example (Surfaces of revolution)

A parametric curve $\alpha(t) = (r(t), z(t))$, $t \in I$, in the (r, z) -plane creates a parametric surface in the following way:

$$\gamma(\theta, t) = \begin{pmatrix} r(t) \cos \theta \\ r(t) \sin \theta \\ z(t) \end{pmatrix}, \quad (\theta, t) \in \mathbb{R} \times I.$$

The surface $S = \gamma(\mathbb{R} \times I)$ is obtained by rotating the curve $t \mapsto \gamma(0, t) = (r(t), 0, z(t))$ (a copy of α in the (x, z) -plane) around the z -axis.

$$\mathbf{J}_\gamma(\theta, t) = \begin{pmatrix} -r(t) \sin \theta & r'(t) \cos \theta \\ r(t) \cos \theta & r'(t) \sin \theta \\ 0 & z'(t) \end{pmatrix}$$

This matrix has rank 2 if $r(t)r'(t) \neq 0$ or $r(t)z'(t) \neq 0$.

$\implies \gamma$ is regular if α is regular (smooth) and $r(t) > 0$ for $t \in I$ (which is a reasonable assumption).

Example (cont'd)

Spheres are surfaces of revolution, e.g., the unit sphere S^2 is obtained by taking α as the half circle

$$\alpha(t) = (\cos t, \sin t), \quad t \in (-\pi/2, \pi/2).$$

Another interesting special case is that of a *torus* ("tire") obtained by rotating a circle contained in the right half of the (x, z) -plane, e.g.,

$$\alpha(t) = (A + a \cos t, a \sin t), \quad t \in \mathbb{R},$$

for some $0 < a < A$.

Like a sphere, such a torus surface T can also be defined algebraically as a level set of some real-valued function on \mathbb{R}^3 : The condition $(r - A)^2 + z^2 = a^2$, $r = \sqrt{x^2 + y^2}$, for points on the circle $\alpha(\mathbb{R})$ shows that T is the solution set of the equation

$$f(x, y, z) = (\sqrt{x^2 + y^2} - A)^2 + z^2 = a^2.$$

Since $\nabla f(x, y, z) \neq (0, 0, 0)$ for all $(x, y, z) \in T$ (as is easily verified), this shows that T is smooth as well.

Example (Smooth curves)

A C^1 -curve $\gamma: I \rightarrow \mathbb{R}^n$, $I \subseteq \mathbb{R}$ an interval, defines an immersion if $\gamma'(t) \neq \mathbf{0}$ for every $t \in I$, i.e., the parametric curve γ is smooth.

Example (Graphs of C^1 -functions)

Suppose $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is a C^1 -map. Then $\gamma: D \rightarrow \mathbb{R}^{m+n}$, $\mathbf{x} \mapsto (\mathbf{x}, f(\mathbf{x}))$ is a C^1 -map as well, and

$$\mathbf{J}_\gamma(\mathbf{x}) = \begin{pmatrix} \mathbf{I}_n \\ \mathbf{J}_f(\mathbf{x}) \end{pmatrix}$$

shows that $\mathbf{J}_\gamma(\mathbf{x})$ has rank n for $\mathbf{x} \in D$, and hence that γ is an immersion. (There is no condition here on the rank of $\mathbf{J}_f(\mathbf{x})$.)

The n -dimensional parametric surface in \mathbb{R}^{m+n} determined by γ is just the graph

$$G_f = \{(\mathbf{x}, f(\mathbf{x}); \mathbf{x} \in D\}$$

of f .

Differentiable Manifolds

The proper setting for surface integration are so-called differentiable manifolds. We only consider manifolds embedded into \mathbb{R}^n for some n .

Definition

A subset $M \subseteq \mathbb{R}^n$ is called a *d-dimensional differentiable manifold* if for every point $\mathbf{a} \in M$ there exist an open neighborhood U of \mathbf{a} and a diffeomorphism $\phi: U \rightarrow V$ onto some open subset $V \subseteq \mathbb{R}^n$ such that

$$\phi(M \cap U) = \{\mathbf{x} \in V; x_{d+1} = x_{d+2} = \cdots = x_n = 0\}.$$

The map ϕ is called a *chart* for M and $M \cap U$ the corresponding *chart region*. An *atlas* for M is a collection of charts whose domains U (or chart regions $M \cap U$) cover M . Intuitively speaking, a *d*-dimensional differentiable manifold looks locally like a “curved” version of a piece of \mathbb{R}^d embedded into \mathbb{R}^n in the standard way.

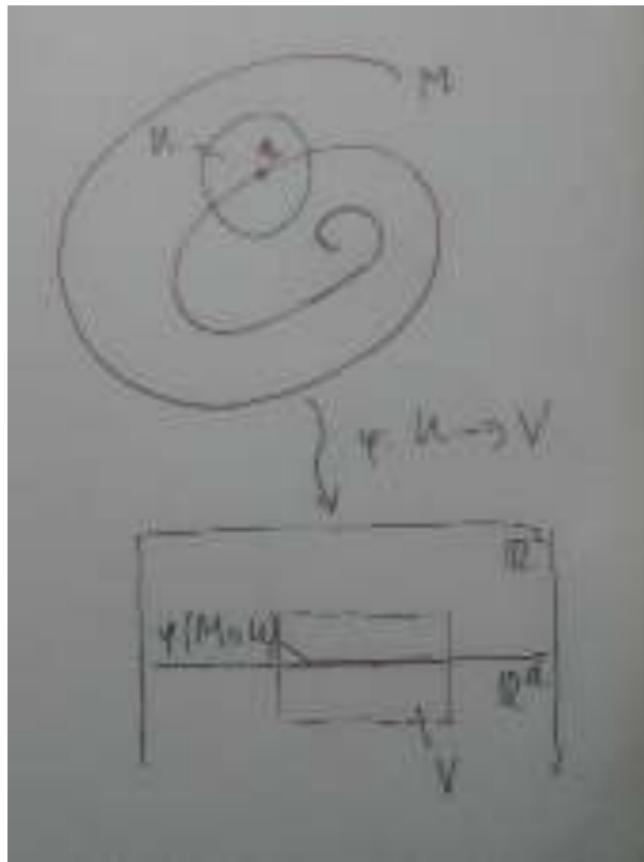


Figure: Illustration of charts for a d -dimensional differentiable manifold

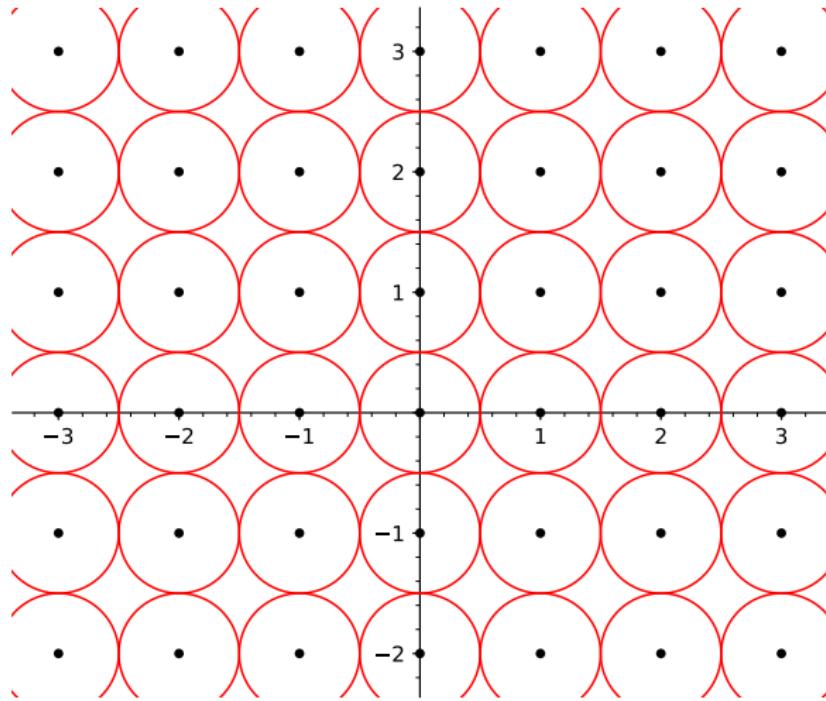


Figure: The union C of all circles in \mathbb{R}^2 centered at the points in \mathbb{Z}^2 and of radius $1/2$ is not a 1-dimensional manifold, since the points where circles touch don't admit a chart as on the previous slide. (Nevertheless we can integrate functions on C with respect to arc length; cf. subsequent section on C^1 -surfaces.)

Examples of Manifolds

Example (Graphs of C^1 -functions)

If $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is a C^1 -map then G_f is an n -dimensional differentiable manifold in \mathbb{R}^{m+n} . In order to see this, let $U = V = D \times \mathbb{R}^m$ and

$$\phi(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y} - f(\mathbf{x})) \quad \text{for } (\mathbf{x}, \mathbf{y}) \in D \times \mathbb{R}^m.$$

$\phi: U \rightarrow V$ is a diffeomorphism and $\phi(G_f) = D \times \{\mathbf{0}\}$, as required.

Example (Level sets of C^1 -functions)

Suppose $f: D \rightarrow \mathbb{R}^m$, $D \subseteq \mathbb{R}^n$, is a C^1 -map, $\mathbf{c} \in \mathbb{R}^m$, and

$$M = N_f(\mathbf{c}) = \{\mathbf{x} \in D; f(\mathbf{x}) = \mathbf{c}\} \neq \emptyset.$$

If for all $\mathbf{x} \in M$ the matrix $\mathbf{J}_f(\mathbf{x})$ has full row rank m then M is an $(n - m)$ -dimensional differentiable manifold in \mathbb{R}^n .

Definition

Two immersions $\gamma_1: \Omega_1 \rightarrow \mathbb{R}^n$ and $\gamma_2: \Omega_2 \rightarrow \mathbb{R}^n$ are said to be *equivalent* if there exists a diffeomorphism $T: \Omega_1 \rightarrow \Omega_2$ satisfying $\gamma_1 = \gamma_2 \circ T$.

Lemma

Suppose M is a d -dimensional differentiable manifold in \mathbb{R}^n and $\phi: U \rightarrow V$ is a chart of M .

- 1 There exists an immersion $\gamma: \Omega \rightarrow \mathbb{R}^n$, $\Omega \subseteq \mathbb{R}^d$, “exactly” parametrizing the chart region $M \cap U$, i.e., γ is bijective and $\gamma(\Omega) = M \cap U$.
- 2 If γ_1, γ_2 are immersions parametrizing $M \cap U$ as in (1) then γ_1 and γ_2 are equivalent.

Note

Part (2) of the lemma is very important, since it will allow us to compare the surface integrals defined by different immersions.

Proof the lemma.

We only prove the easier Part (1).

Let $\Omega \subseteq \mathbb{R}^d$ be defined by $\phi(M \cap U) = \Omega \times \{\mathbf{0}\}$.

The set Ω is open (since V is open in \mathbb{R}^n), and

$$\gamma: \Omega \rightarrow \mathbb{R}^n, \mathbf{u} \mapsto \phi^{-1}(\mathbf{u}, \mathbf{0})$$

maps Ω bijectively to $M \cap U$.

The matrix $\mathbf{J}_\gamma(\mathbf{u})$ consists of the first d columns of the invertible matrix $\mathbf{J}_{\phi^{-1}}(\mathbf{u}, \mathbf{0})$ and hence has rank d . □

Volume of d -Dimensional Parallelepipeds in \mathbb{R}^n

Definition

Let $\mathbf{v}_1, \dots, \mathbf{v}_d$ be vectors in \mathbb{R}^n . The *parallelepiped* $P = P(\mathbf{v}_1, \dots, \mathbf{v}_d) \subset \mathbb{R}^n$ spanned by $\mathbf{v}_1, \dots, \mathbf{v}_d$ is defined as

$$P = \{c_1\mathbf{v}_1 + \cdots + c_d\mathbf{v}_d; 0 \leq c_i \leq 1\}.$$

We would like to assign a d -dimensional volume $\text{vol}_d(P)$ to such parallelepipeds P . For $d < n$ we cannot use the Lebesgue measure, since clearly $\text{vol}(P) = 0$ in this case.

The function $\text{vol}_d(P) = \text{vol}_d(\mathbf{v}_1, \dots, \mathbf{v}_d)$ should have the following properties:

- ① $\text{vol}_d(P) = 1$ if $\mathbf{v}_1, \dots, \mathbf{v}_d$ are orthonormal, i.e.,
 $\mathbf{v}_i \cdot \mathbf{v}_i = |\mathbf{v}_i|^2 = 1$ and $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ for $i \neq j$;
- ② $\text{vol}_d(P)$ does not change if \mathbf{v}_i and \mathbf{v}_j are interchanged.
- ③ $\text{vol}_d(Q) = |\lambda| \text{vol}_d(P)$ if Q arises from P by multiplying \mathbf{v}_i by $\lambda \in \mathbb{R}$.
- ④ $\text{vol}_d(Q) = \text{vol}_d(P)$ if Q arises from P by adding $\lambda \mathbf{v}_i$ to \mathbf{v}_j for some $i \neq j$ and $\lambda \in \mathbb{R}$.

