Ch01. Point-Set Topology and Calculus

Ping Yu

Faculty of Business and Economics
The University of Hong Kong

1. Sets and Set Operations

2. Functions

3. Point-Set Topology in the Euclidean Space
   - Euclidean Spaces
   - Open Sets
   - Compact Sets

4. Single Variable Calculus
   - Limits
   - Continuity
   - Differentiability
   - Higher-order Derivatives
   - Integrability

5. Multivariable Calculus

General Information on the Math Camp

- Instructor: Ping Yu
- Email: pingyu@hku.hk
- Time: 06:45-8:00pm and 8:15-9:30pm, Tuesday
- Location: ATC-B4 (Sept 4, 11&18)/KK315 (Oct 2)/ATC-B12 (Oct 9&16)
- Office Hour: 11:00-12:00pm, Tuesday, KK1108
  - I will not answer questions in email if the answer is long or is not easy to explain exactly by words. Please stop by during my office hour.

- Tutor: TBA
- Email: TBA
- Time/Location: TBA
- Office Hour: TBA

Information on the Content and Evaluation of Math Camp

- Textbook: My lecture notes (LNs) posted on Moodle.
  - Others: Rudin (1976) for Chapter 1, Simon and Blume (1994) and Sundaram (1996) for Chapter 2-4, and Casella and Berger (2002) for Chapter 5-6.
- Exercises in LNs: no need to turn in, and for practice only, so no answer key will be posted.
- Evaluation: One Assignment (40%) and One Exam (60%)/only materials in slides
  - Assignment: six problems/Chapter 1 contributes two and Chapter 2-5 each contributes one.
  - Exam: four problems and each of Chapter 2-5 contributes one/closed-book and closed-note and mimic the assignment.
- Time and Location of the Exam: TBA

Course Policy

- In Class: (i) turn off your cell phone and keep quiet; (ii) come to class and return from the break on time; (iii) you can ask me freely in class, but if your question is far out of the course or will take a long time to answer, I will answer you after class.
- Assignment: The assignment must be typed. Turn in your assignment online through moodle before the due day (5:30pm of October 26 - Friday). Late assignment is not acceptable for whatever reasons. To avoid any risk, start your assignment early.
- Tutorial: The answer key to the assignment would NOT be posted on moodle and will be taught by the tutor. Two tutorial classes would be provided before the exam.
  - Time and Location of the Tutorials: TBA

Overview This Course

- Chapter 1: Set, Function, Point-Set Topology, Single and Multivariable Calculus
  - too long, so we will only cover the topics that are necessary for future chapters.

- Chapter 2: Existence of Optimizer, Equality- and Inequality-Constrained Optimization/Necessary Conditions
- Chapter 3: Convex Set, Concave and Convex Function, Uniqueness of Optimizer, Sufficient Conditions
- Chapter 4: Maximum Theorem, Implicit Function Theorem, Envelope Theorem

- Chapter 5: Basics for Probability Theory
- Chapter 6: Basics for Statistics
  - This chapter would be detailed in Econ6001 and Econ6005, so no lecture note is posted, no exercises are given in the assignment, and it will not be tested. Just follow the slides!

- Order of Learning Process: slides in class $\longrightarrow$ the lecture notes $\longrightarrow$ the references
- Benefit Future Students: check typos of LNs and suggest topics to be taught in the future after finishing Econ6001 and Econ6021.

Sets and Set Operations

Sets

- Examples: the consumption set and production set will be used in the optimizing decisions of consumers and firms.
- A set A is a collection of distinct objects.
  - Elements in a set are **not ordered** and must be **distinct**, so the following three sets are the same: $\{1, 2\} = \{2, 1\} = \{1, 2, 1\}$.
- Notations: An element x in A is denoted as $x \in A$. An empty set is often denoted as $\emptyset$.
  - Sets are represented by uppercase italic, e.g., X, and their elements by lower case italic, e.g., x.
- Terms: In mathematics, "**collection**", "**class**" and "**family**" all mean "set". An "object" in a set is often called a "**point**" although it can be a function defined in the following section or any mathematical object.

Set Operations

1. Subset: set A is contained in B.
   - $A \subset B$: set A is contained in B, and $A \neq B$.
   - $A \subseteq B$: set A is contained in B, and A and B may be equal.

   - Quite often, $\subset$ means $\subseteq$. To emphasize $A \neq B$, $\subsetneqq$ is often used. We will use the convention of $\subseteq$ and $\subset$.
2. Union: $A \cup B = \{x | x \in A \text{ or } x \in B\}$. All points in either A or B.
   - "$|$" is read as "such that", and is often used exchangeably with ":" in this course.
3. Intersection: $A \cap B = \{x | x \in A \text{ and } x \in B\}$. All points in both A and B.
4. Complement: $A^c = \{x | x \notin A\}$. All points not in A. Here, a total set is implicitly defined.
5. Relative Complement: $B \setminus A = \{x \in B | x \notin A\} = B \cap A^c$: all points that are in B, but not in A. [Figure here]

- De Morgan's Law: $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$.

