Econ 600: taught by Prof. Shaowei Ke

Linfeng Li
llinfeng@umich.edu

September 23, 2015

Disclaimer

This is a personal note of mine. I will try to follow professor Ke's lecture as close as possible. However, neither is this an official lecture note, nor will Linfeng be responsible for any errors + typos. Nevertheless, corrections and suggestions are always welcomed.

As this lecture note will be maintained on Github, PLEASE:

- Use the "Issues" feature on Github to post suggestions;
- Feel free to fork this repo and send me pull requests.

Paragraphs starting with "Note that ..." are most likely my personal reflections. Please be aware of this.

1 Lecture 1: Logic, Sets and some Real Analysis[1]

1.1 Logic

Definition 1.1. Proposition is a sentence that is either *true* or *false*. It cannot be both true and false.

Note: "true" and "false" may not necessarily be based on any (objective/subjective) factual basis. However, to give a concrete example, contextually correct propositions are usually employed.

Definition 1.2. Logic Connectives: $\wedge$ and $\vee$. Let P and Q be propositions

- Conjunction of P and Q is denoted as $P \wedge Q$;
- Disjunction of P and Q is denoted as $P \vee Q$.
- Negation of P is denoted as: $\neg P$.

P	Q	$P \wedge Q$	$P \vee Q$	$\neg P$
1	1	1	1	0
1	0	0	1	0
0	1	0	1	0
0	0	0	0	1

Table 1: Truth Table for logic connectives

Truth Table is vaguely defined, with each row being a possible "state of the world". On top of this,

Definition 1.3 (Conditionals and Biconditionals)**.** Let P, Q, R be propositions,

1. Conditional of P and Q is $P \implies Q$;
2. Biconditional of P and Q is $P \iff Q$.

P	Q	$P \implies Q$	$P \iff Q$
1	1	1	1
1	0	0	0
0	1	1	0
0	0	1	1

Table 2: Truth Table for Conditionals and Biconditionals

Note that, the two 1's are obtained for free. Conditional of P and Q are trivially true if P is false (thus the conditional is not entered, thereby cannot be disproved?). ⇐Check This.

Additionally, from an external source (← click me!):

> Conditionals are FALSE only when the first condition (if) is true and the second condition (then) is false. All other cases are TRUE.

Definition 1.4. Two propositions are **equivalent** if they have the same truth table, denoted using "≡".

Example 1. Claim: that $P \implies Q$ and $\neg Q \implies \neg P$ are equivalent.

Proof. Refer to table 3: that by definition, the truth table of the two conditionals are the same. □

Note, (it seems that)[a] truth tables are the same if the two "column vectors" denoting the true/false status are the same.

[a]Since "truth table" was not explicitly defined.

Definition 1.5 (Tautology)**.** A proposition whose truth table consists only 1's is called **tautology**.

[1]Relation, Function, Correspondence and Sequences in $\mathbb{R}$

Table 3: Truth Table: equivalence of $P \implies Q$ and $\neg Q \implies \neg P$

P	Q	$P \implies Q$	$\neg Q \implies \neg P$
1	1	1	1
1	0	0	0
0	1	1	1
0	0	1	1

Example 2. Claim: $Q \implies (P \implies Q)$ is a tautology.

Proof. Refer to Table 4 □

Table 4: Truth Table: Tautology

P	Q	$P \implies Q$	$Q \implies (P \implies Q)$
1	1	1	1
1	0	0	1
0	1	1	1
0	0	1	1

Remark 1.6. We introduce the following 4 types of proof:

1. Direct proof: to follow the direction of the statement.
   - **Proposition**: For odd integers x, y, $x + y$ is an even integer.
2. Proof by contrapositive: (restate the proposition and prove the easier direction).
   - **Proposition**: If n^2 is an odd integer (P), then n is an odd integer.

     Proof. Prove instead that: "if n is an even integer, then n^2 is an even integer". □
3. Proof by contradiction: (construct a structure that leads to contradiction between derived conditions and given conditions.).
   - That $\sqrt{2}$ is rational number[2].
4. Proving a "if and only if" statement/proposition to be true: either one of the following 4 are valid strategies:
   (a) $P \implies Q$ and $Q \implies P$;
   (b) $P \implies Q$ and $\neg P \implies \neg Q$;
   (c) $\neg Q \implies \neg P$ and $Q \implies P$;
   (d) $\neg Q \implies \neg P$ and $\neg P \implies \neg Q$.

[2]The set of rational numbers is denoted as Q.

1.2 Sets

Remark 1.7 (Russell's paradox)**.** The barber is a man who shaves all those and only those who do not shave themselves.

In terms of set theory, let $R = \{x : x \notin x\}$, then:

$$R \in R \iff R \notin R$$

which is very problematic.

Definition 1.8 (Sets)**.** There are two definition of sets:

1. (Enumerating all elements)

   A set is a collection of objects, e.g. $\{1, 2, \ldots\}$ [3] or $\{1, 2\}$ [4].

2. (Describing properties to be satisfied by elements in the set)

   If A is a set of all objects that satisfies property P, then we can write

   $$A = \{x : P(x)\}$$

   where the colon means "such that", and $P(x)$ means that x satisfies property P.

Now, we can define the following **sets** using the two definitions of sets:

- (Natural Number) $N = \{1, 2, \ldots\}$;
- (Integer) $Z = \{x : x = n \text{ or } x = -n \text{ or } x = 0, \text{ for some } n \in N\}$;
- (Rational number) $Q = \{x : x = \frac{m}{n}, m, n \in Z\}$.

Definition 1.9 (Set Equality)**.** Two sets A and B are equal if they have the same elements. That is:

$$A = B \text{ if and only if } x \in A \iff x \in B, \forall x$$

Note, that the notion $\forall x$ was used sloppily here. Without loss of generality, it shall better be $\forall x \in A \bigcup B$.

Definition 1.10 (Set Containment)**.** A set A is contained in a set B, denoted by $A \subseteq B$, if $\forall x \in A \implies x \in B$.

As a consequence, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

Definition 1.11 (Cardinality (finite case))**.** If a set A has $n \in N$[5] distinct elements, then n is the cardinality of A and we call A a finite set. The **cardinality of** A is denoted by $|A|$.

Definition 1.12 (Empty set $\emptyset$)**.** The empty set is the set with no element.

[3] a countably infinite set.

[4] a finite set.

[5] Natural number.

Definition 1.13 (Power set 2^A)**.** Let A be a set. The **power set of** A is the collection of all subsets of A.

Note that, A is an arbitrary set. It could be finite, in which case 2^A easy to envision; At the other extreme, it could be a uncountable set. Nevertheless, the following equality shall hold:

$$|2^A| = 2^{|A|}$$

Example 3. Let $A = \{1, 3\}$, then $2^A = \big\{\emptyset, \{1\}, \{3\}, \{1, 3\}\big\}$. In terms of notation, note that 1 is an element in A, thus $1 \in A$; yet, $\{1\}$ is a subset of A, thus $\{1\} \subset A$.

Definition 1.14 (Operations on sets: $\bigcup$, $\bigcap$, $\setminus$ and $\cdot^c$.)**.** Let A and B be two sets:

- Union: $A \bigcup B \coloneqq \{x : x \in A \vee x \in B\}$;
- Intersection: $A \bigcap B \coloneqq \{x : x \in A \wedge x \in B\}$;
- A and B is disjoint if $A \bigcup B = \emptyset$;
- Difference of A and B is defined as: $A \setminus B \coloneqq \{x \in A \wedge x \notin B\}$;
- Complements of A: $A^c \coloneqq \{x : x \notin A\}$.

Side note: **Index set** I is a countable set.

$$\bigcup_{i \in I} A_i = \{x : x \in A_i \text{ for some } i \in I\}$$

Definition 1.15 (de Morgan's law)**.**

$$\left(\bigcup_{i \in I} A_i\right)^c = \bigcap_{i \in I} (A_i^c) \text{ and } \left(\bigcap_{i \in I} A_i\right)^c = \bigcup_{i \in I} (A_i^c)$$

Exercise 1.16. Prove that $(A \bigcup B)^c = A^c \bigcap B^c$.

Proof. Prove mutual containment using element argument. $\square$

Counters reset

1.3 Relation, Function and Correspondence

Definition 1.1 (Ordered pair)**.** For two sets A and B, an ordered pair is (a, b) such that $a \in A$ and $b \in B$.

Definition 1.2 (n-taple)**.** Let there be n sets: $A_1, \ldots, A_n$, an n-taple is $(a_1, \ldots, a_n)$ such that $a_i \in A_i$, $\forall i = 1, 2, \ldots n$.