Theorem

Let n and d be positive integers with $1 \leq d \leq n$. There exists exactly one function $\text{vol}_d: (\mathbb{R}^n)^d \rightarrow \mathbb{R}$ having the indicated 4 properties. The function is given by

$$\text{vol}_d(\mathbf{v}_1, \dots, \mathbf{v}_d) = \sqrt{\det(\mathbf{A}^\top \mathbf{A})}, \quad \text{where } \mathbf{A} = (\mathbf{v}_1 | \dots | \mathbf{v}_d).$$

Notes

- The entries of the matrix $\mathbf{A}^\top \mathbf{A} \in \mathbb{R}^{d \times d}$ are the pairwise dot products $\mathbf{v}_i \cdot \mathbf{v}_j$. It is known as the *Gram matrix* of $\mathbf{v}_1, \dots, \mathbf{v}_d$.
- The special case $n = 3, d = 2$ gives a hint for finding this formula. In this case we know from our discussion of the cross product that $\text{vol}_2(\mathbf{v}_1, \mathbf{v}_2) = |\mathbf{v}_1 \times \mathbf{v}_2|$. This gives

$$\begin{aligned}\text{vol}_2(\mathbf{v}_1, \mathbf{v}_2)^2 &= |\mathbf{v}_1 \times \mathbf{v}_2|^2 = |\mathbf{v}_1|^2 |\mathbf{v}_2|^2 \sin^2 \phi \\ &= |\mathbf{v}_1|^2 |\mathbf{v}_2|^2 (1 - \cos^2 \phi) = |\mathbf{v}_1|^2 |\mathbf{v}_2|^2 - (\mathbf{v}_1 \cdot \mathbf{v}_2)^2 \\ &= \begin{vmatrix} \mathbf{v}_1 \cdot \mathbf{v}_1 & \mathbf{v}_1 \cdot \mathbf{v}_2 \\ \mathbf{v}_2 \cdot \mathbf{v}_1 & \mathbf{v}_2 \cdot \mathbf{v}_2 \end{vmatrix}.\end{aligned}$$

Proof of the theorem.

Prop. (3) and (4) together imply that $\text{vol}_d(P) = 0$ if $\mathbf{v}_1, \dots, \mathbf{v}_d$ are linearly dependent. Since $\text{rk}(\mathbf{A}^T \mathbf{A}) \leq \text{rk } \mathbf{A} < d$, we also have $\det(\mathbf{A}^T \mathbf{A}) = 0$ in this case. Thus we can assume from now on that $\mathbf{v}_1, \dots, \mathbf{v}_d$ are linearly independent.

Now we use Gram-Schmidt orthogonalization of \mathbf{A} . Using a sequence of elementary column operations we can transform \mathbf{A} into a matrix $\mathbf{Q} = (\mathbf{u}_1 | \dots | \mathbf{u}_d)$ satisfying

$\mathbf{Q}^T \mathbf{Q} = (\mathbf{u}_i \cdot \mathbf{u}_j)_{1 \leq i, j \leq d} = \mathbf{I}_d$. The column operations are afforded by corresponding elementary matrices $\mathbf{E}_1, \dots, \mathbf{E}_r$ (i.e.,

$$\mathbf{A} \mapsto \mathbf{A} \mathbf{E}_1 \mapsto \mathbf{A} \mathbf{E}_1 \mathbf{E}_2 \mapsto \cdots \mapsto \mathbf{A} \mathbf{E}_1 \mathbf{E}_2 \cdots \mathbf{E}_r = \mathbf{Q}.$$

Inverting, we get

$$\mathbf{A} = \mathbf{Q} \mathbf{E}_r^{-1} \cdots \mathbf{E}_2^{-1} \mathbf{E}_1^{-1} = \mathbf{Q} \mathbf{R} \quad \text{with} \quad \mathbf{R} = \mathbf{E}_r^{-1} \cdots \mathbf{E}_2^{-1} \mathbf{E}_1^{-1}.$$

On one hand we now have

$$\sqrt{\det(\mathbf{A}^T \mathbf{A})} = \sqrt{\det(\mathbf{R}^T \mathbf{Q}^T \mathbf{Q} \mathbf{R})} = \sqrt{\det(\mathbf{R}^T \mathbf{R})} = \sqrt{\det(\mathbf{R})^2} = \det(\mathbf{R})$$

and on the other

Proof cont'd.

$$\begin{aligned}\text{vol}_d(\mathbf{v}_1, \dots, \mathbf{v}_d) &= \text{vol}(\mathbf{u}_1, \dots, \mathbf{u}_d) \det(E_r^{-1}) \cdots \det(E_1^{-1}) \\ &= \det(E_r^{-1}) \cdots \det(E_1^{-1}) \\ &= \det(\mathbf{R}),\end{aligned}$$

since for the elementary column operations used and their inverses, respectively, the volume change equals the determinant of the corresponding elementary matrix. (Here we use that during Gram-Schmidt orthogonalization of a matrix columns are only multiplied by scalars $\lambda > 0$. This fact also accounts for $\det(\mathbf{R}) > 0$.) □

Remark

We have only considered “linear” parallelepipeds, which have one vertex at the origin. A general (“affine”) parallelepiped is obtained from a linear one by translation. It has the form

$P = \{\mathbf{a} + \sum_{i=1}^d c_i \mathbf{v}_i; 0 \leq c_i \leq 1\}$ and of course the same volume as its linear counterpart $P - \mathbf{a}$.

Integration over Chart Regions

Suppose that $M_0 \subseteq M$ is a chart region in a d -dimensional manifold $M \subseteq \mathbb{R}^n$, $\gamma: \Omega \rightarrow M_0$ is parametrization of M_0 by a bijective immersion and $f: M_0 \rightarrow \mathbb{R}$ is a function.

Idea

Reduce surface integration of f over M_0 to ordinary Lebesgue integration of the composition $f \circ \gamma: \Omega \rightarrow \mathbb{R}$ over Ω .

We cannot use $f \circ \gamma$ directly, as the special case of a characteristic function f and a (one-to-one) linear map γ shows:

Lemma

Suppose γ is linear, $\gamma(\mathbf{u}) = \mathbf{A}\mathbf{u}$ for $\mathbf{u} \in \Omega$. Then for every d -dimensional parallelepiped $Q \subseteq \Omega$ we have

$$\text{vol}_d(\gamma(Q)) = \sqrt{\det(\mathbf{A}^\top \mathbf{A})} \cdot \text{vol}(Q).$$

Proof.

We may assume that $\mathbf{0} \in \Omega$ and Q is linear. If Q is spanned by $\mathbf{b}_1, \dots, \mathbf{b}_d \in \mathbb{R}^d$ then $P = \gamma(Q)$ is spanned by $\mathbf{A}\mathbf{b}_1, \dots, \mathbf{A}\mathbf{b}_d \in \mathbb{R}^n$ (in particular P is a parallelepiped as well, so that $\text{vol}_d(P)$ is defined).

Proof cont'd.

The vectors \mathbf{Ab}_i are the columns of \mathbf{AB} , and hence

$$\begin{aligned}\text{vol}_d(P) &= \sqrt{\det((\mathbf{AB})^T \mathbf{AB})} = \sqrt{\det(\mathbf{B}^T \mathbf{A}^T \mathbf{AB})} \\ &= \sqrt{\det(\mathbf{B}^T) \det(\mathbf{A}^T \mathbf{A}) \det(\mathbf{B})} \quad (\text{since } \mathbf{B} \text{ is square}) \\ &= \sqrt{\det(\mathbf{A}^T \mathbf{A})} \cdot |\det \mathbf{B}| \quad (\text{since } \det(\mathbf{B}^T) = \det(\mathbf{B})) \\ &= \sqrt{\det(\mathbf{A}^T \mathbf{A})} \cdot \text{vol}(Q).\end{aligned}$$

□

The lemma shows that volumes under linear immersions scale by a nontrivial factor, which has to be accounted for when defining surface integrals.

Nonlinear case

Locally γ is well approximated by its differential $d\gamma$, which is a linear map, so that the preceding lemma can be applied.

We can decompose Ω into small d -dimensional intervals (special parallelepipeds) Q_i , choose $\mathbf{u}_i \in Q_i$, and approximate the surface part $\gamma(Q_i)$ “above Q_i ” by the parallelepiped $P_i = d\gamma(\mathbf{u}_i)(Q_i)$, which has volume $\text{vol}_d(P_i) = \sqrt{\det(\mathbf{J}_\gamma(\mathbf{u}_i)^T \mathbf{J}_\gamma(\mathbf{u}_i))} \cdot \text{vol}(Q_i)$.

Definition

The quantity $g^\gamma(\mathbf{u}) = \det(\mathbf{J}_\gamma(\mathbf{u})^T \mathbf{J}_\gamma(\mathbf{u}))$ is called *Gram determinant* of γ in $\mathbf{u} \in \Omega$.

The preceding considerations motivate taking

$$\sum_i f(\mathbf{p}_i) \text{vol}_d(P_i) = \sum_i f(\gamma(\mathbf{u}_i)) \sqrt{g^\gamma(\mathbf{u}_i)} \cdot \text{vol}(Q_i)$$

as an approximation for the surface integral of f over M_0 . Since the right-hand side also provides an approximation to the integral of $\mathbf{u} \mapsto f(\gamma(\mathbf{u})) \sqrt{g^\gamma(\mathbf{u})}$ over Ω , it is reasonable to define $\int_{M_0} f \, dS = \int_\Omega f(\gamma(\mathbf{u})) \sqrt{g^\gamma(\mathbf{u})} \, d\mathbf{u}$. (In place of $d^d \mathbf{u}$ we just write $d\mathbf{u}$.)

Question

Does this definition depend on the choice of the surface parametrization?

Fortunately the answer is “No”, as shown in the following fundamental

Lemma

Suppose M_0 is a chart region of a d -dimensional manifold $M \subseteq \mathbb{R}^n$, $\gamma_1: \Omega_1 \rightarrow M_0$ and $\gamma_2: \Omega_2 \rightarrow M_0$ are bijective immersions parametrizing M_0 , and $f: M_0 \rightarrow \mathbb{R}$ is a function. Then $(f \circ \gamma_1)\sqrt{g^{\gamma_1}}$ is integrable over Ω_1 iff $(f \circ \gamma_2)\sqrt{g^{\gamma_2}}$ is integrable over Ω_2 . If this is the case, we have

$$\int_{\Omega_1} f(\gamma_1(\mathbf{u})) \sqrt{g^{\gamma_1}(\mathbf{u})} d\mathbf{u} = \int_{\Omega_2} f(\gamma_2(\mathbf{v})) \sqrt{g^{\gamma_2}(\mathbf{v})} d\mathbf{v}.$$

Proof.

Since γ_1 and γ_2 are equivalent, there exists a diffeomorphism $T: \Omega_1 \rightarrow \Omega_2$ such that $\gamma_1 = \gamma_2 \circ T$.

$$\begin{aligned}\implies & \mathbf{J}_{\gamma_1}(\mathbf{u}) = \mathbf{J}_{\gamma_2}(T(\mathbf{u})) \mathbf{J}_T(\mathbf{u}) \\ \implies & \mathbf{J}_{\gamma_1}(\mathbf{u})^T \mathbf{J}_{\gamma_1}(\mathbf{u}) = \mathbf{J}_T(\mathbf{u})^T \mathbf{J}_{\gamma_2}(T(\mathbf{u}))^T \mathbf{J}_{\gamma_2}(T(\mathbf{u})) \mathbf{J}_T(\mathbf{u}) \\ \implies & \sqrt{g^{\gamma_1}(\mathbf{u})} = \sqrt{g^{\gamma_2}(T(\mathbf{u}))} \cdot |\det \mathbf{J}_T(\mathbf{u})|\end{aligned}$$

With this at hand, an application of the change-of-variables theorem finishes the proof:

$$\begin{aligned}\int_{\Omega_2} f(\gamma_2(\mathbf{v})) \sqrt{g^{\gamma_2}(\mathbf{v})} d\mathbf{v} &= \int_{\Omega_1} f(\gamma_2(T(\mathbf{u}))) \sqrt{g^{\gamma_2}(T(\mathbf{u}))} |\det \mathbf{J}_T(\mathbf{u})| d\mathbf{u} \\ &= \int_{\Omega_1} f(\gamma_1(\mathbf{u})) \sqrt{g^{\gamma_1}(\mathbf{u})} d\mathbf{u},\end{aligned}$$

since $\gamma_2 \circ T = \gamma_1$



Definition (Integration over a chart region)

Let M_0 be a chart region of a d -dimensional manifold $M \subseteq \mathbb{R}^n$ and $\gamma: \Omega \rightarrow M_0$ be any bijective immersion parametrizing M_0 .