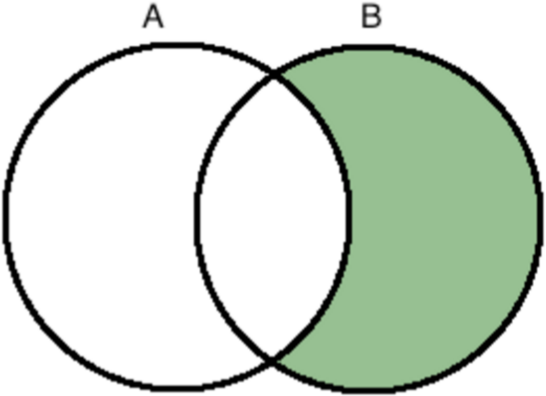


Figure: $B \setminus A = B \cap A^c$

Functions

Functions

- A function (or mapping) $f : X \longmapsto Y$ is a rule that associates each element of X with a unique element of Y; in other words, for each $x \in X$ there exists a specified element $y \in Y$, denoted as $f(x)$.
- x is called the argument of f, and $f(x)$ is called the value of f at x.
- X is called the domain of f, and Y the codomain.
- For $A \subseteq X$, the set

$$f(A) = \{f(x) | x \in A\} \subseteq Y$$

is called the image of A under f, and for $B \subseteq Y$, the set

$$f^{-1}(B) = \{x | f(x) \in B\} \subseteq X$$

is called the inverse image (or pre-image) of B under f. The set $f(X)$ is called the range of f.

Terms on Functions

- The term **function** is usually reserved for cases when the codomain is the set of **real** numbers. That is why we term utility functions and production functions.
- The term **correspondence** is used for a rule connecting elements of X to elements of Y where the latter are not necessarily unique. For example, f^{-1} is a correspondence, but not a function in general. If $f^{-1} : f(X) \to X$ is a function, then we call it the inverse function of f.
- Let $f : X \longmapsto Y$ and $g : Y \longmapsto Z$ are two mappings. The composite function (or mapping) $g \circ f : X \longmapsto Z$ takes each $x \in X$ to the element $g(f(x)) \in Z$.

Point-Set Topology in the Euclidean Space

History of Euclidean Spaces

Euclid (325-265, B.C.), Greek, "father of geometry"

Euclidean Spaces

- $\mathbb{R}^n$ is the Cartesian product of $\mathbb{R}$ with itself n times.
  - For two sets X and Y, the Cartesian product of X and Y is $X \times Y \equiv \{(x,y) \,|\, x \in X, y \in Y\}$, where "$\equiv$" is read as "defined as". [Figure here]
  - $\mathbb{R}^1$ is the real line; $\mathbb{R}^2$ is the plane; $\mathbb{R}^3$ is the three-dimensional space.
- The Euclidean space is $\mathbb{R}^n$ with the **Euclidean structure** imposed on.
  - We will still use $\mathbb{R}^n$ to denote the Euclidean space.
- The Euclidean structure is best described by the standard **inner product** on $\mathbb{R}^n$.
- The inner product (or dot product) of any two real n-vectors $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \text{ [a real number]},$$

  where x_i and y_i are ith coordinates of vectors $\mathbf{x}$ and $\mathbf{y}$ respectively, and $\mathbf{x} \cdot \mathbf{y}$ is often written as $\mathbf{x}'\mathbf{y}$ with $\mathbf{x}'$ meaning the transpose of $\mathbf{x}$ or $\langle \mathbf{x}, \mathbf{y} \rangle$.
- Notations: Real numbers (or scalars) are written using lower case italics, e.g., x. Vectors are defined as **column** vectors and represented using lowercase bold, e.g., $\mathbf{x}$.

History of Cartesian Product

René Descartes (1596-1650), French

- The Cartesian product is named after the French philosopher Descartes - from his Latinized name Cartesius.