Definition 1.3 (Cartesian Product)**.** Let $A_1, \ldots, A_n$ be non-empty sets. Cartesion product of $A_1, \ldots, A_n$ is $A_1 \times \cdots \times A_n$, defined as:

$$\Pi_{i=1}^{n} A_i = \{(a_1, \ldots, a_n) : a_i \in A_i, \forall i = 1, \ldots, n\}$$

Definition 1.4 (Relation)**.** A relation from set A to set B is a subset of $A \times B$, denoted by R.

$$aRb \iff (a, b) \in R$$

A relation on A is a subset of $A \times A$.

Definition 1.5. A relation $R \subseteq A \times A$ is said to be:

- *reflective* if aRa $\forall a \in A$. (That is, $(a, a) \in R$, $\forall a \in A$.);
- *complete* if either aRb or bRa, $\forall a, b \in A$;
- *symmetric* if $\forall a, b \in A$, $aRb \implies bRa$;
- *antisymmetric* if $\forall a, b \in A$, aRb and $bRa \implies a = b$.
- *transitive* if $\forall a, b, c \in A$ s.t. aRb and bRc, aRc (is implied).

Table 5: Property of common relations

	$<$	$\leq$	$\in$	$\subseteq$	$\succeq$
reflective	✗	✓	✗	✓	✓
complete	✗	✓	✗	✗	✓
symmetric	✗	✗	✗	✗	✗
antisymmetric	✓	✓	✓	✓	✗
transitive	✓	✓	✗	✓	✓

Note that, $<$ and $\leq$ are defined on $\mathbb{R}$; $\in$ and $\subseteq$ are defined on sets; $\succeq$ is preference relation that represents "weakly prefer".

Also note that, completeness implies reflectiveness.

Definition 1.6 (Equivilance relation)**.** An **equivalence** on set A is a relation E that is *reflective, symmetric and transitive.* It is denoted as $\sim$.

For any $a \in A$, the **equivalence class** of a with respect to $\sim$ is defined to be the set

$$E_{\sim}(a) = \{b \in A, b \sim a\}$$

Remark: by construction in Definition 1.4, equivalence ($\sim$) is defined as "a relation on A", which is thereby defined in the Cartesian space.

Definition 1.7 (Function: defined using Relation from A to B)**.** A function from set A to set B is a relation f from A to B such that:

(i) $\forall a \in A$, $\exists b \in B$ such that $(a, b) \in f$, i.e. afb

(ii) $\forall a \in A$, if $(a, b) \in f$ and $(a, c) \in f$ for some $b, c \in B$, then $b = c$.

Note that, alternatively, the two conditions could be written in short as:

(iii) $\forall a \in A$, $\exists! b \in B$ such that $(a, b) \in f$, i.e. afb

Convention for f: If $(a, b) \in f$, we write $f(a) = b$. And, f could be interpreted as a "mapping": "$f : A \to B$".

Definition 1.8 (Domain and Rnage)**.** If f is a function from A to B, then A is called the **domain** of f and B is the **codomain** of f. The **range** of f is the set:

$$Ran(f) = \{b \in B : \exists a \in A \text{ s.t. } f(a) = b\}.$$

Definition 1.9 (Propoteries of functions)**.** Let f be a function, then:

(i) f is **surjective** if $Ran(f) = B$; onto

(ii) f is **injective** if $a_1 \neq a_2 \in A \implies f(a_1) \neq f(a_2)$; 1-to-1

(iii) f is bijective if f is surjective and injective.

Side note: a *indicator function* is defined as following: for A being a set and $S \subseteq A$,

$$\mathcal{X}_S(a) = \begin{cases} 1 & \text{if } a \in S \\ 0 & \text{otherwise} \end{cases}$$

Definition 1.10 (Image and Preimage)**.** For $f : A \to B$ and $C \subseteq A$, the **image** of C under f is

$$f(C) = \{b \in B : \exists a \in C \text{ s.t. } f(a) = b\}$$

The **preimage** of $D \subseteq B$ is

$$f^{-1}(D) = \{a \in A : f(a) \in D\}$$

Exercise. Prove that

1. $f^{-1}(f(A)) = A$, and
2. $f(f^{-1}(B)) = B$ if and only if f is surjective.

Proposition 1.11. Given $f : A \to B$, then $f^{-1} : B \to A$ is a function if and only if f is bijective.

Definition 1.12 (Sequence)**.** A sequence is a function $f : N \to A$, denoted by $\{a_1, a_2, \ldots\} = \{a_i\}_{i=1}^{\infty}$[6] i.e. the set of all sequence is the following set:

$$A^\infty = A \times A \times \cdots$$

Definition 1.13 (Cardinality, for (infinite) sequences)**.** Two sets A, B have the same cardinality if $\exists$ a bijective function $f : A \to B$.

Then, $|A| \geq |B|$ if there exists an injective function $f : B \to A$. (Example: $|Z| \geq |N|$ by using identify mapping from N to Z; $|N| \geq |Z|$ by enumerating elements in Z using N. Thus, $|Z| = |N|$.) Eventually, we have:

$$|\mathbb{R}^2| = |\mathbb{R}| > |Q| = |Z| = |N|$$

Definition 1.14 (Correspondence)**.** $T : A \rightrightarrows B$ is a correspondence such that $T : A \to 2^A \backslash \emptyset$.

1.4 Sequences

Definition 1.1 (Sequence in $\mathbb{R}$)**.** A sequence of real number is a function $a : N \to \mathbb{R}$ s.t. $a(i) = a_i$ is the i-th component of the sequence $\{a_j\}_{j=1}^{\infty}$.

Definition 1.2 (Increasing sequence)**.** A real sequence is increasing if $a_{n+1} \geq a_n \ \forall n \in N$.

Definition 1.3 (Bounded and Bounded (from) above/below)**.** A real sequence is

- **bounded above** if $\exists \bar{m} \in \mathbb{R}$ s.t. $a_n \leq \bar{m} \ \forall n \in N$.
- **bounded below** if $\exists \underline{m} \in \mathbb{R}$ s.t. $a_n \geq \underline{m} \ \forall n \in N$.
- **bounded** if it is bounded above and bounded below.

Definition 1.4 (Least upper bound)**.** $a \in \mathbb{R}$ is the least upper bound of a sequence $\{a_n\}$ if

(i) a is an upper bound;

(ii) a is the smallest upper bound, i.e. $\nexists b \in \mathbb{R}$ s.t. $b < a$ and b is a upper bound of $\{a_n\}$.

Axiom 1.5 (Axiom of Real Number: completeness axiom)**.** If S is a nonempty set of real numbers that is bounded above, then there exists a least upper bound ~~that is also a real number~~.

Note, that, claiming that the upper bound is in $\mathbb{R}$ is redundant.

Definition 1.6 (Convergence sequences)**.** A real sequence $\{a_n\}$ converges to the limit $a \in \mathbb{R}$ if $\forall \varepsilon > 0$, $\exists N$ s.t. $\forall n \geq N$

$$|a_n - a| < \varepsilon$$

We write $\lim\limits_{n\to\infty} a_n = a$ or $a_n \to a$.

- If a sequence does not converge, then it diverges. (To $+\infty$ or $-\infty$.)

Theorem 1. A monotone bounded sequence converges.

Proof. Discuss two cases where 1) $\{a_n\}$ is an increasing sequence, and 2) $\{a_n\}$ is a decreasing sequence. Then, proof is completed through using either least upper bound (for increasing sequence) or largest lower bound (for decreasing sequence). □

[6] This is an ordered set.

2 Lecture 2: convergence and more

2.1 Sequence and Convergence

Definition 2.1. A set $S \subset X$ is a linearly ordered set if there is a relation "$\leq$" on X s.t.

$$\leq \text{ is complete, transitive and antisymmetric.}$$

Note that, given the linear ordering, we can define $<$ accordingly. (For arbitrary $a, b \in X$ and $a \leq b$, then we say $a < b$ if $a \leq b$ and $a \neq b$.)

Definition 2.2 (Boundedness for an arbitrary set.)**.** Let X be a linearly ordered set and $S \subset X$, then $a \in X$ is the **supremum** (or *least upper bound)* of X if:

1. a itself is an upper bound of S, i.e.
2. for $b \in X$, $b < a$, then b is not an upper bound of S.

Corollary: For $a = \sup X$, $\forall \varepsilon > 0$, there exists $x \in S$ s.t. $x > a - \varepsilon$.

Axiom 2.3 (Completeness Axiom)**.** If S is a nonempty set of real numbers that is bounded above, then there exists a least upper bound.

Definition 2.4 (Sequence in $\mathbb{R}$)**.** A sequence of real number is a function $a : N \to \mathbb{R}$ s.t. $a(i) = a_i$ is the i-th component of the sequence $\{a_j\}_{j=1}^{\infty}$.

Remark 2.5. $\{a_n\}$ is bounded if $a(N)$ is bounded.

Note, here N is the set of all natural numbers $\{1, 2, \ldots, \}$. Thus, we hereby define the boundedness of a sequence using the our previous definition of set-boundedness.

Lemma 2.6. A monotone bounded sequence converges.