- ① A function $f: M_0 \rightarrow \mathbb{R} \cup \{\pm\infty\}$ is said to be integrable over M_0 if $(f \circ \gamma)\sqrt{g^\gamma}$ is integrable over Ω . If this is the case, the surface integral of f over M_0 is defined as

$$\int_{M_0} f \, dS = \int_{\Omega} f(\gamma(\mathbf{u})) \sqrt{g^\gamma(\mathbf{u})} \, d\mathbf{u}.$$

- ② If the constant function 1 is integrable over M_0 then the d -dimensional volume of M_0 is defined as

$$\text{vol}_d(M_0) = \int_{M_0} 1 \, dS = \int_{\Omega} \sqrt{g^\gamma(\mathbf{u})} \, d\mathbf{u}.$$

Example (Volume of the (slotted) unit sphere in \mathbb{R}^3)

The slotted unit sphere

$$S_0^2 = \{(x, y, z) \in \mathbb{R}^3; x^2 + y^2 + z^2 = 1, y \neq 0 \vee x < 0\}$$

is a chart region of S^2 . We have seen that it is parametrized by the bijective immersion

$$\gamma(\theta, \phi) = \begin{pmatrix} \cos \theta \sin \phi \\ \sin \theta \sin \phi \\ \cos \phi \end{pmatrix}, \quad (\theta, \phi) \in (0, 2\pi) \times (0, \pi).$$

From

$$\mathbf{J}_\gamma(\theta, \phi) = \begin{pmatrix} -\sin \theta \sin \phi & \cos \theta \cos \phi \\ \cos \theta \sin \phi & \sin \theta \cos \phi \\ 0 & -\sin \phi \end{pmatrix}$$

we obtain

$$g^\gamma(\theta, \phi) = \det \begin{pmatrix} \sin^2 \phi & 0 \\ 0 & 1 \end{pmatrix} = \sin^2 \phi, \quad \sqrt{g^\gamma(\theta, \phi)} = \sin \phi,$$

$$\text{vol}_2(S_0^2) = \int_{(0,2\pi) \times (0,\pi)} \sin \phi \, d^2(\theta, \phi) = 2\pi \int_0^\pi \sin \phi \, d\phi = 4\pi.$$

Example (cont'd)

As an example for a surface integral that is not a volume we compute $\int_{S_0^2} x^2 dS$.

$$\begin{aligned}\int_{S_0^2} x^2 dS &= \int_{(0,2\pi) \times (0,\pi)} (\cos \theta \sin \phi)^2 \sin \phi d^2(\theta, \phi) \\ &= \int_0^{2\pi} \cos^2 \theta d\theta \int_0^\pi \sin^3 \phi d\phi \\ &= \pi \times \frac{4}{3}.\end{aligned}$$

Remark

The 0-meridian $S^2 \setminus S_0^2$ turns out to have 2-dimensional measure zero, and the preceding computations give in fact also $\text{vol}_2(S^2) = \text{vol}_2(S_0^2) = 4\pi$ and $\int_{S^2} x^2 dS = \int_{S_0^2} x^2 dS = 4\pi/3$.

Question

Can you derive the relation $\int_{S^2} x^2 dS = \frac{1}{3} \text{vol}_2(S^2)$ directly?

Example (Volume of a torus)

We have seen that a smooth curve $\alpha(t) = (r(t), z(t))$, $t \in I$, with $r(t) \neq 0$ for all $t \in I$ generates a surface of revolution parametrized by the immersion

$$\gamma(\theta, t) = \begin{pmatrix} r(t) \cos \theta \\ r(t) \sin \theta \\ z(t) \end{pmatrix}, \quad (\theta, t) \in \mathbb{R} \times I.$$

For surface integration we must assume that α is one-to-one and, since $\gamma(\theta, t) = \gamma(\theta + 2\pi, t)$, restrict the parameter (θ, t) to $\Omega = (0, 2\pi) \times I$, say, which cuts out the 0-meridian.

The volume (surface area) of the slotted surface of revolution S is then obtained as follows:

$$\mathbf{J}_\gamma(\theta, t) = \begin{pmatrix} -r(t) \sin \theta & r'(t) \cos \theta \\ r(t) \cos \theta & r'(t) \sin \theta \\ 0 & z'(t) \end{pmatrix},$$

$$g^\gamma(\theta, t) = \det \begin{pmatrix} r(t)^2 & 0 \\ 0 & r'(t)^2 + z'(t)^2 \end{pmatrix} = r(t)^2 |\alpha'(t)|^2$$

Example (cont'd)

$$\begin{aligned}\text{vol}_2(S) &= \int_{(0,2\pi) \times I} r(t) |\alpha'(t)| d^2(\theta, t) = 2\pi \int_I r(t) |\alpha'(t)| dt \\ &= 2\pi \int_I r(t) \sqrt{r'(t)^2 + z'(t)^2} dt.\end{aligned}$$

For the (slotted) torus $T = T(A, a)$ we have $I = (0, 2\pi)$,
 $\alpha(t) = (r(t), z(t)) = (A + a \cos t, a \sin t)$.

$\implies |\alpha'(t)| = a$ and

$$\text{vol}_2(T) = 2\pi a \int_0^{2\pi} A + a \cos t dt = 4\pi^2 a A.$$

Example (Integration over Graphs)

Suppose $g: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, is a C^1 -function. We have seen that the graph G_g is an n -dimensional manifold in \mathbb{R}^{n+1} (hypersurface) and is parametrized by the (trivially bijective) immersion $\gamma(\mathbf{x}) = (\mathbf{x}, g(\mathbf{x}))$.

$$\begin{aligned}\implies \mathbf{J}_\gamma(\mathbf{x}) &= \begin{pmatrix} \mathbf{I}_n \\ \nabla g(\mathbf{x})^\top \end{pmatrix} \\ \implies g^\gamma(\mathbf{x}) &= \det \left[\begin{pmatrix} \mathbf{I}_n & \nabla g(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \mathbf{I}_n \\ \nabla g(\mathbf{x})^\top \end{pmatrix} \right] \\ &= \det(\mathbf{I}_n + \nabla g(\mathbf{x}) \nabla g(\mathbf{x})^\top) \\ &= 1 + |\nabla g(\mathbf{x})|^2.\end{aligned}\quad \text{(after some work)}$$

Hence a function $f: G_g \rightarrow \mathbb{R}$ is integrable over G_g iff

$\mathbf{x} \mapsto f(\mathbf{x}, g(\mathbf{x})) \sqrt{1 + |\nabla g(\mathbf{x})|^2}$ is integrable over D , and if this is the case then

$$\int_{G_g} f \, dS = \int_D f(\mathbf{x}, g(\mathbf{x})) \sqrt{1 + |\nabla g(\mathbf{x})|^2} d^n \mathbf{x}.$$

Example (cont'd)

As an example we compute the volume σ_n of the $(n - 1)$ -dimensional unit sphere $S^{n-1} = S_1(\mathbf{0}) \subseteq \mathbb{R}^n$.

The positive halfsphere

$$S_+^{n-1} = \{\mathbf{x} \in \mathbb{R}^n; x_1^2 + \cdots + x_{n-1}^2 + x_n^2 = 1, x_n > 0\}$$

is the graph of

$$g: B_1(\mathbf{0}) \rightarrow \mathbb{R}, (x_1, \dots, x_{n-1}) \mapsto \sqrt{1 - x_1^2 - \cdots - x_{n-1}^2},$$

where $B_1(\mathbf{0})$ denotes the unit ball in \mathbb{R}^{n-1} . We compute

$$\nabla g(x_1, \dots, x_{n-1}) = -\frac{1}{\sqrt{1 - x_1^2 - \cdots - x_{n-1}^2}}(x_1, \dots, x_{n-1}),$$

$$1 + |\nabla g(x_1, \dots, x_{n-1})|^2 = \frac{1}{1 - x_1^2 - \cdots - x_{n-1}^2}, \quad \text{and hence}$$

$$\text{vol}_{n-1}(S_+^{n-1}) = \int_{S_+^{n-1}} 1 \, dS = \int_{B_1(\mathbf{0})} \frac{1}{\sqrt{1 - x_1^2 - \cdots - x_{n-1}^2}} \, d^{n-1} \mathbf{x}$$

Example (cont'd)

Using, e.g., the formula for integrating rotation-invariant functions, this reduces to a 1-dimensional integral:

$$\begin{aligned}\text{vol}_{n-1}(S_+^{n-1}) &= (n-1)\beta_{n-1} \int_0^1 \frac{r^{n-2}}{\sqrt{1-r^2}} dr \\ &= (n-1)\beta_{n-1} \int_0^{\pi/2} \sin^{n-2} t dt \quad (\text{Subst. } r = \sin t)\end{aligned}$$

Earlier we had shown that $S_n = \int_0^{\pi/2} \sin^n t dt$ satisfies the recurrence relation $S_n = \frac{n-1}{n} S_{n-2}$ and is related to β_n by $\beta_n = 2\beta_{n-1} S_n$. Moreover, $\sigma_n = \text{vol}_{n-1}(S_+^{n-1}) + \text{vol}_{n-1}(S_-^{n-1}) = 2 \text{vol}_{n-1}(S_+^{n-1})$, since the missing part (a sphere of dimension $n-2$) has $(n-1)$ -dimensional volume zero; cf. subsequent discussion.

$$\implies \sigma_n = 2(n-1)\beta_{n-1} S_{n-2} = 2n\beta_{n-1} S_n = n\beta_n.$$

n	1	2	3	4	5	6	7	8	9	10
σ_n	2	2π	4π	$2\pi^2$	$\frac{8}{3}\pi^2$	π^3	$\frac{16}{15}\pi^3$	$\frac{1}{3}\pi^4$	$\frac{32}{105}\pi^4$	$\frac{1}{12}\pi^5$
\approx	2	6.28	12.57	19.74	26.32	31.01	33.07	32.47	29.69	25.50

Integration over Manifolds

Integration over a d -dimensional manifold $M \subseteq \mathbb{R}^n$ can be reduced to integration over its chart regions by means of a technical device called “partition of unity”.

Definition

Suppose M_1, M_2, M_3, \dots is a (finite or) denumerable covering of M (i.e., $\bigcup_{k=1}^{\infty} M_k = M$) by chart regions. (The existence of such an atlas can be shown for any manifold in \mathbb{R}^n). There exists a sequence $\epsilon_1, \epsilon_2, \epsilon_3, \dots$ of continuous functions $\epsilon_i: M \rightarrow [0, 1]$ having the following properties:

- ① Every point $\mathbf{x} \in M$ has a neighborhood on which only finitely many functions ϵ_i are nonzero.
- ② $\sum_{i=1}^{\infty} \epsilon_i(\mathbf{x}) = 1$ for all $\mathbf{x} \in M$ (by (1), these sums are finite);
- ③ $\epsilon_i(\mathbf{x}) = 0$ if $\mathbf{x} \in M \setminus M_i$

The sequence of functions (ϵ_i) is called a *partition of unity*. From (1), (2) we have that every function $f: M \rightarrow \mathbb{R}$ can be written as $f = \sum_{i=1}^{\infty} f \epsilon_i$ (no questions of convergence arise), and from (3) that $f \epsilon_i$ vanishes outside the chart region M_i .

Definition

A function $f: M \rightarrow \mathbb{R}$ is said to be *integrable* over M if

- ① for each $i = 1, 2, 3, \dots$ the function $f\epsilon_i$ is integrable over the chart region M_i , and
- ② $\sum_{i=1}^{\infty} \int_{M_i} |f| \epsilon_i dS < \infty$.

If this is the case then we define

$$\int_M f dS = \sum_{i=1}^{\infty} \int_{M_i} f\epsilon_i dS.$$

Notes

- Neither the property “ f is integrable” nor the value of $\int_M f dS$ assigned in the definition depend on the particular choice of atlas or partition of unity.
- The surface integral for d -dimensional manifolds inherits many of the properties of the Lebesgue integral. For example, the Monotone Convergence Theorem and the Bounded Convergence Theorem also hold for the surface integral. For the definition of measurability and volume see the next slide.

Surface Volume

The following closely parallels the definition of Lebesgue measure in terms of the Lebesgue integral, as we have stated it earlier.
First we need a familiar technical concept.