Inner Product, Norm and Metric on $\mathbb{R}^n$

- Length of $\mathbf{x}$ - the Euclidean norm:

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}\cdot\mathbf{x}} = \sqrt{\sum_{i=1}^{n} x_i^2} \geq 0 \text{ [nonnegative]}.$$

- Metric (or distance) on $\mathbb{R}^n$:

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \text{ [nonnegative]}.$$

- Angle between $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$: inner product implies the **Cauchy-Schwarz inequality**:

$$|\langle \mathbf{x},\mathbf{y}\rangle| \leq \|\mathbf{x}\|\cdot\|\mathbf{y}\|,$$

  so we can define

$$\text{angle}(\mathbf{x},\mathbf{y}) = \arccos\frac{\langle \mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|\cdot\|\mathbf{y}\|},$$

  where the value of the angle is chosen to be in the interval $[0,\pi]$.
  - If $\langle \mathbf{x},\mathbf{y}\rangle = 0$, $\text{angle}(\mathbf{x},\mathbf{y}) = \frac{\pi}{2}$; we call $\mathbf{x}$ is orthogonal to $\mathbf{y}$ and denote it as $\mathbf{x}\perp\mathbf{y}$.
  [Figure here]

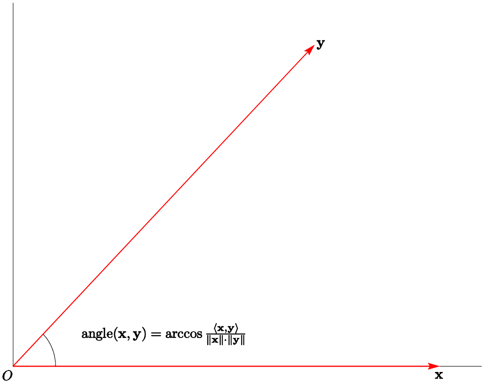


Figure: Angle in $\mathbb{R}^2$

Inner Product Space, Normed Space and Metric Space

- A space is a set plus some structure on it. So a set imposed a metric, a norm or an inner product is called a metric space, a normed space or an inner product space, respectively.
  - We will not study such general spaces in this course but only $\mathbb{R}^n$.
  - The relationship between these spaces is as follows

$\mathbb{R}^n$	$\subset$	inner product space	$\subset$	normed space	$\subset$	metric space
		$\langle x,y\rangle$ angle		$\|x\| = \sqrt{\langle x,x\rangle}$ length		$d(x,y) = \|x-y\|$ distance

Open Sets

- **Open sets** are the basic building blocks of the topological structure of $\mathbb{R}^n$. Roughly speaking, **topology** is about the properties of open sets.
- An n-dimensional open ball (or open sphere) with center $\mathbf{x}$ and radius r is defined by
$$B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^n | d(\mathbf{x}, \mathbf{y}) < r\},$$
which is the collection of points of distance less than r from a fixed point in $\mathbb{R}^n$.
- The open ball for $n = 1$ is called an open interval.
- A subset U of $\mathbb{R}^n$ is called open if for every $\mathbf{x}$ in U there exists an $r > 0$ such that $B_r(\mathbf{x})$ is contained in U.
- A set is open iff it contains an open ball around each of its points, i.e., open balls are the **base** of all open sets,[1] where "iff" is read as "if and only if".
- Intuitively, an open set is "fat" and does not contain its own "boundary".

[1] It is not hard to show that if U is open, then $U = \bigcup_{\mathbf{x} \in U} B_{r_{\mathbf{x}}}(\mathbf{x})$, where we use $r_{\mathbf{x}}$ to indicate that the radius depends on $\mathbf{x}$.

Neighborhoods and Closed Sets

- A neighborhood of the point $\mathbf{x}$ is any subset N of $\mathbb{R}^n$ that contains an open ball about $\mathbf{x}$ as a subset.
  - Intuitively speaking, a neighborhood of a point is a set of points containing that point where one can move some amount away from that point without leaving the set.
  - The neighborhood N need not be an open set itself. If N is open it is called an open neighborhood. Some books require that neighborhoods be open; we will follow this convention.
- The complement of an open set is called closed.
  - Intuitively, a closed set contains its own "boundary".

Example

$(0,1)$ is open, $[0,1]$ is closed, and $[0,1)$ is neither open nor closed.

Compact Sets

- We will not state the general definition of compactness, but only state the famous Heine-Borel theorem.

Theorem (Heine-Borel Theorem)

A set $E \subseteq \mathbb{R}^n$ is compact iff it is bounded and closed.

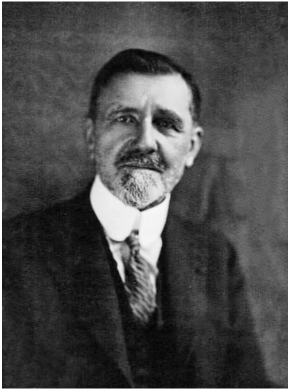
- A set E is bounded if there is real number r and a point $q \in E$ such that $d(p,q) < r$ for all $p \in E$.
  - E is bounded means it can be covered by an open ball $B_r(q)$ of finite radius.

Example

$[0,\infty)$ and $(0,1)$ are not compact by the Heine-Borel theorem, but $[0,1]$ is.