Definition 2.7 (Subsequence)**.** A subsequence $\{a_{n_i}\}$ of $\{a_n\}$ is a sequence s.t. $1 \leq n_1 \leq n_2 \leq \ldots$. That is:

$\exists$ conversion function $\Phi : N \to N$ s.t. $n_i = \Phi(i)$ and $\Phi(i) < \Phi(j)$ whenever $i < j$. We can also write: $a_{n_i} = a_{\Phi(i)}$.

Lemma 2.8. Every sequence of $\mathbb{R}$ has a monotone subsequence.

Proof. Proof by doodling: try to construct a decreasing sequence first, if failed (cannot identify infinitely many of elements as candidate of the sequence), construct an increasing one.

Formally: let $S = \{i : \text{ if } j > i, \text{ then } a_j < a_i\}$.

- if $|S| = |N|$ (countably infinite) [1], we have found a monotone (decreasing) sequence.
- If $|S| < \infty$, let $\max S = N$, then by construction, $\exists n_1$ s.t. $a_{n_1} \geq a_{N+1}$. Since $n_1 \notin X$, there exists $n_2 > n_1$ s.t. $a_{n_2} \geq a_{n_1} \geq a_N$.

  We can construct an increasing sequence in this fashion.

[1] Writing $|S| = \infty$ is not rigorous enough, since uncountably infinite could also be denoted similarly.

□

Theorem 2.9 (Bolzano-Weierstrass Theorem)**.** A bounded sequence of $\mathbb{R}$ has a convergent subsequence.

Proof. By Lemma 2.8, such bounded sequence of $\mathbb{R}$ has a monotone subsequence, which inebriates the boundedness property.

Thus, by Lemma 2.6, such bounded monotone sequence converges. □

Remark 2.10 (Properties of Limits)**.** For $a_n \to a$ and $b_n \to b$ (two convergent sequences):

(i) $c \cdot a_n \to c \cdot a$, for $c \in \mathbb{R}$;

(ii) $a_n + b_n \to a + b$

(iii) $a_n \cdot b_n \to a \cdot b$

(iv) $\frac{a_n}{b_n} \to \frac{a}{b}$ s.t. $b \neq 0$ and $b_n \neq 0$ $\forall n$.

(v) $\forall n \in N$, if $c \leq a_n$, then $c \leq a$. (Note that we have defined only one linear ordering $\leq$.)
However, $a_n > c$ does not imply $a > c$. (e.g.: $\frac{1}{n} > 0, \forall n$, yet $\frac{1}{n} \to 0 = 0$.)

(vi) $\forall n$, if $b_n \leq a_n$, then $b \leq a$.

Definition 2.11 (Cauchy sequence)**.** $\{a_n\}$ is a Cauchy sequence if $\forall \varepsilon > 0$, $\exists N$ s.t. $\forall m, n \geq N$, $|a_m - a_n| < \varepsilon$.

Note that, since the definition of convergent sequence relies on knowing the limit a, when such limit is not attainable, Cauchy becomes handy.

Theorem 2.12. Every convergent sequence is Cauchy.

Proof. Given $\{a_n\} \to a$, thus $\forall \frac{\varepsilon}{2} > 0$ $\exists N$ s.t. $|a_n - a| < \frac{\varepsilon}{2}$, $\forall n > N$.

Now, for any $m, n \geq N$, we have:

$$\begin{aligned} |a_m - a_n| &= |a_m - a + a - a_n| \\ &\leq |a_m - a| + |a_n - a| < \varepsilon \end{aligned}$$

□

Example : Prove that $a_{n+1} = \frac{a_n + 2a_{n-1}}{3}$ converges for $a_1 = 0$, $a_2 = 1$.

Proof. Step 1 First observe that: a_n is an average of two real numbers that are in $[0, 1]$. Thus, $a_n \in [0, 1]$.

Step 2 Also observe that by rearranging the terms in the equality, we have:

$$\frac{a_{n+1} - a_n}{a_n - a_{n-1}} = -\frac{2}{3}$$

At this point, we check definition of Cauchy sequence by showing that: for arbitrary ε, we can find a N such that $|a_m - a_n| < \varepsilon$. Deriving the functional form of $|a_m - a_n|$ suffices. (We can then use this functional form to find a proper N.)

Without loss of generality, let $m > n$, then:

$$\begin{aligned}
|a_m - a_n| &= |a_n - a_{n+1} + a_{n+1} - \cdots - a_m| \\
&\leq |a_n - a_{n+1}| + |a_{n+1} - a_{n+2}| + \cdots + |a_{m-1} - a_m| \\
&\leq \left(\frac{2}{3}\right)^{n-1} + \left(\frac{2}{3}\right)^{n} + \cdots + \left(\frac{2}{3}\right)^{m-2} \\
&= \frac{\left(\frac{2}{3}\right)^{n-1}\left(1 - \left(\frac{2}{3}\right)^{m-n+2}\right)}{1 - \frac{2}{3}} \\
&= O(\left(\frac{2}{3}\right)^{n})
\end{aligned}$$

By now, we can easily demonstrate that the definition of Cauchy sequence could be satisfied by choosing a proper N for any given ε. □

Lemma 2.13. Every Cauchy sequence is bounded.

Proof. Let $\{a_n\}$ be an arbitrary Cauchy sequence. Then, for arbitrary $\varepsilon > 0$, we know that $\exists N_\varepsilon > 0$ such that $\forall m, n > N$, $|a_m - a_n| < \varepsilon$.

Now, to construct an upper bound for $\{a_n\}$, without loss of generality, let $\varepsilon = 1$. Then, we know that there exists $N_1 > 0$ such that $\forall n, m > N_1$, $|a_n - a_m| < 1$. Then, let M_1 denote the bound (either upper or lower). Then, in absolute value, we can define it to be:

$$|M_1| = \max\{|a_1|, \ldots, |a_{N_1}|, |a_{N_1+1}| + 1\}$$

Through more careful, yet unnecessary, discussions, we can derive the exact bound using the absolute value $|M_1|$.

Note that, the bound we found above is only *one of the upper bound*. It is not necessarily the sup nor inf. □

Theorem 2.14. Every Cauchy sequence `in` $\mathbb{R}$ [2] converges.

Proof. Let $\{a_n\}$ be an arbitrary Cauchy sequence. We want to show $\{a_n\}$ converges to some $a \in \mathbb{R}$. That is to show: $\forall \varepsilon > 0$, $\exists N_0 \in \mathbb{N}$ s.t. $\forall n \geq N_0$, $|a_n - a| < \varepsilon$.

(Step 1:) For arbitrary $\varepsilon > 0$, given that $\{a_n\}$ is a Cauchy sequence, for $\frac{\varepsilon}{2} > 0$, $\exists N_1 \in \mathbb{N}$ s.t.

$$|a_m - a_n| < \varepsilon, \qquad \forall m, n > N_1$$

[2] Note that, for $\{\frac{1}{n}\}$ defined on $(0, 1]$, it does not converge in this space since $0 \notin (0, 1]$.

(Step 2:) By Lemma 2.13, we know that every Cauchy sequence is bounded. Thus, by Bolzano-Weierstrass Theorem (Theorem 2.9), we know that $\exists\{a_{n_i}\} \to a$ for some certain real number $a \in \mathbb{R}$. $\Leftarrow$ a is Limit!

By definition of convergence of (sub)sequence, for the arbitrary ε that we started with, $\exists I \in \mathbb{N}$ s.t.

$$|a_{n_j} - a| < \frac{\varepsilon}{2}, \qquad \forall j > I$$

Now, let $N_0 = \max\{n_I, N_1\}$, we see that $\forall n > N_0$ and $n_j > N_0$, we have:

$$\begin{aligned} |a_n - a \ &= |a_n - a_{n_j} + a_{n_j} - a| \\ &\leq |a_n - a_{n_j}| + |a_{n_j} - a| < \varepsilon \end{aligned}$$

Note: a_{n_j} is an arbitrary element of the subsequence $\{a_{n_i}\}$ that we found convergent through B-W Theorem.

Also note that, $|a_n - a_{n_j}| < \frac{\varepsilon}{2}$ follows from Step 1 that $\{a_n\}$ is Cauchy to start with. $\square$

Definition 2.15 (Cauchy Criterion)**.** A sequence in $\mathbb{R}$ is a convergent sequence if and only if it is a Cauchy Sequence.

Demonstration: Theorem 2.12 applies in $\mathbb{R}$, thereby convergent sequence in $\mathbb{R}$ is Cauchy; Theorem 2.14 completes the proof.

Remark 2.16 (Useful limits)**.** Limits of sequences as $n \to \infty$:

- $\lim\limits_{n\to\infty} \frac{n^\alpha}{(1+p)^n} = 0$ for $p > 0$ and $\alpha > 0$. (This demonstrates exponential function dominates polynomials in the limit.)