Definition

Let M be a d -dimensional manifold in \mathbb{R}^n and $A \subseteq M$. A function $f: D \rightarrow \mathbb{R}$ with domain $D \supseteq A$ contains A is said to be *integrable over A* if $f_A: M \rightarrow \mathbb{R}$, $\mathbf{x} \mapsto f(\mathbf{x})\chi_A(\mathbf{x})$ is integrable over M . If this is the case, we define $\int_A f dS = \int_M f_A dS$.

Definition (d -dimensional volume)

Let M be a d -dimensional manifold in \mathbb{R}^n . A subset $A \subseteq M$ is *measurable* if the constant function 1 is integrable over A . If applicable, we call

$$\text{vol}_d(A) = \int_A 1 dS$$

the *d -dimensional volume of A* .

Integration over C^1 -Surfaces

Not all interesting surfaces in \mathbb{R}^n are manifolds. For example, we would like to define the 2-dimensional volume of the surface of a cube in \mathbb{R}^3 . The edges and vertices of the cube do not have neighborhoods cutting chart regions out of the cube.

⇒ A cube is not a manifold.

Definition

Let n, d be integers with $1 \leq d \leq n$. A set $N \subset \mathbb{R}^n$ is said to have *d-dimensional measure* (or volume) zero if for every $\epsilon > 0$ there exists a sequence Q_1, Q_2, Q_3, \dots of n -dimensional cubes with side lengths r_1, r_2, r_3, \dots satisfying $N \subseteq \bigcup_{k=1}^{\infty} Q_k$ and $\sum_{k=1}^{\infty} r_k^d < \epsilon$.

Notes

- In the case $d = n$ this definition reduces to a characterization of sets in \mathbb{R}^n of Lebesgue measure zero given earlier.
- If N has d -dimensional measure zero and $d' > d$ then N has d' -dimensional measure zero as well.
- Smooth surfaces (manifolds) of dimension $d - 1$ or less have d -dimensional measure zero. In particular the vertices and edges of a cube have 2-dimensional measure zero.

Notes cont'd

- If N is a subset of a d -dimensional manifold M in \mathbb{R}^n and N has d -dimensional measure zero, then N behaves with respect to surface integration on M in the same way as sets of Lebesgue measure zero do with respect to ordinary Lebesgue integration. For example, changing the values of a function $f: M \rightarrow \mathbb{R}$ on such a set N does not affect integrability of f over M , nor does it change $\int_M f \, dS$ in case that f is integrable.

The last note provides the main justification for the preceding definition, which is not easy to understand. (Note that r_k^d is not the volume of Q_k , which is r_k^n , but rather the volume of its d -dimensional projections.)

The last note also shows that our previous computation $\text{vol}_2(S^2) = 4\pi$, which in fact showed only that the slotted unit sphere parametrized using spherical coordinates has 2-dimensional volume 2π , was nevertheless correct, since the meridian left out has 2-dimensional volume zero.

Definition

A subset $X \subseteq \mathbb{R}^n$ is called a d -dimensional C^1 -surface if there exists a d -dimensional manifold $M \subseteq X$ such that

- ① $X \setminus M$ has d -dimensional measure zero.
- ② The boundary of M in X is equal to $X \setminus M$.

The largest subset M of X having these properties is called the *regular* (or *smooth*) part of X and denoted by $M(X)$. The complementary set $X \setminus M(X)$ is called *singular* part of X .

Definition (Integration over C^1 -surfaces)

Let $X \subseteq \mathbb{R}^n$ be a d -dimensional C^1 -surface.

- A function $f: X \rightarrow \mathbb{R}$ is said to be integrable over X if f is integrable over $M(X)$. If this is the case, we define
$$\int_X f \, dS = \int_{M(X)} f \, dS.$$
- X is said to *measurable* if $M(X)$ is measurable. If applicable, we set $\text{vol}_d(X) = \text{vol}_d(M(X))$.

Note

It is not necessary to know $M(X)$. In the preceding definitions $M(X)$ can be replaced by any d -dimensional manifold $M \subset X$ having Properties (1) and (2) above.

Example

We compute the surface integral of $f(x, y, z) = xyz$ over the boundary ∂C of the 3-dimensional unit cube $C = [0, 1]^3$.

∂C is a C^1 -surface with singular part formed by the vertices and edges of the unit cube. The smooth part $M(\partial C)$ consists of 6 pieces $M_1, M_2, M_3, M_4, M_5, M_6$, which are the open faces of the unit cube and correspond to the six choices $x \in \{0, 1\}$ (M_1, M_2), $y \in \{0, 1\}$ (M_3, M_4), $z \in \{0, 1\}$ (M_5, M_6).

$$xyz = 0 \text{ on } M_1, M_3, M_5 \implies \int_{M_1} xyz \, dS = \int_{M_3} xyz \, dS = \int_{M_5} xyz \, dS = 0.$$

M_2 admits the parametrization $\gamma(y, z) = (1, y, z)$, which has

$$\mathbf{J}_\gamma(y, z) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ and Gram determinant 1.}$$

$$\implies \int_{M_2} xyz \, dS = \int_{[0,1]^2} 1yz \, d^2(y, z) = 1/4.$$

Similarly we obtain $\int_{M_4} xyz \, dS = \int_{M_6} xyz \, dS = 1/4$ (this also follows by symmetry), and hence $\int_{\partial C} xyz \, dS = 3/4$.

Final Examination

Date/Time/Venue

Wed Jan 10 2024, 9:00–12:00

Instructions to candidates

- This examination paper contains six (6) questions.
- Please answer every question and subquestion, and **justify** your answers.
- For your answers please use the space provided after each question. If this space is insufficient, please continue on the blank sheets provided.
- This is a **CLOSED BOOK** examination, except that you may bring 1 sheet of A4 paper (hand-written only) and a Chinese-English dictionary (paper copy only) to the examination.

The End

We wish you every success in the final examination!

Math 241
Calculus III

Thomas
Honold

Differential
Forms
(optional)

Alternating Forms
Differential Forms
Oriented Manifolds
Integration of
Differential Forms
The General Stokes
Theorem

Math 241 Calculus III

Thomas Honold



ZJU-UIUC Institute



Fall Semester 2023

Differential
Forms
(optional)

Alternating Forms
Differential Forms
Oriented Manifolds
Integration of
Differential Forms
The General Stokes
Theorem

Outline I

1 Differential Forms (optional)

- Alternating Forms
- Differential Forms
- Oriented Manifolds
- Integration of Differential Forms
- The General Stokes Theorem

Differential
Forms
(optional)

Alternating Forms
Differential Forms
Oriented Manifolds
Integration of
Differential Forms
The General Stokes
Theorem

Today's Lecture:

Definition

An *alternating k-form* (“constant differential k-form”) on \mathbb{R}^n is a map $\omega: (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ satisfying the following conditions:

- A1 ω is linear in each argument (*k-multilinear*), i.e.,
 $\omega(\mathbf{v}_1 + \mathbf{v}'_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = \omega(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) + \omega(\mathbf{v}'_1, \mathbf{v}_2, \dots, \mathbf{v}_k),$
 $\omega(c\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = c\omega(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ for all
 $\mathbf{v}_1, \mathbf{v}'_1, \mathbf{v}_2, \dots, \mathbf{v}_k \in \mathbb{R}^n, c \in \mathbb{R}$, and similar for the other arguments.
- A2 $\omega(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k) = 0$ if there exist $i < j$ such that $\mathbf{v}_i = \mathbf{v}_j$.

Notes

- An alternating *k*-form changes sign if two variables are interchanged, e.g.,

$$\omega(\mathbf{v}_2, \mathbf{v}_1, \mathbf{v}_3, \dots, \mathbf{v}_k) = -\omega(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_k).$$

This follows from

$$\begin{aligned} 0 &= \omega(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_2, \dots) \\ &= \omega(\mathbf{v}_1, \mathbf{v}_1, \dots) + \omega(\mathbf{v}_1, \mathbf{v}_2, \dots) + \omega(\mathbf{v}_2, \mathbf{v}_1, \dots) + \omega(\mathbf{v}_2, \mathbf{v}_2, \dots) \\ &= \omega(\mathbf{v}_1, \mathbf{v}_2, \dots) + \omega(\mathbf{v}_2, \mathbf{v}_1, \dots). \end{aligned}$$

Notes cont'd

- For fixed k the set of all alternating k -forms on \mathbb{R}^n forms a vector space over \mathbb{R} with respect to the “point-wise” operations $(\omega_1 + \omega_2)(\mathbf{v}_1, \dots, \mathbf{v}_k) = \omega_1(\mathbf{v}_1, \dots, \mathbf{v}_k) + \omega_2(\mathbf{v}_1, \dots, \mathbf{v}_k)$, $(c\omega)(\mathbf{v}_1, \dots, \mathbf{v}_k) = c\omega(\mathbf{v}_1, \dots, \mathbf{v}_k)$. This vector space is commonly denoted by $\text{Alt}^k(\mathbb{R}^n)$.
- An alternating 0-form, by convention, is simply a constant $c \in \mathbb{R}$. For $k = 1$ condition (A2) is void, so that the alternating 1-forms are just the linear forms.

Definition (Wedge product of linear forms)

For linear forms $\phi_1, \dots, \phi_k: \mathbb{R}^n \rightarrow \mathbb{R}$ we define their *wedge product* $\phi_1 \wedge \cdots \wedge \phi_k \in \text{Alt}^k(\mathbb{R}^n)$ by

$$(\phi_1 \wedge \cdots \wedge \phi_k)(\mathbf{v}_1, \dots, \mathbf{v}_k) = \begin{vmatrix} \phi_1(\mathbf{v}_1) & \phi_1(\mathbf{v}_2) & \cdots & \phi_1(\mathbf{v}_k) \\ \phi_2(\mathbf{v}_1) & \phi_2(\mathbf{v}_2) & \cdots & \phi_2(\mathbf{v}_k) \\ \vdots & \vdots & & \vdots \\ \phi_k(\mathbf{v}_1) & \phi_k(\mathbf{v}_2) & \cdots & \phi_k(\mathbf{v}_k) \end{vmatrix}.$$

Since the determinant is an alternating k -form of its rows (or columns), this is indeed an alternating k -form.

Recalling that $dx_i(\mathbf{v}) = v_i$, we can write

$$\begin{aligned}\phi(\mathbf{v}) &= \phi(v_1\mathbf{e}_1 + \cdots + v_n\mathbf{e}_n) = v_1\phi(\mathbf{e}_1) + \cdots + v_n\phi(\mathbf{e}_n) \\ &= \phi(\mathbf{e}_1)dx_1(\mathbf{v}) + \cdots + \phi(\mathbf{e}_n)dx_n(\mathbf{v}) \\ &= (\phi(\mathbf{e}_1)dx_1 + \cdots + \phi(\mathbf{e}_n)dx_n)(\mathbf{v}),\end{aligned}$$

i.e. $\phi = \phi(\mathbf{e}_1)dx_1 + \cdots + \phi(\mathbf{e}_n)dx_n$.

This shows that the differentials dx_1, \dots, dx_n form a basis of $\text{Alt}^1(\mathbb{R}^n) = (\mathbb{R}^n)^*$ (the dual space of \mathbb{R}^n).

Lemma

The wedge products $dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}$, $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ form a basis of $\text{Alt}^k(\mathbb{R}^n)$. In particular we have

$$\dim \text{Alt}^k(\mathbb{R}^n) = \begin{cases} \binom{n}{k} & \text{if } 0 \leq k \leq n, \\ 0 & \text{if } k > n. \end{cases}$$

Here $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ denotes the binomial coefficient “ n choose k ”.

Proof.

We give the proof only for $n = 2$ and $n = 3$.

$n = 2$

The case $k = 1$ was done before the lemma. For an alternating 2-form $\omega: \mathbb{R}^2 \rightarrow \mathbb{R}$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ we have

$$\begin{aligned}\omega(\mathbf{a}, \mathbf{b}) &= \omega(a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2, b_1 \mathbf{e}_1 + b_2 \mathbf{e}_2) \\ &= a_1 b_1 \omega(\mathbf{e}_1, \mathbf{e}_1) + a_1 b_2 \omega(\mathbf{e}_1, \mathbf{e}_2) + a_2 b_1 \omega(\mathbf{e}_2, \mathbf{e}_1) + a_2 b_2 \omega(\mathbf{e}_2, \mathbf{e}_2) \\ &= (a_1 b_2 - a_2 b_1) \omega(\mathbf{e}_1, \mathbf{e}_2).\end{aligned}$$

On the other hand, we have

$$(dx_1 \wedge dx_2)(\mathbf{a}, \mathbf{b}) = \begin{vmatrix} dx_1(\mathbf{a}) & dx_1(\mathbf{b}) \\ dx_2(\mathbf{a}) & dx_2(\mathbf{b}) \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1.$$

Thus $\omega = c dx_1 \wedge dx_2$ with $c = \omega(\mathbf{e}_1, \mathbf{e}_2)$, showing that $\text{Alt}^2(\mathbb{R}^2)$ is 1-dimensional and generated by $dx_1 \wedge dx_2$.