History of the Heine-Borel Theorem

H.E. Heine (1821 - 1881), German



Émile Borel (1881-1956), French

Single Variable Calculus

Calculus

- In calculus, we study functions with domain being a subset of $\mathbb{R}^n$.
- The "marginal effect analysis" in economics is usually captured by the derivative of a particular function (e.g., marginal utility, marginal cost, marginal revenue, etc.).
- This is why we would review properties of continuous and "smooth" functions.
- The foundation and starting point of calculus is the concept "limit".
- We first define the limit of a sequence and then the limit of a function.
- A sequence $\{x_n\}_{n=1}^{\infty}$ is a mapping from $\mathbb{N}$, the set of natural numbers, to some range space.
  - A sequence is automatically ordered, but its **terms** need not be distinct (like a set). Typically the range space is $\mathbb{R}$ although can be extended to any other space.

Limit of a Sequence and Limit of a Function

- A sequence $\{x_n\}_{n=1}^{\infty}$ is said to converge if there is a value $x \in \mathbb{R}$ such that $\forall\ \varepsilon > 0$, $\exists\ n_0 \in \mathbb{N}$ which may depend on ε, such that for all $n > n_0$, $|x_n - x| < \varepsilon$, where "$\forall$" means "Any", and $\exists$ means "Exist". x is called the limit of $\{x_n\}_{n=1}^{\infty}$, and we write $x_n \to x$ or $\lim_{n\to\infty} x_n = x$.
  - Intuitively, $\lim_{n\to\infty} x_n = x$ means that for n large enough, x_n will stay in an arbitrary small neighborhood of x.

Example

Does the sequence $\{x_n\}_{n=1}^{\infty}$ with $x_n = 1 + (-1)^n/n$ converge?

- Let $f : [a,b] \to \mathbb{R}$. For any $x \in [a,b]$, we claim $\lim_{t\to x} f(t) = y$ if $\forall\ \varepsilon > 0$, $\exists\ \delta > 0$ which may depend on ε, such that $|f(t) - y| < \varepsilon$ for all t with $|t - x| < \delta$.
  - $\lim_{t\to x} f(t) = y$ is equivalent to that for **any** sequence $\{t_n\}$ (with t_n not equal to x for all n) converging to x a the sequence $f(t_n)$ converges to y.

$\varepsilon - \delta$ Language

Augustin-Louis Cauchy (1789-1857), French, a pioneer of analysis

Continuity

- Let $f:[a,b]\to\mathbb{R}$. For any $x\in[a,b]$, f is said to be continuous at x if $\lim_{t\to x} f(t)=f(x)$. If f is continuous at every point on $[a,b]$, then f is said to be continuous on $[a,b]$ and denote $f\in C[a,b]$.
  - Some books define $t=x+\Delta$, so $\lim_{t\to x} f(t)=f(x)$ is equivalently written as $\lim_{\Delta\to 0} f(x+\Delta)=f(x)$, which can be understood as for **any** sequence $\Delta_n\to 0$, $f(x+\Delta_n)\to f(x)$.

Theorem (Intermediate Value Theorem)

Let $f:[a,b]\to\mathbb{R}$ be continuous with $f(a)<f(b)$. Then for any value $M\in(f(a),f(b))$, there is a $c\in(a,b)$ such that $f(c)=M$.

- c need not be unique. [Figure here]
- Intuition: the graph of a continuous function on a closed interval can be drawn without lifting your pencil from the paper.

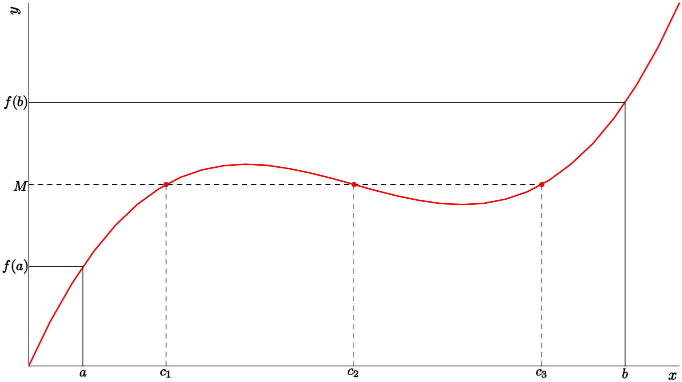


Figure: Intermediate Value Theorem

Differentiability

- Let $f:[a,b]\to\mathbb{R}$. For any $x\in[a,b]$ form the quotient

$$\phi(t)=\frac{f(t)-f(x)}{t-x}\ (a<t<b,t\neq x),$$

and define the derivative of f at x as

$$f'(x)=\lim_{t\to x}\phi(t)$$

provided this limit exists. If f' is defined at a point x, we say f is differentiable at x. If f' is defined at every point of $[a,b]$, we say f is differentiable on $[a,b]$. If f' is further continuous on $[a,b]$, we say f is continuously differentiable or smooth on $[a,b]$ and denote $f\in C^1[a,b]$. (Exercise)