- $\lim\limits_{n\to\infty} \sqrt[n]{n} = 1$

- $\lim\limits_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = e$; then $\lim\limits_{n\to\infty} \left(1 + \frac{t}{n}\right)^n = e^t$.

- $\lim\limits_{n\to\infty} \sqrt[n]{p} = 1$ if $p > 0$.

Refer to page 57 of [Rudin(1976)] Theorem 3.20 for detailed proofs.

Definition 2.17 (limsup, liminf)**.** Let $\{a_n\}$ be a sequence in $\mathbb{R}$, we say: $\limsup\{a_n\} = a$ if $\sup S = a$, where $S = \{b \in \mathbb{R} : \exists$ subsequence $\{a_{n_i}\}$ s.t. $a_{n_i} \to b\}$.

Not surprisingly, we can define

$$\liminf\{a_n\} = -\limsup\{-a_n\}$$

Exercise: equivalent definition of limsup Prove that $\limsup a_n = a$ if and only if:

(i) $\forall \varepsilon > 0$, $\exists N > 0$ s.t. $a_n < a + \varepsilon$, $\forall n > N$;

(ii) $\forall \varepsilon > 0$, $\forall n \in N$, $\exists k > n$ s.t. $a_k > a - \varepsilon$.

Note that, (i) specified a property for subsequence; and (ii) is merely about the existence of one element in the sequence, to be found for all $(\varepsilon, n) \in \mathbb{R}_{++} \times N$.

Proof. The iff statement will be established in the following three steps:

- Prove that $\limsup a_n = a$ implies (i).

WTS: $\forall \varepsilon > 0$, $\exists N > 0$ s.t. $a_n < a + \varepsilon$, $\forall n > N$;

First, suppose that $a = +\infty$, that is $\{a_n\}$ is not bounded from above. Then we are done.

Then, suppose that $\{a_n\}$ is bounded from above. We now prove by contradiction. Suppose that $\exists \varepsilon > 0$ s.t. no such $N \in \mathbb{N}$ exists. Then, we know that 1 cannot serve the role of N. So, for some $n_1 > 1$,

$$a_{n_1} \geq a + \varepsilon$$

Still, $n_1 + 1$ cannot serve the role of N, then for some $n_2 > n_1 + 1$,

$$a_{n_2} \geq a + \varepsilon$$

By induction, we can construct a subsequence that is bounded from below by $a + \varepsilon$. Note that, the original sequence is bounded from above, by Bolzano-Weierstrass Theorem, we know that a bounded sequence converges. However, the limit of such subsequence shall be larger than a, contradicting $\limsup a_n = a$.

Thus what we assumed is wrong. We thereby proved the original claim in (ii).

> Note that, the converse of the following claim:
>
> $$\forall \varepsilon > 0, \quad \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N, \text{ proposition } P \text{ is true.}$$
>
> $$\xRightarrow{\text{converse}} \qquad \exists \varepsilon > 0, \quad \forall N \in \mathbb{N} \text{ s.t. } \exists n \geq N, \text{ proposition } P \text{ is } \textbf{not} \text{ true.}$$

- Prove that $\limsup a_n = a$ implies (ii).

WTS: Given that $\limsup a_n = a$, $\forall \varepsilon > 0$, $\forall n \in N$, $\exists k > n$ s.t. $a_k > a - \varepsilon$.

Now, for arbitrary $\varepsilon > 0$, by definition of limsup, we know that $\exists a' \in (a - \frac{\varepsilon}{2}, a)$ s.t. $\exists \{a_{n_j}\}$ (a subsequence of $\{a_n\}$) s.t. $a_{n_j} \to a'$.

For this convergent subsequence per se, given the arbitrary ε we have specified in the very beginning, we know that $\exists J > 0$ s.t.

$$|a_{n_j} - a'| < \frac{\varepsilon}{2}, \qquad \text{for all } j > J$$

Now, for arbitrary $n \in N$, we can always find a $k = n_i$ with $i > J$, such that $a_k = a_{n_i}$ is within $\frac{\varepsilon}{2}$ distance away from a'. Combining this fact with the construction that $a' \in (a - \frac{\varepsilon}{2}, a)$, it is clear the a_k we found specifically for ε and $n \in N$ satisfies: $a_k > a - \varepsilon$.

- Prove that (i) and (ii) implies that $\limsup a_n = a$.

To prove that $\limsup a_n = a$, we first show that a is the limit of a subsequence of $\{a_n\}$; then we show that $\nexists a' > a$ s.t. a' is the limit of a subsequence of $\{a_n\}$.

Firstly, by (i) and (ii), for arbitrary $\varepsilon > 0$, we can find a subsequence $\{a_{n_j}\}$ with certain $N \in \mathbb{N}$ such that $a - \varepsilon < a_{n_j} < a + \varepsilon$, $\forall n_j > N$. (Step 1: by (i), we can find a N^ε for arbitrary $\varepsilon > 0$, so that: $a_n < a + \varepsilon \ \forall n > N^\varepsilon$; Step 2, for the ε and all $\tilde{n} \geq N^\varepsilon$, we can find a $a_{k_{\tilde{n}}}$ s.t. $a - \varepsilon < a_{k_{\tilde{n}}}$. Thus, we have composed a subsequence $\{a_{k_{\tilde{n}}}\}$.)

Then, suppose $\exists a' > a$ as the limsup, then $\forall \varepsilon > 0 \ \exists N'$ s.t. $\forall n' > N'$, $|a_{n'} - a'| < \varepsilon$. However, (i) is violated when $\varepsilon < \frac{a'-a}{2}$: suppose that $a_{n_k} \to a'$. Then, $\exists N' > 0$ s.t. $\forall k > N'$, $|a_{n_k} - a'| < \varepsilon$. Given that $\varepsilon < \frac{a'-a}{2}$, there does not exist a N that may satisfy (i). (The "$\forall n > N$" statement is violated due to the subsequence that converges to a'.)

Alternatively, one can prove the statement using an equivalent definition of limsup:

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup_{k\geq n} a_k$$

Thus, (i) implies that $\forall \varepsilon > 0$, $\exists N$ s.t. $\forall n > N$:

$$\sup_{k\geq n} a_k < a + \varepsilon$$

Therefore, $\limsup_{n\to\infty} a_n < a + \varepsilon \iff \limsup_{n\to\infty} a_n \leq a$;

At the same time, (ii) implies that $\forall \varepsilon > 0$, for arbitrary $n \in N$, $\exists k > n$ s.t. $a_k > a - \varepsilon$. Then:

$$\sup_{j\geq n} a_j > a - \varepsilon$$

Therefore, $\limsup_{n\to\infty} a_n > a - \varepsilon \iff \limsup_{n\to\infty} a_n \geq a$; $\square$

Definition 2.18 (Infinite series)**.** Given a sequence $\{a_n\}$, let $s_n = \sum_{i=1}^{n} a_i$ be a sequence $\{s_n\}$, it is called **infinite series**. We write $\sum_{n=1}^{\infty} a_n = a$ if $\{s_n\}$ converges to a.

> **Example 4.** For $a_n = \frac{1}{2^n}$, we can obtain an expression for $\sum_{n=1}^{M} a_n$; and $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$.
> Also note that the sum of arbitrary segment of $\{\frac{1}{n}\}$ can be arbitrarily large if the length of such segment is long enough.

Definition 2.19 (Rearrangement)**.** $\{n_i\}_{i=1}^{\infty}$ is a sequence of natural numbers in which each natural number appears exactly once. Let $b_i = a_{n_i}$, then b_i is **a rearrangement** of $\{a_i\}_{i=1}^{\infty}$.

Definition 2.20 (Absolute convergence)**.** If $\sum_{n=1}^{\infty} |a_n|$ converges, we say that $\sum_{n=1}^{\infty} a_n$ converges absolutely. (e.g. for $a_n = (-1)^n \frac{1}{n}$, $\sum_{n=1}^{\infty} < \infty$, yet $\sum_{n=1}^{\infty} |a_n| \to \infty$.)

Proposition 2.21. *If $\sum_{n=1}^{\infty} a_n$ converges absolutely, then $\sum_{n=1}^{\infty} b_i = \sum_{n=1}^{\infty} a_n$, where $\{b_n\}$ is a rearrangement of $\{a_n\}$.*

Note that, rearranging $\{(-1)^n\}_{n=1}^{\infty}$ can give raise to arbitrary partial sum $\in \mathbb{Z}$.

Review: subsets in $\mathbb{R}$ Epistemic-wise, we established the construction of following sets sequentially:

1. $\mathbb{N}$: The set of natural number; [It is countable.]

2. $\mathbb{Z}$: The set of integers; [It is also countable. In fact, $|\mathbb{N}| = |\mathbb{Z}|$.

3. $\mathbb{Q}$: The set of rational number; [It is also countable, and **dense**.]

4. $\mathbb{R}$: The real line. [Completeness Axiom]

Definition 2.22 (Principle of Mathematical Induction)**.** The set of natual numbers is the smallest set that satisfies the axiom of Mathematical Induction.