In general we can expand $\omega \in \text{Alt}^k(\mathbb{R}^2)$ into a linear combination of terms $\omega(\mathbf{v}_1, \dots, \mathbf{v}_k)$, where each \mathbf{v}_i is either \mathbf{e}_1 or \mathbf{e}_2 . For $k \geq 3$ this implies $\omega = 0$ on account of Axiom (A2).

Proof cont'd.

 $n = 3$

Here the only interesting cases are $k = 2, 3$. (For $k \geq 4$ one shows $\omega = 0$ in the same way as for $n = 2$, now using the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ of \mathbb{R}^3 .)

For $\omega \in \text{Alt}^2(\mathbb{R}^3)$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ we obtain

$$\begin{aligned}\omega(\mathbf{a}, \mathbf{b}) &= \sum_{i,j=1}^3 a_i b_j \omega(\mathbf{e}_i, \mathbf{e}_j) \\ &= (a_1 b_2 - a_2 b_1) \omega(\mathbf{e}_1, \mathbf{e}_2) + (a_1 b_3 - a_3 b_1) \omega(\mathbf{e}_1, \mathbf{e}_3) + (a_2 b_3 - a_3 b_2) \omega(\mathbf{e}_2, \mathbf{e}_3)\end{aligned}$$

The left factors are $dx_1 \wedge dx_2$, $dx_1 \wedge dx_3$, $dx_2 \wedge dx_3$ evaluated at (\mathbf{a}, \mathbf{b}) in this order, so that

$$\begin{aligned}\omega &= \omega(\mathbf{e}_1, \mathbf{e}_2) dx_1 \wedge dx_2 + \omega(\mathbf{e}_1, \mathbf{e}_3) dx_1 \wedge dx_3 + \omega(\mathbf{e}_2, \mathbf{e}_3) dx_2 \wedge dx_3. \\ \implies \text{Alt}^2(\mathbb{R}^3) \text{ has basis } &\{dx_1 \wedge dx_2, dx_1 \wedge dx_3, dx_2 \wedge dx_3\}.\end{aligned}$$

Proof cont'd.

For $\omega \in \text{Alt}^3(\mathbb{R}^3)$ and $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ we obtain

$$\begin{aligned}\omega(\mathbf{a}, \mathbf{b}, \mathbf{c}) &= \sum_{i,j,l=1}^3 a_i b_j c_l \omega(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_l) \\ &= \left(\sum_{\pi \in S_3} (-1)^\pi a_{\pi(1)} b_{\pi(2)} c_{\pi(3)} \right) \omega(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \\ &= \omega(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.\end{aligned}$$

This shows that $\text{Alt}^3(\mathbb{R}^3)$ is 1-dimensional and generated by the determinant function $(\mathbf{a}, \mathbf{b}, \mathbf{c}) \mapsto \det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \det(\mathbf{a}|\mathbf{b}|\mathbf{c})$. □

The wedge product $dx_1 \wedge dx_2 \wedge dx_3$ is equal to the determinant:

$$(dx_1 \wedge dx_2 \wedge dx_3)(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \begin{vmatrix} dx_1(\mathbf{a}) & dx_1(\mathbf{b}) & dx_1(\mathbf{c}) \\ dx_2(\mathbf{a}) & dx_2(\mathbf{b}) & dx_2(\mathbf{c}) \\ dx_3(\mathbf{a}) & dx_3(\mathbf{b}) & dx_3(\mathbf{c}) \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

In the general case $(dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k})(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ is equal to the determinant of the $k \times k$ submatrix of $(\mathbf{v}_1 | \mathbf{v}_2 | \dots | \mathbf{v}_k)$ formed by rows i_1, i_2, \dots, i_k , and $\omega \in \text{Alt}^k(\mathbb{R}^n)$ has the representation

$$\omega = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \omega(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_k}) dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}.$$

This representation can be used to define the *wedge product* $\omega \wedge \eta$ of two arbitrary differential forms $\omega \in \text{Alt}^k(\mathbb{R}^n)$, $\eta \in \text{Alt}^l(\mathbb{R}^n)$ by

$$(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) \wedge (dx_{j_1} \wedge \cdots \wedge dx_{j_l}) = dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_l}$$

and bilinear extension, with the additional convention that $\omega \wedge \eta = \eta \wedge \omega = c \omega$ if $\eta = c \in \text{Alt}^0(\mathbb{R}^n)$.

One easily shows that $(\omega, \eta) \mapsto \omega \wedge \eta$ is well-defined, bilinear, associative, and in place of the commutative law satisfies the identity

$$\omega \wedge \eta = (-1)^{kl} \eta \wedge \omega \quad \text{for } \omega \in \text{Alt}^k(\mathbb{R}^n), \eta \in \text{Alt}^l(\mathbb{R}^n) = \mathbb{R}.$$

Physical Interpretation ($n = 3$)

$k = 1$

We have already seen that a linear form

$$\omega = c_1 dx + c_2 dy + c_3 dz \in \text{Alt}^1(\mathbb{R}^3), \text{ i.e.,}$$

$\omega(\mathbf{v}) = c_1 v_1 + c_2 v_2 + c_3 v_3 = \mathbf{c} \cdot \mathbf{v}$ for $\mathbf{v} \in \mathbb{R}^3$, is the work done by the constant force field $\mathbf{F}(\mathbf{x}) = \mathbf{c}$ on an object displaced by \mathbf{v} , i.e., moved from \mathbf{x} to $\mathbf{x} + \mathbf{v}$ for any point $\mathbf{x} \in \mathbb{R}^3$.

$k = 2$

An alternating 2-form $\omega \in \text{Alt}^2(\mathbb{R}^3)$ has a unique representation

$\omega = c_1 dy \wedge dz + c_2 dz \wedge dx + c_3 dx \wedge dy$ with $c_1, c_2, c_3 \in \mathbb{R}$. (The correspondence with the standard representation is

$$c_1 = \omega(\mathbf{e}_2, \mathbf{e}_3), c_2 = -\omega(\mathbf{e}_1, \mathbf{e}_3), c_3 = \omega(\mathbf{e}_1, \mathbf{e}_2).$$

$$\implies \omega(\mathbf{a}, \mathbf{b}) = c_1(a_2 b_3 - a_3 b_2) + c_2(a_3 b_1 - a_1 b_3) + c_3(a_1 b_2 - a_2 b_1)$$

$$= (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

the oriented volume of the parallelepiped spanned by $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

$k = 2$ cont'd

⇒ If \mathbf{c} is the vectorial velocity of a constant flow then $\omega(\mathbf{a}, \mathbf{b})$ is the oriented flow per unit time through the parallelogram spanned by \mathbf{a}, \mathbf{b} , i.e., with vertices $\mathbf{x}, \mathbf{x} + \mathbf{a}, \mathbf{x} + \mathbf{b}, \mathbf{x} + \mathbf{a} + \mathbf{b}$ for any point $\mathbf{x} \in \mathbb{R}^3$. The flow is positive if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is positively oriented (i.e., the vertices of the parallelogram are ordered counter-clock-wise when watched from the side into which the flow flows).

$k = 3$

An alternating 3-form $\omega \in \text{Alt}^2(\mathbb{R}^3)$ has a unique representation $\omega = c dx \wedge dy \wedge dz$ with $c \in \mathbb{R}$. The form $dx \wedge dy \wedge dz$ is the 3×3 determinant, and hence $\omega(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is equal to c times the oriented volume of a parallelepiped spanned by $\mathbf{a}, \mathbf{b}, \mathbf{c}$, i.e., with vertices $\mathbf{x}, \mathbf{x} + \mathbf{a}, \mathbf{x} + \mathbf{b}, \mathbf{x} + \mathbf{c}, \mathbf{x} + \mathbf{a} + \mathbf{b}, \mathbf{x} + \mathbf{a} + \mathbf{c}, \mathbf{x} + \mathbf{b} + \mathbf{c}, \mathbf{x} + \mathbf{a} + \mathbf{b} + \mathbf{c}$ for any point $\mathbf{x} \in \mathbb{R}^3$.

Definition

Suppose $D \subseteq \mathbb{R}^n$ is open. A *differential k-form* on D is a map $\omega: D \rightarrow \text{Alt}^k(\mathbb{R}^n)$, i.e., for every $\mathbf{x} \in D$ the image $\omega(\mathbf{x})$ is an alternating k -form on \mathbb{R}^n .

Notes

- This definition is in sync with our earlier definition of differential 1-forms.
- In view of our convention regarding alternating 0-forms, a differential 0-form on D is simply a function $f: D \rightarrow \mathbb{R}$.
- The standard representation of $\omega(\mathbf{x})$ in terms of the differentials $dx_{i_1} \wedge \cdots \wedge dx_{i_k}$ gives functions $f_{i_1, \dots, i_k}: D \rightarrow \mathbb{R}$ such that

$$\omega(\mathbf{x}) = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} f_{i_1, i_2, \dots, i_k}(\mathbf{x}) dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}$$

for every $\mathbf{x} \in D$. This is written as

$\omega = \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} f_{i_1, i_2, \dots, i_k} dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}$ and called
standard representation of ω .

Notes cont'd

- The wedge product of a differential k -form and a differential l -form on D is defined point-wise, $(\omega \wedge \eta)(\mathbf{x}) = \omega(\mathbf{x}) \wedge \eta(\mathbf{x})$ for $\mathbf{x} \in D$, and satisfies the same rules as in the constant case (i.e, the case of alternating forms). In particular we have $f \wedge \omega = \omega \wedge f = f\omega$ if $f: D \rightarrow \mathbb{R}$ is a function (differential 0-form).

Physical Interpretation

For continuous differential forms on open subsets of \mathbb{R}^3 the observations made for alternating forms (constant differential forms) carry over:

- For a differential 1-form $\omega = f_1 dx + f_2 dy + f_3 dz$ the quantity $\omega(\mathbf{x})(\mathbf{v})$ is approximately equal to the work done by the force field $\mathbf{F} = (f_1, f_2, f_3)$ on an object moved along the line segment $[\mathbf{x}, \mathbf{x} + \mathbf{v}]$, provided \mathbf{v} is sufficiently small.
- For a differential 2-form $\omega = f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy$ the quantity $\omega(\mathbf{x})(\mathbf{a}, \mathbf{b})$ is approximately equal to the oriented flow per unit time of a flow field with vectorial velocity $\mathbf{F} = (f_1, f_2, f_3)$ through the parallelogram $P = \{\mathbf{x} + \lambda \mathbf{a} + \mu \mathbf{b}; 0 \leq \lambda, \mu \leq 1\}$, provided P is sufficiently small.
- For a differential 3-form $\omega = f dx \wedge dy \wedge dz$ the quantity $\omega(\mathbf{x})(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is approximately equal to $f(\mathbf{x})$ times the oriented volume of the parallelepiped $P = \{\mathbf{x} + \lambda \mathbf{a} + \mu \mathbf{b} + \nu \mathbf{c}; 0 \leq \lambda, \mu, \nu \leq 1\}$, provided P is sufficiently small.

Pullback of Differential Forms

It is obvious that for an alternating k -form $\omega \in \text{Alt}^k(\mathbb{R}^m)$ and a linear map $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ the map

$$T^*\omega: \begin{cases} (\mathbb{R}^n)^k \rightarrow \mathbb{R}, \\ (\mathbf{v}_1, \dots, \mathbf{v}_k) \mapsto \omega(T(\mathbf{v}_1), \dots, T(\mathbf{v}_k)) \end{cases}$$

("composition of ω and T ") is an alternating k -form in $\text{Alt}^k(\mathbb{R}^n)$.

Definition

Suppose $D \subseteq \mathbb{R}^m$, $\Omega \subseteq \mathbb{R}^n$ are open, $T: \Omega \rightarrow D$ is differentiable, and $\omega: D \rightarrow \text{Alt}^k(\mathbb{R}^m)$ is a differential k -form. Then the differential k -form $T^*\omega: \Omega \rightarrow \text{Alt}^k(\mathbb{R}^n)$ defined by

$$\begin{aligned} (T^*\omega)(\mathbf{x})(\mathbf{v}_1, \dots, \mathbf{v}_k) &= \omega(T(\mathbf{x}))(\mathrm{d}T(\mathbf{x})(\mathbf{v}_1), \dots, \mathrm{d}T(\mathbf{x})(\mathbf{v}_k)) \\ &= \omega(T(\mathbf{x}))(\mathbf{J}_T(\mathbf{x})\mathbf{v}_1, \dots, \mathbf{J}_T(\mathbf{x})\mathbf{v}_k) \end{aligned}$$

for $\mathbf{x} \in \Omega$ and $(\mathbf{v}_1, \dots, \mathbf{v}_k) \in (\mathbb{R}^n)^k$ is called *pullback of ω along T* .