- Intuition: the derivative is the limit of local slopes.
- The notation of $f'(x)$ is attributed to Newton. The corresponding Leibniz's notation is $\frac{dy}{dx}$ or $\frac{df}{dx}(x)$, where $y\equiv f(x)$.
  - One advantage of Leibniz's notation is that we can intuitively write $dy=f'(x)dx$.
  - If we want to emphasize that the derivative is taken at a specific point, say x_0, then we may write $f'(x_0)$ as $\left.\frac{dy}{dx}\right|_{x=x_0}$.

Derivatives

Example

Suppose $f(x) = x^2$. We want to calculate its derivative at point $x = 2$. Using the definition of derivative, we have

$$f'(2) = \lim_{t\to 2} \frac{f(t) - f(2)}{t-2} = \lim_{t\to 2} \frac{t^2 - 2^2}{t-2} = \lim_{t\to 2} \frac{(t+2)(t-2)}{t-2} = \lim_{t\to 2} (t+2) = 4.$$

- To avoid the burden of calculating the derivative using its definition, summarize the derivatives of popular functions in the following table.

$f(x)$	$f'(x)$	$f(x)$	$f'(x)$
c	0	$\exp(x)$	$\exp(x)$
cx	c	a^x	$a^x \ln(a)$
x^2	$2x$	$\ln(x)$	$1/x$
x^n	nx^{n-1}	$\log_a(x)$	$1/(x \ln a)$
x^{-1}	$-1/x^2$	$\sin(x)$	$\cos(x)$
$\sqrt{x}$	$\frac{1}{2}\frac{1}{\sqrt{x}}$	$\cos(x)$	$-\sin(x)$

Table: Derivatives of Popular Functions

Rules of Differentiation

Theorem

Suppose f and g are defined on $[a,b]$ and are differentiable at a point $x \in (a,b)$. Then $f+g$, fg, and f/g are differentiable at x, and

(i) (Sum Rule) $(f+g)'(x) = f'(x) + g'(x)$;

(ii) (Product Rule) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$;

(iii) (Quotient Rule) $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2}$.

- In the quotient rule, if $f = 1$, then we get the **reciprocal rule**: $\left(\frac{1}{g}\right)'(x) = -\frac{g'(x)}{g(x)^2}$.

Monotone Functions

- The sign of the derivative of a function can be used to check whether it is monotone.
- A function f is said to be non-decreasing (or increasing) if $f(y) \geq f(x)$ whenever $y > x$. It is non-increasing (or decreasing) if $-f$ is nondecreasing (increasing). A strictly increasing (strictly decreasing) function changes the above inequality to be strict. A monotone (or monotonic) function is either non-decreasing or non-increasing. A strictly monotone function is either strictly increasing or strictly decreasing.
  - Some books use "increasing" for our "strictly increasing".

Example

$f(x) = 2x$ is monotone on $\mathbb{R}$. $f(x) = x^2$ is not monotone on $\mathbb{R}$, but is monotone on $\mathbb{R}_+ \equiv [0, \infty)$.

Checking Monotonicity by Derivative

Theorem

Let $f : [a,b] \to \mathbb{R}$ *be continuous on* $[a,b]$ *and differentiable on* (a,b)*.*

(i) $f'(x) \geq 0$ *for* $x \in (a,b)$ *iff* $f(x)$ *is non-decreasing;*

(ii) $f'(x) \leq 0$ *for* $x \in (a,b)$ *iff* $f(x)$ *is non-increasing;*

(iii) *if* $f'(x) > 0$ *for* $x \in (a,b)$*, then* $f(x)$ *is strictly increasing;*

(iv) *if* $f'(x) < 0$ *for* $x \in (a,b)$*, then* $f(x)$ *is strictly decreasing.*

- A strictly increasing function f need not have $f'(x) > 0$ for any $x \in (a,b)$, e.g., for $f(x) = x^3$ on $\mathbb{R}$, $f'(0) = 0$.

Chain Rule and the Mean Value Theorem

Theorem (Chain Rule)

If g is a function that is differentiable at a point c and f is a function that is differentiable at $g(c)$, then the composite function $f \circ g$ is differentiable at c, and the derivative is

$$(f \circ g)'(c) = f'(g(c)) \cdot g'(c),$$

or in short, $(f \circ g)' = (f' \circ g) \cdot g'$.