Example 5. Prove that $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$.

Proof method: To prove by induction:

- When $n = 1$, LHS = RHS;
- Suppose that, for some $n_0 \in \mathbb{N}$, LSH = RHS $\forall n \in N$ s.t. $n \leq n_0$, then we show that LHS = RHS for $n = n_0 + 1$.

$\square$

2.2 Real Value Functions

Definition 2.1. A real valued function defined on X (an arbitrary set) is represented as following, with $\mathbb{R}$ as the codomain:

$$f : x \to \mathbb{R}$$

Notation 2.2. For $a \in \mathbb{R}$ and f, g being real value functions, "$=, \geq, >, \gg$, function addition and (scalar) multiplcation" are defined as follows:

- If $f(x) = a \ \forall x \in X$, we write $f = a$;
- If $f(x) \geq g(x) \ \forall x \in X$, we write $f \geq g$;
- If $f \geq g$, but not the other way, then $f > g$. ($f(x) = g(x)$ is permissible for some $x \in X$).
- If $f(x) > g(x) \ \forall x \in X$, then we write $f \gg g$.
- $(f + g)(x) \coloneqq f(x) + g(x)$;
- $(a \cdot f)(x) \coloneqq a \cdot f(x)$;
- $(f \cdot g)(x) \coloneqq f(x) \cdot g(x)$.

Note that, $f > g$ is a "weakly hight" relationship.

Definition 2.3 (strictly/weakly increasing/decreasing)**.** Construction is intuitive and thereby omitted.

Definition 2.4 (Limit of function)**.** A function $f : x \to \mathbb{R}$ converges to $a \in \mathbb{R}$ as x approaches some $x_0 \in X$ if

$$\forall \varepsilon > 0,\ \exists \delta > 0 \text{ s.t. } \forall x \in (x_0 - \delta, x_0 + \delta)$$
$$|f(x) - a| < \varepsilon$$

in which case we write $\lim_{x \to x_0} f(x) = a$.

Definition 2.5 (Right limit)**.** A function $f : x \to \mathbb{R}$ converges to $a \in \mathbb{R}$ from right as x approaches x_0 if

$$\forall \varepsilon > 0,\ \exists \delta > 0 \text{ s.t. } \forall x \in (x_0, x_0 + \delta),$$
$$|f(x) - a| < \varepsilon$$

We write the right limit as: $\lim_{x \to x_0^+} f(x) = a$.

Proposition 2.6. Suppose $f : X \to \mathbb{R}$ and $g : X \to \mathbb{R}$, with $\lim_{x \to x_0} f(x) = a$ and $\lim_{x \to x_0} g(x) = b$.

(i) $\lim_{x \to x_0} f(x) \pm g(x) = a + b$;

(ii) $\lim_{x \to x_0} f(x) \cdot g(x) = a \cdot b$;

(iii) $\lim_{x \to x_0} \frac{f(x)}{g(x)} = \frac{a}{b}$ if $g \neq 0$ and $b \neq 0$.

Definition 2.7 (Continuity)**.** A function $f : X \to \mathbb{R}$ is continuous at $x_0 \in X$ if

$$\lim_{x \to x_0} f(x) = f(x_0)$$

Note that, we can draw definition of *limit of function* to formalize an $\varepsilon - \delta$ argument that defines a continuous function.

3 Lecture 3: Linear Space and $f : X \to \mathbb{R}$, continued

3.1 Linear spaces and linear algebra

Definition 3.1 (Vector Space)**. Vector space** V over a field F is a set V together with *vector addition* and *scalar multiplication.*

- A field F is a set with addition and multiplication operation defined among its own elements.

  Example: $\mathbb{R}$ with normal $+$ and $\cdot$ is a field, denoted as: "$F : \mathbb{R}, +, \cdot$".

  Formally, a field is also established using a set of axiom. Note that field is "equipped with": $0, (-1)$ elements.

Axiomatically, $\forall u, v, w \in V$ and $a, b \in F$, the following shall be satisfied:

Axiom 1 $u + (v + w) = (u + v) + w$;

Axiom 2 $u + v = v + u$;

Axiom 3 $\exists \theta \in V$ s.t. $u + \theta = u$;

Axiom 4 $\exists \phi(u) \in V$ s.t. $u + \phi(u) = \theta$;

Axiom 5 $a \cdot (u + v) = a \cdot u + a \cdots v$;

Axiom 6 $(a + b) \cdot u = a \cdot u + b \cdot u$;

Axiom 7 $a \cdot (b \cdot u) = b \cdot (a \cdot u)$;

Axiom 8 V is closed under vector addition and scalar multiplication;

Axiom 9 $1 \cdot u = u$, where 1 is the identity in F.

Note that, the last axiom was not stated in lecture.

Proposition 3.2. Using the axioms, we can show the following equalities hold:

1. $0 \cdot u = \theta$;
2. $\phi(u) = (-1) \cdot u$;
3. $a\theta = \theta$;
4. θ is unique.

Proof. Relies heavily on algebraic tricks. Omitted as of 2015-08-29 15:01:15. □

Example 6 (Example for vector spaces). 1. $V = \mathbb{R}^n$ and $F : R, +, \cdot$;

2. $V = \{ax^2 + bx + c : a, b, c \in \mathbb{R}, x \in [0, 1]\}$, for $F : \mathbb{R}, +, \cdot$.

Definition 3.3. A vector space cna also be called a linear space.

Definition 3.4 (Linear subspace)**.** For V being a linear space and $U \subseteq V$, if U itself is a linear space with *the same* vector additions and scalar multiplication, then we say U is a **linear subspace of** V.

Note that, this definition admits the case where $U = V$, i.e. though trivially, V is a linear subspace of itself.

3.1.1 *Finite* Linear combination, span and linear independence of vectors

From now on, we limit the discussion to the following case:

1. Adopt $\mathbb{R}$ with normal addition and multiplication to be the field F;
2. Consider only finite operations when defining linear combination and span;
3. Note that: it is still permissible for V to be an arbitrary set.

Definition 3.5 (Linear Combination)**.** For $U \subseteq V$,

(i) If $U = \{v_1, \ldots, v_n\}$ for some $n \in \mathbb{N}$, i.e. U is a finite subset of V, then a linear combination of U is a new vector:

$$v = \sum_{i=1}^{n} a_i v_i,\ a_i \in \mathbb{R},\ \ i = 1, \ldots, n$$

(ii) If U is no longer finite, regardless of whether is is countably infinite or uncountable, a **linear combination of** U is a vector that is *a linear combination of finitely many vector of* U.

Definition 3.6 (span of a set of vectors)**.** For $A = \{v_1, \ldots, v_n\}$,

$$span(A) = \{\sum_{i=1}^{n} a_i v_i : a_i \in \mathbb{R},\ i = 1, \ldots, n\}$$

Proposition 3.7. The span of any $U \subset V$ is a linear subspace of linear space V.

Scatch of proof. Note that, by construction of $span(A)$, arbitrary coefficient is allows. Letting all coefficients to be 0 gives raise to the θ; other properties may follow from standard algebra in $\mathbb{R}$ (the field). ☐

Definition 3.8 (Linear independence)**.** A (finite) set of vectors A is linearly independent if $\nexists v \in A$ can be written as linear combinations of the others. Formally,

$$A = \{v_1, \ldots, v_n\} \text{ is linearly independent if } \sum_{i=1}^{n} a_i v_i = \varepsilon \implies a_i = 0 \forall i$$

Proof. Suppose not, that is $\sum_{i=1}^{n} a_i v_i = \varepsilon$ yet $a_j \neq 0$ for some j, then we can write:

$$-a_j v_j = \sum_{k \neq j} a_k v_k$$

where, upon simplification, v_j could be written as a linear combination of the other vectors. ☐

Proposition 3.9. For $A \subseteq V$, $span(A)$ is the smallest linear space that contains A.

Alternatively, one can define $span(A)$ to be the intersection of all linear subspaces of V that contains A.

Definition 3.10 (base and dimension of V)**.** If $\{v_1, \ldots, v_n\}$ are linearly independent and $span(\{v_1, \ldots, v_n\} = V$ (the linear space), then $\{v_1, \ldots, v_n\}$ is called a **base of** V.

In this case, the **dimension** of V is $dim(V) = n$.

Theorem 3.11. If A and B are two bases of V and A, B are finite, then $|A| = |B|$.

Idea of the proof. Suppose $A = \{u_1, u_2\}$ and $B = \{v\}$. Then one can write:

$$u_1 = av; \qquad u_2 = bv \text{ for some } a, b \in \mathbb{R}$$

Therefore, u_1 and u_2 are not linearly independent.

□

Note that, it seems to me that the finiteness assumption only serves the need of simplifying the proof.