Note that $(T^*\omega)(\mathbf{x}) = (\mathrm{d}T(\mathbf{x}))^* \omega(T(\mathbf{x}))$, i.e., for each point $\mathbf{x} \in \Omega$ the differential k -form $\omega(T(\mathbf{x})): \Omega \rightarrow \text{Alt}^k(\mathbb{R}^m)$ is pulled back using the linear map $\mathrm{d}T(\mathbf{x}): \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Writing $T = (T_1, \dots, T_m)$, it is easy to see that in the constant case we have

$$T^*(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) = T^*dx_{i_1} \wedge \cdots \wedge T^*dx_{i_k} = T_{i_1} \wedge \cdots \wedge T_{i_k},$$

$$T^*\omega = \sum_{1 \leq i_1 < \cdots < i_k \leq n} c_{i_1, \dots, i_k} T_{i_1} \wedge \cdots \wedge T_{i_k},$$

where $c_{i_1, \dots, i_k} = \omega(\mathbf{e}_1, \dots, \mathbf{e}_k)$.

The pullback of a differential k -form $\omega: D \rightarrow \text{Alt}^k(\mathbb{R}^m)$ along $T: \Omega \rightarrow D$, $\Omega \subseteq \mathbb{R}^n$, therefore admits the representation

$$(T^*\omega)(\mathbf{x}) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} f_{i_1, \dots, i_k}(T(\mathbf{x})) dT_{i_1}(\mathbf{x}) \wedge \cdots \wedge dT_{i_k}(\mathbf{x}),$$

$\mathbf{x} \in \Omega$ or, for short, $T^*\omega = \sum_{1 \leq i_1 < \cdots < i_k \leq n} (f_{i_1, \dots, i_k} \circ T) dT_{i_1} \wedge \cdots \wedge dT_{i_k}$.

Examples

- ① In the case $k = m = n$ we have $\omega = f \, dx_1 \wedge \cdots \wedge dx_n$ with $f: D \rightarrow \mathbb{R}$ and

$$T^* \omega = (f \circ T) \det \mathbf{J}_T \, dx_1 \wedge \cdots \wedge dx_n.$$

This follows from

$$(dT_1(\mathbf{x}) \wedge \cdots \wedge dT_n(\mathbf{x}))(\mathbf{e}_1, \dots, \mathbf{e}_n) = \begin{vmatrix} dT_1(\mathbf{x})(\mathbf{e}_1) & \dots & dT_1(\mathbf{x})(\mathbf{e}_n) \\ \vdots & & \vdots \\ dT_n(\mathbf{x})(\mathbf{e}_1) & \dots & dT_n(\mathbf{x})(\mathbf{e}_n) \end{vmatrix} = \det(dT(\mathbf{x})(\mathbf{e}_1) | \dots | dT(\mathbf{x})(\mathbf{e}_n)) = \det \mathbf{J}_T(\mathbf{x}).$$

- ② Suppose $\omega = f_1 dx_1 + \cdots + f_n dx_n$ is a differential 1-form on $D \subseteq \mathbb{R}^n$ and $\gamma: (a, b) \rightarrow D$ is a differentiable curve in D .

$$\begin{aligned} \implies (\gamma^* \omega)(t) &= f_1(\gamma(t)) d\gamma_1(t) + \cdots + f_n(\gamma(t)) d\gamma_n(t) \\ &= f_1(\gamma(t)) \gamma'_1(t) dt + \cdots + f_n(\gamma(t)) \gamma'_n(t) dt \\ &= [f_1(\gamma(t)) \gamma'_1(t) + \cdots + f_n(\gamma(t)) \gamma'_n(t)] dt, \end{aligned}$$

the integrand used to compute the line integral $\int_{\gamma} \omega$.

Derivative of Differential Forms

Definition

A differential k -form $\omega: D \rightarrow \text{Alt}^k(\mathbb{R}^n)$, $D \subseteq \mathbb{R}^n$, is said to be *differentiable* if all its coordinate functions (in the standard representation) are differentiable. If applicable, the *derivative* of ω is defined as the differential $(k+1)$ -form

$$\begin{aligned} d\omega &= d \left(\sum_{1 \leq i_1 < \dots < i_k \leq n} f_{i_1, \dots, i_k} dx_{i_1} \wedge \dots \wedge dx_{i_k} \right) \\ &= \sum_{1 \leq i_1 < \dots < i_k \leq n} df_{i_1, \dots, i_k} \wedge dx_{i_1} \wedge \dots \wedge dx_{i_k}, \end{aligned}$$

with the convention that df is the usual differential for differential 0-forms (functions) $f: D \rightarrow \mathbb{R}$.

According to this definition we have $d\omega = 0$ if ω is constant, and for differential 0-forms $f: D \rightarrow \mathbb{R}$ the product rule

$$d(f \wedge \omega) = d(f\omega) = df \wedge \omega + f d\omega = df \wedge \omega + f \wedge d\omega.$$

Together with the linearity requirement these properties

determine the map $\omega \mapsto d\omega$ uniquely, since then necessarily

$$\begin{aligned} d(f dx_{i_1} \wedge \cdots \wedge dx_{i_k}) &= df \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k} + f d(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) \\ &= df \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}. \end{aligned}$$

The product rule for general differential forms is similar, except that the 2nd summand has an additional minus sign if the first factor is a differential k -form with k odd: For $D \subseteq \mathbb{R}^n$, $\omega: D \rightarrow \text{Alt}^k(\mathbb{R}^n)$, $\eta: D \rightarrow \text{Alt}^l(\mathbb{R}^n)$ we have

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.$$

Examples

- 1 For $n = 2$ there is only one interesting case, the derivative of a differential 1-form $\omega = P dx + Q dy$:

$$\begin{aligned} d\omega &= dP \wedge dx + dQ \wedge dy \\ &= (P_x dx + P_y dy) \wedge dx + (Q_x dx + Q_y dy) \wedge dy \\ &= P_x dx \wedge dx + P_y dy \wedge dx + Q_x dx \wedge dy + Q_y dy \wedge dy \\ &= (Q_x - P_y) dx \wedge dy. \end{aligned}$$

Examples (cont'd)

- ② For $n = 3$ there are two interesting cases, the derivatives of a differential 1-form $\omega = f \, dx + g \, dy + h \, dz$ and a differential 2-form $\eta = f \, dy \wedge dz + g \, dz \wedge dx + h \, dx \wedge dy$:

$$\begin{aligned} d\omega &= df \wedge dx + dg \wedge dy + dh \wedge dz \\ &= (f_x \, dx + f_y \, dy + f_z \, dz) \wedge dx + (g_x \, dx + g_y \, dy + g_z \, dz) \wedge dy + (h_x \, dx + h_y \, dy + h_z \, dz) \wedge dz \\ &= (g_x - f_y) \, dx \wedge dy + (h_y - g_z) \, dy \wedge dz + (f_z - h_x) \, dz \wedge dx, \end{aligned}$$

$$\begin{aligned} d\eta &= df \wedge dy \wedge dz + dg \wedge dz \wedge dx + dh \wedge dx \wedge dy \\ &= f_x \, dx \wedge dy \wedge dz + g_y \, dy \wedge dz \wedge dx + h_z \, dz \wedge dx \wedge dy \\ &= (f_x + g_y + h_z) \, dx \wedge dy \wedge dz = \text{div}(f, g, h) \, dx \wedge dy \wedge dz, \end{aligned}$$

the latter since the 3-cycles $(1, 2, 3)$ and $(1, 3, 2)$ are needed to restore the standard order in $dy \wedge dz \wedge dx$ resp. $dz \wedge dx \wedge dy$, and these have sign $+1$.

Using the notation $\omega = f_1 \, dx + f_2 \, dy + f_3 \, dz$, $\partial_1 f_1 = (f_1)_x = g_x$, etc., the first formula can also be written as

$d\omega = (\partial_1 f_2 - \partial_2 f_1) \, dx \wedge dy + (\partial_2 f_3 - \partial_3 f_2) \, dy \wedge dz + (\partial_3 f_1 - \partial_1 f_3) \, dz \wedge dx$, so that the vector field associated with $d\omega$ is $\text{curl}(f, g, h)$.

Clairaut's Theorem for Differential Forms

For a C^2 -function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^n$, we have

$$\begin{aligned} d(df) &= d\left(\sum_{i=1}^n f_{x_i} dx_i\right) = \sum_{i=1}^n df_{x_i} \wedge dx_i \\ &= \sum_{i=1}^n \sum_{j=1}^n f_{x_i x_j} dx_j \wedge dx_i = \sum_{1 \leq i < j \leq n} (f_{x_j x_i} - f_{x_i x_j}) dx_i \wedge dx_j \\ &= 0. \end{aligned} \quad (\text{by Clairaut's Theorem})$$

This generalizes to arbitrary differential forms:

Theorem

For every differential form ω of class C^2 (i.e., the coordinate functions in the standard representation of ω are C^2 -functions) we have $d(d\omega) = 0$.

Proof.

The proof follows readily from Clairaut's Theorem: It suffices to consider $\omega = f dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, in which case

$$d(d\omega) = d(df \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}) = d^2 f \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k} = 0. \quad \square$$

Definition

A differential k -form ($k \geq 1$) ω on $D \subseteq \mathbb{R}^n$ is said to be *closed* if $d\omega = 0$, and *exact* if there is a differential $(k - 1)$ -form η such that $\omega = d\eta$.

This definition of exactness generalizes the earlier one for differential 1-forms. A differential 1-form $\omega = \sum_{i=1}^n f_i dx_i$ is closed iff it satisfies the conditions in Poincaré's Lemma (i.e., iff it is locally exact). This follows from

$$\begin{aligned} d\omega &= \sum_{i=1}^n df_i \wedge dx_i = \sum_{i=1}^n \left(\sum_{j=1}^n \partial_j f_i dx_j \right) \wedge dx_i \\ &= \sum_{1 \leq i < j \leq n} (\partial_i f_j - \partial_j f_i) dx_i \wedge dx_j. \end{aligned}$$

“Exact” implies “closed” because of Clairaut’s Theorem:
 $\omega = d\eta \implies d\omega = d(d\eta) = 0$.

For $n = 3$ there are two nontrivial instances of Clairaut’s Theorem ($k = 0, 1$), which translated into the language of vector fields say:

- 1 For any C^2 -function $f: D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}^3$, we have $\text{curl}(\nabla f) = 0$.
- 2 For any C^2 -vector-field $\mathbf{F} = (f, g, h): D \rightarrow \mathbb{R}^3$, $D \subseteq \mathbb{R}^3$, we have $\text{div}(\text{curl } \mathbf{G}) = 0$.

Poincaré's Lemma for differential 1-forms on star-shaped regions and its generalization to simply-connected regions generalize to differential k -forms with $k > 1$:

Theorem

Suppose $D \subseteq \mathbb{R}^n$ is simply-connected and ω is a continuously differentiable differential k -form on D with $k \geq 1$. If ω is closed then ω is exact.

For $n = 3$ there are three nontrivial instances of this theorem ($k = 1, 2, 3$), which translated into the language of vector fields say: If $D \subseteq \mathbb{R}^3$ is simply-connected then

- ① every vector field F on D of class C^1 satisfying $\operatorname{curl} F = 0$ (i.e., F is *irrotational*) is a gradient field;
- ② every vector field F on D of class C^1 satisfying $\operatorname{div} F = 0$ (i.e., F is *incompressible*) is the curl of some vector field;
- ③ every C^1 -function $f: D \rightarrow \mathbb{R}$ is the divergence of some vector field.

In (3) it suffices to assume that f is continuous; cf. [Ste21], Ch. 16.6, Ex. 41.

Exercise

Suppose $D \subseteq \mathbb{R}^3$ is simply-connected. Show directly that every C^1 -function $f: D \rightarrow \mathbb{R}$ is the divergence of some vector field $G: D \rightarrow \mathbb{R}^3$.

Hint: G can be chosen of the special form $G = (g, 0, 0)$. Show first that locally at $(x_0, y_0, z_0) \in D$ one can take $g(x, y, z) = \int_{x_0}^x f(t, y, z) dt$, and then use “prolongation” along paths in D to define G globally.

Exercise

Let $F = (f_1, f_2, f_3)$, $G = (g_1, g_2, g_3)$ be differentiable vector fields on $D \subseteq \mathbb{R}^3$. Show that the curl of $F \times G: D \rightarrow \mathbb{R}^3$, $\mathbf{x} \mapsto F(\mathbf{x}) \times G(\mathbf{x})$ satisfies

$$\nabla \times (F \times G) = (\nabla \cdot G)F + (G \cdot \nabla)F - (\nabla \cdot F)G - (F \cdot \nabla)G.$$

Here $\nabla \cdot G = \text{div}(G)$ and $G \cdot \nabla = g_1\partial_1 + g_2\partial_2 + g_3\partial_3$ is the operator sending f to $g_1(\partial_1 f) + g_2(\partial_2 f) + g_3(\partial_3 f)$.