Theorem (MVT)

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

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

- Intuition: given a planar arc between two endpoints, there is at least one point at which the tangent to the arc is parallel to the secant through its endpoints. [Figure here]

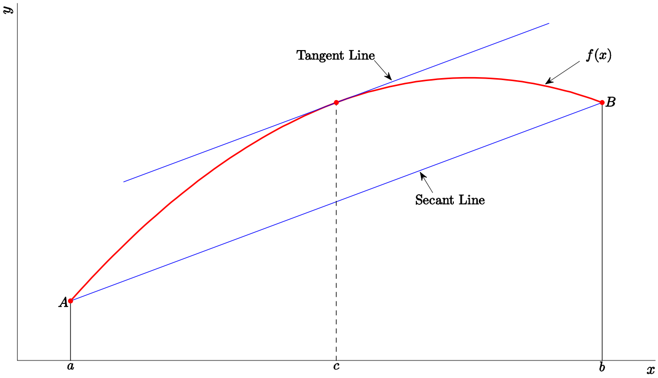


Figure: Mean Value Theorem

Higher-order Derivative

- If f has a derivative f' on an interval, and if f' is itself differentiable, we denote the derivative of f' by f'' and call f'' the **second derivative** of f.
- Continuing in this manner, we obtain functions

$$f, f', f'', f^{(3)}, \cdots, f^{(k)},$$

each of which is the derivative of the preceding one. $f^{(k)}$ is called the kth derivative, or the derivative of order k, of f.

- In Leibniz's notation, $f^{(k)}(x) = \frac{d^k y}{dx^k}$, where $y \equiv f(x)$.

- For a curve, f' means its slope and f'' mean its curvature; that is why monotonicity and concavity of a function (which will be defined in Chapter 3) are related to its first and second order derivatives, respectively.

- Taylor's theorem gives an approximation of a k times differentiable function around a given point by a k-th order Taylor polynomial.

- We will not use this theorem in this course.

History of Riemann Integral

Georg F.B. Riemann (1826 - 1866), German[2]

[2]He was a student of Gauss.

Riemann Integral

- Let $f:[a,b]\to\mathbb{R}$. A partition of $[a,b]$ is a finite sequence $P=\{x_j\}_{j=0}^{n}$ such that $a=x_0<x_1<\cdots<x_n=b$.
- The Riemann sum of f with respect to the partition P is

$$\sum_{i=0}^{n-1} f(t_i)(x_{i+1}-x_i),$$

  where $t_i\in[x_i,x_{i+1}]$.
  - each term represents the (signed) area of a rectangle with height $f(t_i)$ and width $x_{i+1}-x_i$. [Figure here]
- The Riemann integral is the limit of the Riemann sums of a function as the partitions get finer, and is often denoted as $\int_a^b f(x)dx$. If the limit exists then the function is said to be Riemann integrable.

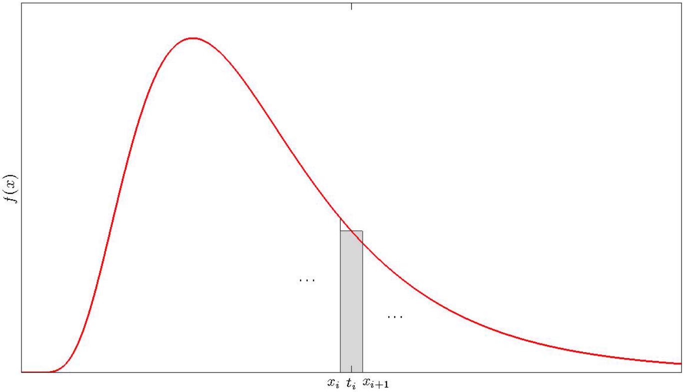


Figure: Riemann Sum

Co-Inventers of Calculus

Isaac Newton (1642-1726), English

Gottfried Leibniz (1646-1716), German

- We usually say Newton and Leibniz invented calculus because they found the **fundamental theorem of calculus** which links the concept of the derivative of a function with the concept of the function's integral.

Fundamental Theorem of Calculus

Theorem (Fundamental Theorem of Calculus)

Part I: Let $f:[a,b]\to\mathbb{R}$ be ***continuous****, and $F:[a,b]\to\mathbb{R}$ be defined, for all $x\in[a,b]$, by*

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

Then, F is differentiable on the open interval (a,b), and

$$F'(x)=f(x)$$

for all $x\in(a,b)$. [Figure here]
Part II: Let $f:[a,b]\to\mathbb{R}$ be ***Riemann integrable****, and $F:[a,b]\to\mathbb{R}$ be continuous and $F'(x)=f(x)$ for all $x\in(a,b)$. Then*

$$\int_a^b f(t)dt=F(b)-F(a).$$

- Part I: guarantees the existence of (infinitely-many) antiderivatives (or indefinite integrals) for continuous functions.
- Part II: simplifies the computation of the definite integral of a function by any of its indefinite integrals.