3.1.2 Matrix

Definition 3.12. A $m \times n$ matrix could be written as:

$$A = \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} = [u_1 \ldots u_n]$$

where v_i is a $1 \times n$ (row) vector, and u_i is a $m \times 1$ (column) vector.

Definition 3.13 (Rank of a matrix)**.** The *maximum number* of linearly independent row/column vectors denotes the rank of a matrix.

Comment: implicitly, by definition, $rank(A) = rank(A^T)$.

Definition 3.13 (Linear transformation)**.** $T : U \to V$ is a linear transformation if

$$T(au_1 + bu_2) = aT(u_1) + bT(u_2), \qquad \forall a, b \in \mathbb{R}$$

Remark 3.14. A $m \times n$ matrix A is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$.

3.2 Real-Valued functions continued

3.2.1 Continuity and its corollaries

Definition 3.1 (Interval)**.** An interval of $\mathbb{R}$ is either $[a, b]$, $(a, b]$, $[a, b)$ or (a, b); where $a, b \in \mathbb{R} \bigcup \{+\infty, -\infty\}$ (the extended real line).

Theorem 3.2 (Intermediate Value Theorem)**.** If I is an interval of $\mathbb{R}$ [1], and $f : I \to \mathbb{R}$ is continuous, then $f(I)$ is also an interval of $\mathbb{R}$[2].

[1] I could be a connected set in Euclidean space ($\mathbb{R}^n$).

[2] Correspondingly, $f(I)$ would be a connected set.

Proposition 3.3. If f is continuous and bijective (thus invertible, i.e. f^{-1} is a function), ten f is either strictly increasing or strictly decreasing.

Note that:

- Continuity forced bijections to be monotone;
- "Strictness" is used to support bijection;
- A stronger statement (yet correct) goes as follows:

  Let I and J be both intervals, then $f : I \to J$ is continuous and bijective if and only if it is strictly monotonic.

Theorem 3.4 (Extreme Value Theorem)**.** If $f : [a, b] \to \mathbb{R}$ is continuous, then $\exists x_1, x_2 \in [a, b]$ s.t.

$$f(x_1) = \sup f([a, b])$$
$$f(x_2) = \inf f([a, b])$$

Comment: using max and inf in the statement would be more precise though.

Definition 3.5 (Uniformly continuity)**.** $f : x \to \mathbb{R}$ is said to be uniformly continuous if $\forall \varepsilon > 0$, $\exists \delta > 0$, s.t.

$$|f(x) - f(y)| < \varepsilon, \qquad \forall |x - y| < \delta$$

Note that:

1. We no longer specify a certain point $x_0 \in X$;
2. Instead, the δ applies to all $x, y \in X$ as long as they are within δ distance away.

Exercise 3.6. Prove that $f(x) = \frac{1}{x}$ $(x > 0)$ is not uniformly continuous.

Professor's Proof. Without loss of generality, suppose $\varepsilon = \frac{1}{2}$. Now we want ot show that $\nexists$ $\delta > 0$ s.t. if $|x - y| < \delta$, $|f(x) - f(y)| < \frac{1}{2}$.

By the property of $f(x)$, we look for a threshold $z^*(\varepsilon, \delta)$ at which:

$$\left| \frac{1}{z^*} - \frac{1}{z^* + \delta} \right| = \frac{1}{2}$$

Then, for arbitrary $\delta > 0$, write $z^* = z^*(\varepsilon, \delta)$, we have:

$$|f(z') - f(z' + \delta)| > \frac{1}{2}, \qquad \forall z' < z^*$$

Thus, we see that for $\varepsilon = \frac{1}{2}$, there does not exist a $\delta > 0$ that satisfies $|f(x) - f(y)| < \frac{1}{2}$ $\forall |x - y| < \delta$. $\square$

Comment: in professor's proof, there is a flaw: choosing z' and $z' + \delta$ won't help disprove the original statement. This could easily be fixed as shown in the alternative proof.

Alternative proof. Without loss of generality, suppose $\varepsilon = \frac{1}{2}$. Now we want ot show that $\nexists$ $\delta > 0$ s.t. if $|x - y| < \delta$, $|f(x) - f(y)| < \frac{1}{2}$.

By the property of $f(x)$, we look for a threshold $z^*(\varepsilon, \delta)$ at which:

$$\left| \frac{1}{z^*} - \frac{1}{z^* + \frac{\delta}{2}} \right| = \frac{1}{2}$$

Then, for arbitrary $\delta > 0$, write $z^* = z^*(\varepsilon, \delta)$, we have:

$$|f(z') - f(z' + \frac{\delta}{2})| > \frac{1}{2}, \qquad \forall z' < z^*$$

Thus, we see that for $\varepsilon = \frac{1}{2}$, there does not exist a $\delta > 0$ that satisfies $|f(x) - f(y)| < \frac{1}{2}$ $\forall |x - y| < \delta$.

Note that, it is the highlighted condition that has been disproved.

□

3.3 Differentiation

Remark 3.1. "Differentiation" is essentially a process of taking linear approximation.

Definition 3.2 (Tangent line)**.** The tangent line to a function $y = f(x)$ at the point $(x_0, f(x_0))$, when exists, is the line through $(x_0, f(x_0)$ with slope

$$\alpha = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

When α exists, the tangent line exists. It could be written as:

$$y = f(x_0) + \alpha(x - x_0)$$

Definition 3.3 (Differentiation)**.** The **derivative of** $f : x \to \mathbb{R}$ **at** $x_0 \in X$ is

$$f'(x) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

We may also write the derivative as:

$$\left(\frac{df(x)}{dx} \bigg|_{x = x_0} \right)$$

The **derivative of** f is denoted by

$$\frac{df(x)}{dx}$$

Remark 3.4 (Properties of derivatives)**.** For f, g as functions and $a, b \in \mathbb{R}$:

(i) $(af + bg)' = af' + bg'$

(ii) $(fg)' = f'g + fg'$

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

Proposition 3.5 (Chain Rule)**.** If $g : x \to \mathbb{R}$ is differentiable at $x_0 \in X$ and $f : Y \to \mathbb{R}$ is differentiable at $g(x_0) \in Y$, then $f(g(x))$ is differentiable at $x = x_0$, we write:

$$\left(\frac{df(g(x))}{dx} \bigg|_{x=x_0} \right) = f'(g(x_0))g'(x_0)$$

Proposition 3.6 (Inverse function theorem)**.** If $f : X \to Y$ is bijective, then derivative of $f^{-1} : Y \to X$ is

$$\frac{df^{-1}(y)}{dy} = \frac{1}{f'(f^{-1}(y))}$$

Proof. Since f is bijective function, we have: $f(f^{-1}(y)) = y$. Differentiating w.r.t.[3] y gives:

$$f'(f^{-1}(y)) \cdot \frac{df^{-1}(y)}{dy} = 1 \implies \frac{df^{-1}(y)}{dy} = \frac{1}{f'(f^{-1}(y))}$$

$\square$

Definition 3.7 (local maximum)**.** Function $f : X \to \mathbb{R}$ has a local maximum at $x_0 \in X$ if $\exists \delta > 0$, s.t.

$$f(x_0) \geq f(x), \qquad \forall x \in \{x \in X : |x - x_0| < \delta\}$$

Proposition 3.8 (Condition for interior local maximum)**.** If $f : X \to \mathbb{R}$ is differentiable, and has a local maximum at *an interor point* $x = x_0$ [4], then $f'(x_0) = 0$

Proof. First consider the right limit: $\lim\limits_{\Delta \to 0^+} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$. For $\Delta x > 0$ and $f(x_0 + \Delta x) - f(x_0) \leq 0$ (by local maximum), we see:

$$\lim_{\Delta \to 0^+} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \leq 0$$

In similar spirit, we conclude that:

$$\lim_{\Delta \to 0^-} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \geq 0$$

Thus we conclude that $f'(x_0) = 0$ due to differentiability of f at x_0.

$\square$

Note that, if x_0 is at the boundary of X, whether this proposition holds (or not) depends on how we define the derivative at the boundary point.

Theorem 3.9 (Rolle's Theorem)**.** If $f : [a, b] \to \mathbb{R}$ is differentiable and $f(a) = f(b) = 0$, then $\exists x_0 \in (a, b)$ s.t. $f'(x_0) = 0$.

---

[3] with respect to

[4] x_0 is an interior point of X, i.e. $\exists \delta > 0$ s.t. $\{x \in X : |x - x_0| < \delta\} \subseteq X$

Proof. Since $f : [a, b] \to \mathbb{R}$ is differentiable and hence continuous, if $\sup f([a, b]) > 0$, then we can locate a x_0 as local maximum. Then, by the previous proposition, $f'(x_0) = 0$;

Alternatively, if $\inf(f[a, b]) < 0$, we can find a x_1 as local minimum. This also gives raise that $f'(x_1) = 0$.

Otherwise, f is flat, and $f'(x) = 0\ \forall x \in [a, b]$. $\square$

Note that, this is like reaching a plateau/basin when leaving at sea-level and reaching another point at sea-level.