Exercise

Show that any vector field $F: D \rightarrow \mathbb{R}^3$, $D \subseteq \mathbb{R}^3$, of class C^2 satisfies the identity

$$\nabla \times (\nabla \times F) = \operatorname{curl}(\operatorname{curl} F) = \nabla(\operatorname{div} F) - \Delta F,$$

where $\Delta = \partial_1^2 + \partial_2^2 + \partial_3^2$ acts coordinate-wise on F .

Exercise

Suppose $D \subseteq \mathbb{R}^3$ is star-shaped with center $\mathbf{0}$, $F: D \rightarrow \mathbb{R}^3$ is a vector field of class C^1 with $\operatorname{div} F = 0$, and $G: D \rightarrow \mathbb{R}^3$ is defined by

$$G(\mathbf{x}) = \int_0^1 F(t\mathbf{x}) \times t\mathbf{x} dt.$$

Show that

$$\operatorname{curl}(F(t\mathbf{x}) \times t\mathbf{x}) = \frac{d}{dt}(t^2 F(t\mathbf{x})) \quad \text{for } \mathbf{x} \in D, t \in [0, 1]$$

(where curl applies only to the variables in \mathbf{x}), and conclude that $\operatorname{curl} G(\mathbf{x}) = F(\mathbf{x})$ for $\mathbf{x} \in D$.

Orientation of Vector Spaces

In finite-dimensional real vector spaces it makes sense to speak of the orientation of an ordered basis.

Definition

- ① An ordered basis $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of the standard vector space \mathbb{R}^n is said to be *positively oriented (negatively oriented)* if the matrix $\mathbf{B} = (\mathbf{b}_1 | \dots | \mathbf{b}_n)$ satisfies $\det(\mathbf{B}) > 0$ (resp., $\det(\mathbf{B}) < 0$). In particular the standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ of \mathbb{R}^n (ordered in the usual way) is positively oriented.
- ② Two ordered bases $B = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$, $B' = \{\mathbf{b}'_1, \dots, \mathbf{b}'_n\}$, of an arbitrary n -dimensional vector space V over \mathbb{R} are said to be *equally oriented (oppositely oriented)* if the change-of-basis matrix \mathbf{A} from B to B' (i.e., $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n \times n}$) is defined by $\mathbf{b}_j = \sum_{i=1}^n a_{ij} \mathbf{b}'_i$ for $1 \leq j \leq n$) satisfies $\det(\mathbf{A}) > 0$ (resp., $\det(\mathbf{A}) < 0$).
- ③ A vector space V as in (2) is said to be *oriented* if one particular ordered basis B_0 of V has been named “positively oriented”. (By convention, for \mathbb{R}^n this is the standard basis.) All ordered bases of V with the same (opposite) orientation as B_0 are then called positively (resp., negatively) oriented.

Notes

- The relation “equally oriented” defined in (2) is an equivalence relation with exactly two equivalence classes $[B_1]$, $[B_2]$. A vector space V as in (2) admits precisely two distinct orientations—one in which B_1 is positively oriented and another one in which B_1 is negatively oriented.
- Linearly independent vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ are positively oriented iff the path $\mathbf{0} \rightarrow \mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{0}$ traverses the boundary of the triangle with vertices $\mathbf{0}, \mathbf{a}, \mathbf{b}$ (or $\mathbf{v}, \mathbf{v} + \mathbf{a}, \mathbf{v} + \mathbf{b}$ with $\mathbf{v} \in \mathbb{R}^2$ arbitrary) counterclock-wise. Linearly independent vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ are positively oriented iff the path $\mathbf{0} \rightarrow \mathbf{a} \rightarrow \mathbf{b} \rightarrow \mathbf{0}$ traverses the boundary of the triangle with vertices $\mathbf{0}, \mathbf{a}, \mathbf{b}$ counterclock-wise when “watched” from the half space containing the point \mathbf{c} .
- Well-known properties of the determinant function imply that the orientation of an ordered basis changes if two basis vectors are interchanged or a basis vector is multiplied by a negative real number. A cyclic permutation of the vectors in an ordered basis of \mathbb{R}^3 preserves the orientation; more generally, this is true for \mathbb{R}^n with n odd (but false for even n).

Notes cont'd

- Suppose V, W are oriented vector spaces over \mathbb{R} and $f: V \rightarrow W$ is a bijective linear map (so in particular $\dim V = \dim W$). The map f is said to be *orientation-preserving* (*orientation-reversing*) if it maps one positively oriented basis of V to a positively oriented basis of W (resp., to a negatively oriented basis of W). In the special case $V = W = \mathbb{R}^n$ a linear map $\mathbf{x} \mapsto \mathbf{Ax}$, $\mathbf{A} \in \mathbb{R}^{n \times n}$, is orientation-preserving iff $\det \mathbf{A} > 0$.

Exercise

Let V be a finite-dimensional vector space over \mathbb{R} . Using further well-known properties of the determinant function, prove in detail that “equal orientation” defines an equivalence relation on V , and that the number of its equivalence classes is two.

Exercise

Show that an orientation-preserving linear map as defined above maps any positively (negatively) oriented basis of V to a positively (resp., negatively) oriented basis of W , and an orientation-reversing linear map does the same with the roles of positively/negatively oriented bases of W interchanged.

Orientation of Manifolds

Now suppose that M is a differentiable k -dimensional manifold in \mathbb{R}^n . Recall that this implies that M has a k -dimensional linear tangent space $T_{\mathbf{a}}(M)$ in every point $\mathbf{a} \in M$, described as follows:

- ① $T_{\mathbf{a}}(M)$ consists of all tangent vectors $\gamma'(0)$ of differentiable curves $\gamma: (-\epsilon, \epsilon) \rightarrow M$ with $\gamma(0) = \mathbf{a}$.
- ② If $\gamma: \Omega \rightarrow M \cap U$, $\Omega \subseteq \mathbb{R}^k$, is a bijective immersion parametrizing a chart region of M containing \mathbf{a} (i.e., γ inverts a chart $\kappa: U \rightarrow V$ with $\kappa(M \cap U) = \Omega \times \{\mathbf{0}\} \subset \mathbb{R}^k \times \mathbb{R}^{n-k}$) then $T_{\mathbf{a}}(M)$ is equal to the column space of $\mathbf{J}_{\gamma}(\omega_0)$, $\mathbf{a} = \gamma(\omega_0)$.
- ③ $T_{\mathbf{a}}(M)$ is the direction space of the affine tangent space of M in \mathbf{a} that we have considered earlier.

Also recall that if $\gamma_i: \Omega_i \rightarrow M \cap U$, $i = 1, 2$ have the property in (2) then there exists a diffeomorphism $T: \Omega_1 \rightarrow \Omega_2$ such that $\gamma_1 = \gamma_2 \circ T$.
 \Rightarrow If $\gamma_i: \Omega_i \rightarrow M \cap U_i$, $i = 1, 2$, parametrize overlapping chart regions (i.e., $M \cap U_1 \cap U_2 \neq \emptyset$) and $\Omega'_i = \gamma_i^{-1}(M \cap U_1 \cap U_2)$ then their restrictions $\gamma'_i = \gamma_i|_{\Omega'_i}$ parametrize the same chart region $M \cap U_1 \cap U_2$ and hence are switched by a diffeomorphism $T: \Omega'_1 \rightarrow \Omega'_2$ in the above sense (i.e., $\gamma'_1 = \gamma'_2 \circ T$).

Definition

- 1 Two parametrizations $\gamma_i: \Omega_i \rightarrow M \cap U_i$, $i = 1, 2$ are said to be equally oriented if their chart regions don't overlap or the diffeomorphism T switching γ'_1 and γ'_2 (cf. previous slide) has positive Jacobi determinant.
- 2 M is said to be *orientable* if there exists a family Γ of equally oriented parametrizations $\gamma: \Omega \rightarrow M \cap U$ whose chart regions $\gamma(\Omega) = M \cap U$ cover M (i.e., $M \subseteq \bigcup_{\gamma \in \Gamma} U_\gamma$, where U_γ denotes the domain of the chart corresponding to γ).

Notes

- W.l.o.g. we may assume that parameter domains (and chart regions) are connected. Then, according to the Intermediate Value Theorem, $\det \mathbf{J}_T(\omega)$ must have the same sign for all $\omega \in \Omega'_1$ (it can't be zero, since $\mathbf{J}_T(\omega)$ is invertible).
- The condition $\det \mathbf{J}_T > 0$ may be rephrased as follows: $T_{\mathbf{a}}(M)$, $\mathbf{a} \in M$, can be oriented in a consistent fashion

Notes cont'd

- Not every manifold is orientable; cf. the Moebius strip, obtained by glueing two opposite sides of a rectangle in the reverse order.
- If M is orientable, it has exactly two different orientations
- A hypersurface (*i.e.*, an $(n - 1)$ -dimensional manifold) in \mathbb{R}^n is orientable iff there exists a continuous map $\mathbf{n}: M \rightarrow \mathbb{R}^n$ with $|\mathbf{n}(\mathbf{x})| = 1$ for all $\mathbf{x} \in M$ and such that $\mathbf{n}(\mathbf{x})$ is orthogonal to the tangent hyperplane of M in \mathbf{x} .

Unlike functions, differential k -forms on $D \subseteq \mathbb{R}^n$ with $k \geq 1$ can be integrated only over oriented k -dimensional surfaces contained in D . First we consider the case $k = n$.

Definition

Suppose $\omega = f dx_1 \wedge \cdots \wedge dx_n$ is a differential n -form on $\Omega \subseteq \mathbb{R}^n$. The form ω is said to be *integrable over Ω* if f is Lebesgue-integrable over Ω , and if applicable we define

$$\int_{\Omega} \omega = \int_{\Omega} f dx_1 \wedge \cdots \wedge dx_n := \int_{\Omega} f(\mathbf{x}) d^n \mathbf{x}.$$

In particular $dx_1 \wedge \cdots \wedge dx_n$ is integrable over Ω iff Ω is measurable, in which case $\int_{\Omega} dx_1 \wedge \cdots \wedge dx_n = \text{vol}(\Omega)$. Accordingly $dx_1 \wedge \cdots \wedge dx_n$ is also known as the *volume form* on \mathbb{R}^n .

Observation

Suppose $\Omega_1, \Omega_2 \subseteq \mathbb{R}^n$ are connected open sets and $T: \Omega_1 \rightarrow \Omega_2$ is a diffeomorphism, and $\omega = f dx_1 \wedge \cdots \wedge dx_n$ is a differential n -form on Ω_2 . Then $\det \mathbf{J}_T(\mathbf{x})$ has the same sign for all $\mathbf{x} \in \Omega_1$, and we have

$$\int_{\Omega_1} T^* \omega = \begin{cases} \int_{\Omega_2} \omega & \text{if } \det \mathbf{J}_T > 0, \\ -\int_{\Omega_2} \omega & \text{if } \det \mathbf{J}_T < 0. \end{cases}$$

Proof.

Since T is a diffeomorphism, we have $\det \mathbf{J}_T(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in \Omega_1$. If $\mathbf{x} \rightarrow \det \mathbf{J}_T(\mathbf{x})$ would attain both positive and negative values then by the Intermediate Value Theorem it would attain the value 0 as well; contradiction. (The function $\mathbf{x} \rightarrow \det \mathbf{J}_T(\mathbf{x})$ is continuous, because T is C^1 .)

From an earlier example we know that

$T^* \omega = (f \circ T) \det \mathbf{J}_T dx_1 \wedge \cdots \wedge dx_n$, and hence

$$\begin{aligned} \int_{\Omega_1} T^* \omega &= \int_{\Omega_1} f(T(\mathbf{x})) \det \mathbf{J}_T(\mathbf{x}) d^n \mathbf{x} && (\text{definition of } \int_{\Omega_1} T^* \omega) \\ &= \pm \int_{\Omega_2} f(\mathbf{y}) d^n \mathbf{y} = \pm \int_{\Omega_2} \omega, && (\text{change of variables}) \end{aligned}$$

Our goal is to define the integral of a differential k -form ω on $D \subseteq \mathbb{R}^n$ over a k -dimensional surface $S \subset D$ (more precisely, a chart region $S = M \cap U$ of a k -dimensional differentiable manifold $M \subset D$) by pulling it back along a parametrization (more precisely, a bijective immersion) $\gamma: \Omega \rightarrow S$ to a differential k -form $\gamma^*\omega$ on $\Omega \subseteq \mathbb{R}^k$ and setting $\int_S \omega = \int_{\Omega} \gamma^*\omega$, which is an instance of the already done case $k = n$.