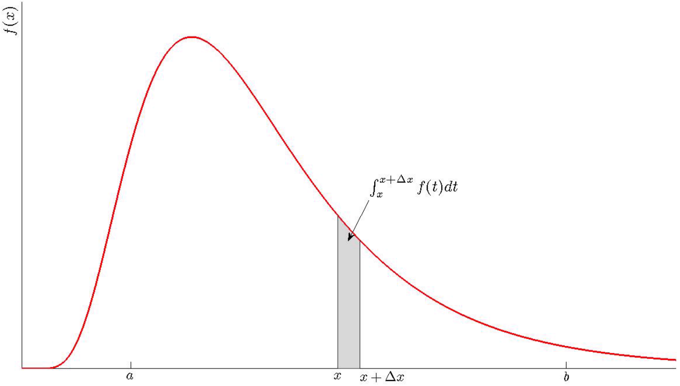


Figure: Fundamental Theorem of Calculus: Part I

- $F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\int_x^{x+\Delta x} f(t)dt}{\Delta x} = f(x).$

Multivariable Calculus

Differentiability

- The derivative $f'(x)$ when $x \in \mathbb{R}$ can be equivalently reexpressed as follows: there exists a linear function $f'(x)h$ such that

$$\lim_{h\to 0}\frac{r(h)}{h} \equiv \lim_{h\to 0}\frac{f(x+h)-f(x)-f'(x)h}{h}=0,$$

  where r is for "remainder".
- A function $\mathbf{f}: E \to \mathbb{R}^m$ is said to be differentiable at $\mathbf{x}$ if there exists a linear map $\mathbf{J}: \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|\mathbf{r}(\mathbf{h})\|_{\mathbb{R}^m}}{\|\mathbf{h}\|_{\mathbb{R}^n}} \equiv \lim_{\mathbf{h}\to\mathbf{0}}\frac{\|\mathbf{f}(\mathbf{x}+\mathbf{h})-\mathbf{f}(\mathbf{x})-\mathbf{J}(\mathbf{h})\|}{\|\mathbf{h}\|}=0,$$

  where E is an open set in $\mathbb{R}^n$, $\|\cdot\|$ is the Euclidean norm, $\mathbf{J}(\mathbf{h})=\mathbf{J}\mathbf{h}$, and the $m\times n$ matrix $\mathbf{J}$ is called the Jacobian matrix at $\mathbf{x}$. We write $\mathbf{f}'(\mathbf{x})=\mathbf{J}$. If $\mathbf{f}$ is differentiable at every $\mathbf{x}\in E$, we say f is differentiable in E.
- Notations: Matrices are represented using uppercase bold, e.g., $\mathbf{J}$; Often, $\mathbf{f}'(\mathbf{x})$ are denoted as $D_{\mathbf{x}}\mathbf{f}(\mathbf{x})$ or $D\mathbf{f}(\mathbf{x})$.
  - Note that $\mathbf{J}$ depends on $\mathbf{x}$ as all these notations indicate.
- When $m=n$, $\mathbf{J}$ is square. Both the matrix and its determinant are referred to as the Jacobian in literature.

Partial Derivatives

- The derivative defined above is often called the total derivative of $\mathbf{f}$ at $\mathbf{x}$.
- Problem: how to find $\mathbf{J}$ in practice?
- Let $\mathbf{f} = (f_1, \cdots, f_m)'$. The partial derivative of f_i at $\mathbf{x}$ with respect to the j-th variable is defined as

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x}) = \lim_{h \to 0} \frac{f_i(x_1, \cdots, x_{j-1}, x_j + h, x_{j+1}, \cdots, x_n) - f_i(x_1, \cdots, x_{j-1}, x_j, x_{j+1}, \cdots, x_n)}{h}$$

  provided the limit exists.
  - $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ measures how much f_i would change when all other variables except x_j are fixed at $\mathbf{x}$ and only x_j changes a little bit, so it is very useful in economics for "ceteris paribus" analysis.
- If $\mathbf{f}$ is differentiable at $\mathbf{x}$, then all the partial derivatives at $\mathbf{x}$ exist, and

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{pmatrix} \equiv \frac{\partial \mathbf{f}}{\partial \mathbf{x}'}(\mathbf{x}).$$

Comparison of the Total and Partial Derivative

- When $m = 1$, $Df(\mathbf{x})$ is a $1 \times n$ (row) vector; we may intuitively express the total derivative in the form of total differential,

$$dy = \frac{\partial f}{\partial x_1}(\mathbf{x})\, dx_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x})\, dx_n,$$

or equivalently,

$$\frac{dy}{dx_1} = \frac{\partial f}{\partial x_1}(\mathbf{x}) + \frac{\partial f}{\partial x_2}(\mathbf{x}) \frac{dx_2}{dx_1} + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{x}) \frac{dx_n}{dx_1}.$$