Theorem 3.10 (Mean Value Theorem)**.** If $f : [a, b] \to \mathbb{R}$ is differentiable, then $\exists x_0 \in (a, b)$ s.t.

$$f'(x_0) = \frac{f(b) - f(a)}{b - a}$$

Proof. Subtract a line function: $y = f(a) - \frac{f(b)-f(a)}{b-a}(x - a)$ from $f(x)$ to get $g(x)$, we can apply Rolle's Theorem and find a x_0 that satisfies $g'(x_0) = 0$.

Note that, one can envision subtracting a line-function as a transformation of coordinate system. $\square$

Theorem 3.11 (Generalized Mean Value Theorem)**.** Let $f : [a, b] \to \mathbb{R}$, $g : [a, b] \to \mathbb{R}$ be both differentiable, then $\exists x_0 \in [a, b]$ s.t.

$$g'(x_0)(f(b) - f(a)) = f'(x_0)(g(b) - g(a))$$

Note that, we can rationalize this theorem as: the ratio of average speed shall equal the ratio of travel speed at some point of time[5].

Theorem 3.12 (L'Hopital Rule)**.** f and g are differentiable, with $g'(x) \neq 0$, $\forall x \in X$. Suppose $\lim\limits_{x\to x_0} \frac{f'(x)}{g'(x)} = q$, then if either of the following conditions is satisfied, $\lim\limits_{x\to x_0} \frac{f(x)}{g(x)} = q$.

(i) If $f(x), g(x) \to 0$ as $x \to x_0$;

(ii) If $f(x), g(x) \to \infty$ as $x \to x_0$.

Proposition 3.13 (Derivative is continuous at x_0)**.** If $\lim\limits_{x\to x_0} f'(x)$ exists, then $f'(x_0) = \lim\limits_{x\to x_0} f'(x)$.

Proof. Define $h(x) = f(x) - f(x_0)$, we see that $h(x) \to 0$ as $x \to x_0$; then, define $g(x) = x - x_0$, we also see that $g(x) \to 0$ as $x \to x_0$.

Thus, by L' Hopital's rule, we have:

$$\lim_{x\to x_0} \frac{h(x)}{g(x)} = \lim_{x\to x_0} \frac{h'(x)}{g'(x)} = \frac{f'(x)}{1} = f'(x)$$

Note that, what we started with is by definition $f'(x_0) = \lim\limits_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0}$. So, we are done. $\square$

[5]Though, Prof Ke did not specify which one is the "time variable".

4 Lecture 4: Differentiation and Linear Algebra

4.1 Differentiation

Definition 4.1 (derivative)**.** If f has a derivative f', and f' itsef is also differnetiable, then we write the derivative of f', f''. The we also derive the other higher order derivatives:

$$f', f'', f^{(3)}, f^{(4)}, \dots, \frac{d^n f(x)}{dx^n} = f^{(n)}$$

Theorem 4.2. For $f : [a, b] \to \mathbb{R}$, suppose $f^{(n-1)}$ is continuous (i.e. $f \in C^{n-1}$), and $f^{(n)}$ exists. Then for $x_0, \bar{x} \in [a, b]$, $\exists \tilde{x} \in (\min\{x_0, \bar{x}\}, \max\{x_n, \bar{x}\})$ s.t.

$$\begin{aligned} f(\bar{x}) =& f(x_0) + f'(x_0)(\bar{x} - x_0) + \frac{1}{2}f''(x_0)(\bar{x} - x_0)^2 + \frac{1}{3!}f^{(3)}(x_0)(\bar{x} - x_0)^3 \\ &+ \ldots + \frac{1}{(n-1)!}f^{(n-1)}(x_0)(\bar{x} - x_0)^{n-1} + \frac{1}{n!}f^{(n)}(\tilde{x})(\bar{x} - x_0)^n \end{aligned}$$

Alternatively, one can write:

$$f(\bar{x}) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(\bar{x} - x_0)^k + \frac{1}{n!}f^{(n)}(\tilde{x})(\bar{x} - x_0)^n$$

Proof. This is a direct proof, which relies on Mean Value Theorem and a tricky construction. $\square$

Remark 4.4 (Limitation of Taylor series expansion)**.** If $f^{(n)}$ exists for all $n \in \mathbb{N}$, can we write down the Taylor series as follows?

$$f(\bar{x}) \stackrel{?}{=} \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(\bar{x} - x_0)^k$$

It turns out that:

(i) RHS may not necessarily converge: for $x_0 = 0$, $f(x) = \frac{1}{1-x} \implies RHS = \sum_{k=0}^{\infty} x^k$. If $|x| > 1$, RHS does not converge.

(ii) RHS converges but LHS $\neq$ RHS:

Let $e^{-\frac{1}{0}} = 0$, then $e^{-\frac{1}{x^2}}$ has a Taylor series expansion at $x_0 = 0$, yet RHS $= 0$.

(iii) The equality may hold, for example:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \implies \text{LHS} \ = \ \text{RHS} \ , \ \forall x$$

4.1.1 Multi-variable Calculus

Definition 4.5 (Differentiable multi-variate vector-valued function)**.** For $f : X \to \mathbb{R}^m$, $X \subseteq \mathbb{R}^n$, we say that f **is differentiable at** $x_0 \in X$ if $\exists A$ (m by n matrix) s.t.

$$\lim_{x \to x_0} \frac{||f(x) - [f(x_0) + A \cdot (x - x_0)]||}{||x - x_0||} = 0$$

where the norm of x is defined as: $||x|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$.

Note that, $f(x_0) + A \cdot (x - x_0)$ is a local approximation of $f(x)$ at x_0. Here, $f(x_0)$ is a $m \times 1$ vector, $x - x_0$ is a $n \times 1$ vector, and A is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Exercise: L'Hopital's Rule

- $\lim\limits_{x \to 0} \frac{e^x - 1}{x} = \lim\limits_{x \to 0} \frac{e^x}{1} = 1$ (By L'Hopital's Rule).
- $\lim\limits_{x \to 0} x \cdot ln(1 + \frac{1}{x}) = \lim\limits_{x \to 0} \frac{ln(1+\frac{1}{x}}{\frac{1}{x}} = \lim\limits_{x \to 0} \frac{\frac{1}{1+\frac{1}{x}} \cdot (-\frac{1}{x^2})}{-\frac{1}{x^2}} = \lim\limits_{x \to 0} \frac{1}{1+\frac{1}{x}} = 0$ (Again, by L'Hopital's Rule)

4.1.2 Partial Derivatives

Definition 4.6 (Partial Derivatives)**.** For $x : x \to \mathbb{R}$ with $x \in \mathbb{R}^n$, the partial derivative of f at $x = (x_1, \ldots, x_n)$ with respect to x_i is defined as:

$$\frac{\partial f(x)}{\partial x_i} := f_i(x) = \lim_{\lambda \to 0} \frac{f(x + \lambda e_i) - f(x)}{\lambda}$$

where $e_i = [\underbrace{0, \ldots, 0}_{i-1}, 1, 0, \ldots]^T \in \mathbb{R}^n$ (1 is at the i-th element).

Definition 4.7 ($\partial x_i \partial x_j$)**.** If we take the partial derivative of $\frac{\partial f(x)}{\partial x_i}$ with respect to x_j, we have:

$$\frac{\partial \left(\frac{\partial f(x)}{\partial x_i} \right)}{\partial x_j} := \frac{\partial^2 f(x)}{\partial x_i \cdot \partial x_j} := f_{ij}$$

Proposition 4.8. If f_i, f_j and f_{ij} exists, and f_{ij} is continuous at $x_0 \in x$, then f_{ji} exists at x_0 and $f_{ij} = f_{ji}$.

4.1.3 Gradient

Definition 4.9 (Gradient ∇)**.** Remember the matrix A we defined in Definition 4.5? Let $A = \nabla f(x)$, then $\nabla f(x)$ is called the gradient of f at x, i.e. $\nabla f(x)$ is a $m \times n$ matrix s.t.

$$\lim_{x \to x_0} \frac{||f(x) - (f(x_0) + \nabla f(x_0)(x - x_0))\,||}{||x - x_0||}$$

It turns out that we are essentially defining $\nabla f(x_0)$ as a $m \times n$ matrix without transpose? Note: $\nabla f(x_0) = [f_1(x_0), f_2(x_0), \ldots, f_n(x_n)]$. (Simon and Blume page 321 defined gradient to be a $n \times m$ matrix.

Definition 4.10 (Inner product)**.** In $\mathbb{R}^2$, the inner product of (x_1, y_1) and (x_2, y_2) is

$$< (x_1, y_1), (x_2, y_2) > \quad := \quad (x_1, y_1) \cdot \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = x_1 x_2 + y_1 y_2$$

Additionally, one can define the norm as $||(x, y)|| = \sqrt{< (x, y), (x, y) >}$. Further more, one can write:

$$< (x_1, y_1), (x_2, y_2) >= ||(x_1, y_1)|| \cdot ||(x_2, y_2)|| \cdot \cos\theta$$

where θ is the angle from vector (x_1, y_1) to (x_2, y_2).