Since $\int_{\Omega} \gamma^*\omega$ should be independent of the chosen parametrization (recall that any two such parametrizations $\gamma_i: \Omega_i \rightarrow S$, $i = 1, 2$, are switched by a diffeomorphism $T: \Omega_1 \rightarrow \Omega_2$ in the sense that $\gamma_1 = \gamma_2 \circ T$), in view of the preceding observation we must restrict consideration to the case $\det \mathbf{J}_T > 0$.

Definition (Integration of a differential k -form on $D \subseteq \mathbb{R}^n$ over a k -dimensional oriented manifold $M \subset D$)

Suppose M is oriented and Γ is a family of equally oriented parametrizations (bijective immersions) representing the given orientation and whose images (chart regions) cover M .

- ① If $M_0 \subseteq M$ is a chart region and $\gamma: \Omega \rightarrow M_0$ is a parametrization with the same orientation as the members of Γ , we define

$$\int_{M_0} \omega := \int_{\Omega} \gamma^* \omega = \int_{\Omega} f dx_1 \wedge \cdots \wedge dx_k,$$

where $\gamma^* \omega = f dx_1 \wedge \cdots \wedge dx_k$ denotes the pullback of ω along γ (which is a differential k -form on $\Omega \subseteq \mathbb{R}^k$).

- ② In the general case $\int_M \omega$ is defined in the same way as for surface integrals using a partition of unity on M .

Physical Interpretation

Suppose $D \subseteq \mathbb{R}^3$ is open, $\omega: D \rightarrow \text{Alt}^k(\mathbb{R}^3)$ is a differential k -form, and M is k -dimensional oriented manifold in D , where $k \in \{1, 2, 3\}$.

$k = 1$ A 1-dimensional oriented manifold in D is essentially the same as a smooth non-parametric curve (which can always be oriented), and $\int_M \omega$ is the line integral of ω along M , which for chart regions is computed using the order of traversal determined by the orientation.

$k = 2$ Here ω has the form $\omega = f \, dy \wedge dz + g \, dz \wedge dx + h \, dx \wedge dy$, M is a 2-dimensional oriented surface in D , whose orientation is determined by a continuous unit normal $\mathbf{n}: M \rightarrow \mathbb{R}^3$, and

$$\int_M \omega = \int_M \mathbf{F}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) \, dS(\mathbf{x}).$$

can be interpreted as the flow with velocity field \mathbf{F} through M .

$k = 3$ Here M is just an open subset of D and $\int_M \omega$ is the ordinary Lebesgue integral.

The General Stokes Theorem

Also known as “Fundamental Theorem of Calculus”

Suppose $M \subseteq \mathbb{R}^n$ is an oriented k -dimensional differentiable manifold, $G \subseteq M$ is a compact subset with smooth boundary, and ∂G carries the orientation induced by M . Then for every continuously differentiable differential $(k - 1)$ -form ω on M we have

$$\int_G d\omega = \int_{\partial G} \omega.$$

Note

As in the definition of surface integrals one may relax the condition on the smoothness of ∂G to the case that the singular boundary of G has $(k - 1)$ -dimensional Lebesgue measure zero.

The Fundamental Theorem for Line Integrals

A special case of the general Stokes Theorem

Suppose $\gamma: [a, b] \rightarrow \mathbb{R}^n$ is a smooth curve parametrizing a 1-dimensional compact differentiable manifold M .

$\implies M$ is oriented (by the unit tangent vector $\mathbf{T} = \gamma / |\gamma|$), and the induced orientation on $\partial M = \{\gamma(a), \gamma(b)\}$ is given by $+1$ at $\gamma(b)$ and -1 at $\gamma(a)$.

If $\omega = f$ is a continuously differentiable differential 0-form (i.e., a function), Stokes' Theorem gives

$$\int_{\gamma} df = \int_M df = \int_{\partial M} f = f(\gamma(b)) - f(\gamma(a)).$$

Thus the special case $k = 1$ of Stokes' Theorem gives the Fundamental Theorem for Line Integrals.

Green's Theorem

A special case of the general Stokes Theorem

Green's Theorem (also called *2-dimensional divergence theorem*) is the case $n = k = 2$ of Stokes' Theorem. It says that for a bounded open subset D of \mathbb{R}^2 with smooth boundary ∂D and any continuously differentiable differential 1-form $\omega = P(x, y) dx + Q(x, y) dy$ defined on $\overline{D} = D \cup \partial D$ we have

$$\int_D (Q_x - P_y) d^2(x, y) = \int_D d\omega = \int_{\partial D} \omega = \int_{\gamma} P dx + Q dy,$$

where γ is a positively oriented parametrization of ∂D . This means that when moving along γ the region D is always on the left and $\mathbb{R}^2 \setminus \overline{D}$ is always on the right.

If ∂D consists of several smooth pieces, each piece has to be parametrized in this way, say by γ_i for $1 \leq i \leq r$, and $\int_{\partial D} P dx + Q dy := \sum_{i=1}^r \int_{\gamma_i} P dx + Q dy$.

Direct proof of Green's Theorem.

(1) First assume that $D = (a, b) \times (c, d)$ is a rectangle. In this case the boundary consists of the 4 vertices, which can be discarded, and 4 smooth pieces (the edges of the rectangle). Positive parametrizations of the edges are

$$\gamma_1(s) = (s, c), \quad s \in (a, b),$$

$$\gamma_2(t) = (b, t), \quad t \in (c, d),$$

$$\gamma_3(s) = (a + b - s, d), \quad s \in (a, b),$$

$$\gamma_4(t) = (a, c + d - t), \quad t \in (c, d).$$

It suffices to prove $\int_D Q_x \, d^2(x, y) = \int_{\partial D} Q \, dy$ and $\int_D -P_y \, d^2(x, y) = \int_{\partial D} P \, dx$ separately.

$$\begin{aligned}\int_D Q_x \, d^2(x, y) &= \int_a^b \int_c^d Q_x(s, t) \, dt \, ds = \int_c^d \int_a^b Q_x(s, t) \, ds \, dt \\ &= \int_c^d Q(b, t) - Q(a, t) \, dt = \int_{\gamma_2} Q \, dy + \int_{\gamma_4} Q \, dy.\end{aligned}$$

Moreover, $\int_{\gamma_1} Q \, dy = \int_{\gamma_3} Q \, dy = 0$, showing the first formula. The second formula is proved in the same way. □

Direct proof of Green's Theorem cont'd.

(2) Since D is open, we can write $D = \bigcup_{i=1}^{\infty} Q_i$ as a countable union of 2-dimensional intervals (rectangles) with mutually disjoint interiors (i.e., with only boundary points in common).

$$\implies \int_D Q_x - P_y \, d^2(x, y) = \sum_{i=1}^{\infty} \int_{Q_i} Q_x - P_y \, d^2(x, y) = \sum_{i=1}^{\infty} \int_{\partial Q_i} P \, dx + Q \, dy.$$

On the other hand we have, writing $D_N = \bigcup_{i=1}^N Q_i$,

$$\sum_{i=1}^N \int_{\partial Q_i} P \, dx + Q \, dy = \int_{\partial D_N} P \, dx + Q \, dy,$$

since line segments making up the edges of Q_1, \dots, Q_N that are not on the boundary of D_N belong to exactly two rectangles Q_i, Q_j and carry opposite orientations with respect to these.

\implies Letting $N \rightarrow \infty$ finishes the proof, since it is easy to see that

$$\lim_{N \rightarrow \infty} \int_{\partial D_N} P \, dx + Q \, dy = \int_{\partial D} P \, dx + Q \, dy. \quad \square$$

A Different View

Replacing Q by f and P by $-g$ in Green's Theorem shows that it is equivalent to

$$\int_D f_x + g_y \, d^2(x, y) = \int_{\partial D} -g \, dx + f \, dy = \int_{\partial D} \mathbf{F} \cdot \mathbf{n} \, ds,$$

where $\mathbf{F} = (f, g)$ and \mathbf{n} is the unit normal vector of ∂D pointing outwards.

$(f, g) \mapsto f_x + g_y$ is a 2-dimensional analogue of the divergence of a 3-dimensional vector field, and $\int_{\partial D} \mathbf{F} \cdot \mathbf{n} \, ds$ is the “2-dimensional flow” of F through ∂D in the given orientation (i.e., out of D).

This analogy with Gauss' Divergence Theorem motivates the name “2-dimensional divergence theorem” for Green's Theorem.

Application

If we set $\mathbf{F}(x, y) = \frac{1}{2}(x, y)$, i.e., $f(x, y) = x/2$, $g(x, y) = y/2$, we have $\operatorname{div} \mathbf{F}(x, y) = 1$, so that $A = \int_D f \, dx + g \, dy$ is just the area of D .
 \implies Green's theorem gives the following formula for the area of D :

$$A = \frac{1}{2} \int_{\partial D} x \, dy - y \, dx. \quad (\text{Leibniz's Sector Formula})$$

Application cont'd

The formula is called “Sector Formula”, since for not necessarily closed plane curves $\gamma: [a, b] \rightarrow \mathbb{R}^2$ it gives the oriented area swept out by the radius vector of the curve. This can be seen using the parametrized form

$$A = \frac{1}{2} \int_a^b x(t)y'(t) - y(t)x'(t) dt = \int_a^b \frac{1}{2} \begin{vmatrix} x(t) & x'(t) \\ y(t) & y'(t) \end{vmatrix} dt$$

and recalling that the integrand gives the oriented area of the triangle with vertices $\mathbf{0}, \gamma(t), \gamma(t) + \gamma'(t)$.

If γ is a simple closed curve (i.e., looks like a deformed circle), parts of the sector exterior to γ are swept out twice with opposite directions—draw a picture with $(0, 0)$ exterior to γ and check this property!—and hence don't contribute to the integral.

It should be noted that there are other ways to compute the area enclosed by a simple closed curve, since there are many vector fields $\mathbf{F} = (f, g)$ with $\operatorname{div} \mathbf{F} \equiv 1$. For example we can take

$\mathbf{F} = (x, 0)$ or $(0, y)$ in Green's Theorem to obtain

$$A = \frac{1}{2} \int_{\partial D} x dy - y dx = \int_{\partial D} x dy = - \int_{\partial D} y dx.$$

Exercise

Show directly from the definition that for closed curves γ the line integral $\int_{\gamma} x \, dy - y \, dx$ is translation-invariant, i.e., if $\gamma: [a, b] \rightarrow \mathbb{R}^2$ is closed and $\mathbf{v} \in \mathbb{R}^2$ then the line integrals of $x \, dy - y \, dx$ along γ and $\gamma + \mathbf{v}$ are the same.

Gauss's Divergence Theorem

A special case of the general Stokes Theorem

Gauss's Divergence Theorem is the case $n = k = 3$ of Stokes' Theorem. It says that for a bounded open subset D of \mathbb{R}^3 with smooth boundary surface ∂D and any differential 2-form $\omega = f \, dy \wedge dz + g \, dz \wedge dx + h \, dx \wedge dy$ of class C^1 on the (compact) solid $\overline{D} = D \cup \partial D$ we have

$$\int_D \operatorname{div} \mathbf{F} d^3(x, y, z) = \int_D \omega = \int_{\partial D} d\omega = \int_{\partial D} \mathbf{F} \cdot \mathbf{n} dS,$$

where $\mathbf{n}: \partial D \rightarrow \mathbb{R}^3$ denotes the *outer unit normal* of D .

Application

Following the reasoning after Green's Theorem, it is clear that we can use the Divergence Theorem with a suitably chosen vector field to express the volume of a solid $B \subset \mathbb{R}^3$ as a surface integral over the boundary ∂S . We use this idea to derive the relation $\sigma_n = n \beta_n$ between the unit ball's volume and surface area in a less painful way.

Consider the vector field $\mathbf{F}(x, y, z) = (x, y, z)$. Then $\operatorname{div} \mathbf{F} \equiv 3$, and \mathbf{F} provides the outer unit normal on the surface S^2 of the unit ball $B_1(\mathbf{0})$. (For a general sphere of radius r centered at the origin the outer unit normal would be $\mathbf{F}/|\mathbf{F}|$.)

Hence the Divergence Theorem gives

$$3\beta_3 = \int_B \operatorname{div} \mathbf{F} d^3(x, y, z) = \int_{S^2} \mathbf{F} \cdot \mathbf{F} dS = \int_{S^2} 1 dS = \sigma_3.$$

The relation $\sigma_n = n \beta_n$ follows in the same way by applying the Divergence Theorem in dimension n (which has mutatis mutandis the same form) to the vector field $\mathbf{F}(\mathbf{x}) = \mathbf{x}$, $\mathbf{x} = (x_1, \dots, x_n)$.