- Obviously, the total derivative must take into account of the change of $(x_2, \cdots, x_n)$ as x_1 changes, which is dramatically different from the partial derivative.
- In other words, the path of $\mathbf{h} \to \mathbf{0}$ in $\mathbb{R}^n$ in the definition of total derivative is not restricted while the path of $\mathbf{h} \to \mathbf{0}$ in the definition of partial derivative $\partial f_i / \partial x_j (\mathbf{x})$ is restricted to be along the axis (i.e., $h_j \to 0$ and $h_k = 0$ if $k \neq j$). [Figure here, $m = 1$]
- Even if all partial derivatives $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ exist at a given point $\mathbf{x}$, f need not be (totally) differentiable, or even continuous in the sense that $\lim_{\mathbf{h} \to \mathbf{0}} \mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x})$ (Exercise).
- If all partial derivatives exist in a neighborhood of $\mathbf{x}$ and are continuous there, then $\mathbf{f}$ is (totally) differentiable in that neighborhood and the total derivative is continuous. In this case, it is said that $\mathbf{f}$ is a C^1 function.

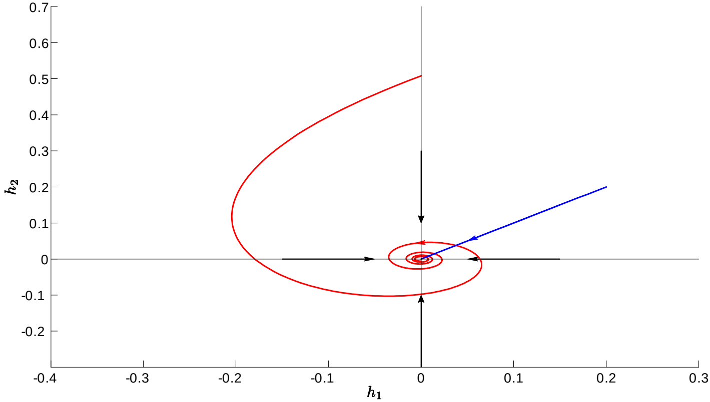


Figure: Total Derivative, Partial Derivtive and Directional Derivative: Red Arrow for Total Derivative, Blue Arrow for Directional Derivative, Black Arrows along y-axis for $\partial f / \partial x_2 (\mathbf{x})$ and Black Arrows along x-axis for $\partial f / \partial x_1 (\mathbf{x})$

Young's Theorem

- The partial derivative $\partial f_i / \partial x_j$ can be seen as another function defined on E and can again be partially differentiated.
- To simplify notations, suppose $m = 1$, i.e., $f : E \to \mathbb{R}$.
- Problem: whether the order of differentiation matters, that is, whether $\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) \equiv \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(\mathbf{x})\right)$ equals $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \equiv \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}(\mathbf{x})\right)$.

Theorem (Young's Theorem)

*Let $f : E \to \mathbb{R}$, where E is an open set in $\mathbb{R}^n$. If f has **continuous** second partial derivatives at $\mathbf{x}$, then*

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}).$$

- If all mixed second order partial derivatives are continuous at a point (or on a set), f is termed a C^2 function at that point (or on that set).
  - Intuition: C^∞-plump/fat, C^0 but not C^1-slim/thin

Hessian Matrix

- Young's theorem implies that for a C^2 function f at $\mathbf{x}$, the Hessian matrix f at $\mathbf{x}$, $\mathbf{H}(\mathbf{x})$, is symmetric, where

$$\mathbf{H}(\mathbf{x}) \equiv \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{pmatrix} = \frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}'}(\mathbf{x})$$

  with the (i,j)th element being $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$.
- The Hessian matrix of f at $\mathbf{x}$ is often denoted as $D_{\mathbf{x}}^2 f(\mathbf{x})$ or $D^2 f(\mathbf{x})$.
- As in the case of the Jacobian, the term "Hessian" unfortunately appears to be used both to refer to this matrix and to the determinant of this matrix.

Higher Order Partial Derivatives and Multiple Integral

- As the higher-order derivatives in single variable calculas, we can similarly define higher-order partial derivatives in multivariate calculus. Also, Taylor's theorem can be extended to the multivariate case.
- We can also similarly define the multiple integral

$$\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(\mathbf{x}) dx_1 \cdots dx_n$$

as in the single variable case.

- Integrals of a function of two variables over a region in $\mathbb{R}^2$ are called double integrals, and integrals of a function of three variables over a region of $\mathbb{R}^3$ are called triple integrals. They can be used to calculate areas and volumes of regions in the plane, respectively.