Remark 4.11 (Tangential plane example)**.** For $f : \mathbb{R}^2 \to \mathbb{R}$, let $X \times Y = \mathbb{R}^2$ be the domain and $Z = \mathbb{R}$ be the codomain. Then, we can define a tangential plane at $x_0 = (x', y')$ as follows:

- Fix the y component, we can define the following plane that is parallel to $Z - X$ plane:

$$z = f(x_0) + \frac{\partial f(x_0)}{\partial x}(x - x')$$

- Fix the x component, we can define the following plane that is parallel to $Z - Y$ plane:

$$z = f(x_0) + \frac{\partial f(x_0)}{\partial y}(y - y')$$

Then, the tangential plane is defined as:

$$f(x_0) + \left(\frac{\partial f(x_0)}{\partial x}, \frac{\partial f(x_0)}{\partial y} \right) \cdot (x - x_0)$$

Projection (Continued from the previous $f : \mathbb{R}^2 \to \mathbb{R}$ function) In $X \times Y$, we can draw a set of contour-lines at which $f(x)$ achieves the same Z-value.

Under the projection, the tangential plane we defined previous now becomes a tangent line. More importantly, the projection of *gradient vector* is orthogonal to such tangent line in $X - Y$.

In general, the gradient is also orthogonal to the tangent plane.

Proposition 4.12. The gradient is the direction along which $f(x)$ increases the fastest.

Implication of gradient Formally, Remark 4.11 could be stated as follows:

- $\frac{\partial f(\bar{x})}{\partial x_i}$ indicates the slope of the function f restricted to the subset of $\mathbb{R}^n$ s.t.

$$x_j = \bar{x}_j, \qquad \forall j \neq i$$

  Therefore, the vector of the slopes $\left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right)$ denotes the tangent plane's "slope".

- The gradient $\left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \right)$ points out the direction in $\mathbb{R}^n$ along which f increases the fastest.

Example interpreting properties of gradient: $z = 2x + y$

1. Plot it in $x - y - z$;
2. Calculate the gradient at $(1, 2)$ as: $\nabla f(1, 2) = (2, 1)$ through definition.
3. Demonstrate the solution to the following optimization problem

$$\lim_{\Delta x, \Delta y} 2(\bar{x} + \Delta x) + (\bar{y} + \Delta y)$$

$$s.t. ||(\Delta x, \Delta y)|| = 1$$

4.2 Vector & Matrix Differentiation

Definition 4.1. For $f : \mathbb{R}^n \to \mathbb{R}^m$, write $f(x) = \begin{bmatrix} f^{(1)}(x) \\ \vdots \\ f^{(m)}(x) \end{bmatrix}$. Define the derivative of f as $f'(x) = Df(x) = D \begin{bmatrix} f^{(1)}(x) \\ \vdots \\ f^{(m)}(x) \end{bmatrix} = \begin{bmatrix} f_1^{(1)}(x) & \cdots & f_n^{(1)}(x) \\ \vdots & \ddots & \vdots \\ f_1^{(m)}(x) & \cdots & f_n^{(m)}(x) \end{bmatrix}$, where $f'(x)$ is defined to be a $m \times n$ matrix.

Convention here is different from what we used for multi-variable single-valued functions, where $f^{(n)} := \frac{d^n f}{dx^n}$ for $f : \mathbb{R}^n \to \mathbb{R}$. **Here**, $f^{(m)}(x)$ is the m-th in the codomain.

Example: for $f(x, y) = \begin{bmatrix} 2x + y \\ 3x^2 \end{bmatrix}$, then $Df(x, y) = \begin{bmatrix} 2 & 1 \\ 6x & 0 \end{bmatrix}$.

Definition 4.2 (Higher order (partial) derivatives for single-valued function)**.** The Higher order (partial) derivatives for $f : \mathbb{R}^n \to \mathbb{R}$ is a $1 \times n$ vector defined as:

$$\frac{df}{dx} = Df = [f_1(x), \cdots, f_n(x)].$$

For the second order derivative:

$$\frac{d}{dx}\left(\frac{df}{dx}\right) = D^2 f = \begin{bmatrix} f_{11}(x) & \cdots & f_{1n}(x) \\ \vdots & \ddots & \vdots \\ f_{n1}(x) & \cdots & f_{nn}(x) \end{bmatrix} = \frac{\partial^2 f}{\partial x \partial x^T}$$

Here, x is a vector, and $\partial x \partial x^T \simeq x^2$ in $\mathbb{R}^1$, where $\simeq$ is defined in terms of equivalent expression. **Note that,** the goal here is to let $f' \cdot x$ yield a scalar, instead of a matrix.

Additionally, for $f : \mathbb{R}^{m \times n} \to \mathbb{R}$, suppose $A \in \mathbb{R}^{m \times n}$, then:

$$Df(A) := \begin{bmatrix} \frac{f(A)}{\partial A_{11}} & \cdots & \frac{\partial f(A)}{\partial A_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(A)}{\partial A_{m1}} & \cdots & \frac{\partial f(A)}{\partial A_{m \times n}} \end{bmatrix}$$

Note it here that $\mathbb{R}^{2 \times 2} \neq \mathbb{R}^4$, in terms of expression of the elements in such Euclidean space. However, the dimensionality of the two spaces are the same, thereby there exists a bijection between $\mathbb{R}^{2 \times 2}$ and $\mathbb{R}^4$.

Exercise 4.3. (Example: for taking derivative of composite functions.)

1. $f(x, y)$ is a real-valued function, and $x = u + \log v$ and $y = u - \log v$. Show that:

$$\frac{\partial^2 f}{\partial u^2} = \frac{\partial^2 f}{\partial x^2} + 2\frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2}$$

Proof. Write first taht $f(x(u, v), y(u, v))$ where $x(u, v) = u + \log v$; $y(u, v) = u - \log v$.

$$\frac{\partial f(x(u, v), y(u, v))}{\partial u} = f_x \frac{\partial x}{\partial u} + f_y \frac{\partial y}{\partial u} = f_x + f_y$$

(note that, $\frac{\partial x}{\partial u} = 1$ and $\frac{\partial y}{\partial u} = 1$, by evaluating the functional form.)
Now, take this simplified functional form for $\frac{\partial f}{\partial u} = f_x + f_y$, look into:

$$\frac{\partial^2 f}{\partial u^2} = f_{xx} \cdot \frac{\partial x}{\partial u} + f_{xy} \cdot \frac{\partial y}{\partial u} + f_{yx} \cdot \frac{\partial x}{\partial u} + f_{yy} \cdot \frac{\partial y}{\partial u}$$

Simplify this by evaluating $\frac{\partial x}{\partial u} = 1 = \frac{\partial y}{\partial u}$ and get the expression we wanted to show. □

2. $f(x, y) = g(\frac{x}{y})$, g is differentiable, show that:

$$x \cdot \frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y}$$

Proof. Take the original equality and differentiated w.r.t x and y, we have:

$$\frac{\partial f}{\partial x} = g' \frac{\partial \left(\frac{x}{y}\right)}{\partial x} = \frac{1}{y} \cdot g'$$

$$\frac{\partial f}{\partial y} = g' \frac{\partial \left(\frac{x}{y}\right)}{\partial y} = \frac{x}{-y^2} g'$$

□

Definition 4.4 (Derivative of a inner product)**.** For matrices and vectors that are made of real numbers, inner product is defined as: $a^T \cdot x = \sum_{i=1}^{n} a_i x_i$, where $x = (x_1, \ldots, x_n)^T$ (as a $n \times 1$ (column) vector); and $a = (a_1, \ldots, a_n)^T$ (as a $n \times 1$ (column) vector). Then, the derivative w.r.t to x (a vector) is defined as:

$$\frac{\partial a^T \cdot x}{\partial x} = a$$

Definition 4.5 (Derivative for Ax)**.** Let A be a $m \times n$ matrix, then

$$\frac{\partial (Ax)}{\partial x} = A^T, \text{ simillarly, we have: } \frac{\partial (Ax)}{\partial x^T} = A$$

Definition 4.6 (Derivative for x^TAx)**.** Let A be a $n \times n$ matrix, and $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, then:

$$\frac{\partial\left(x^TAx\right)}{\partial x} = (A + A^T)x$$

On the other hand, we write:

$$\frac{\partial\left(x^TAx\right)}{\partial A} = xx^T \quad \Leftarrow n \times n \text{ square matrix.}$$

Remember that x is defined to be a $n \times 1$ vector, upon differentiating w.r.t. a square matrix, we get a square matrix in return. (Though, it is yet unclear what each element in the matrix shall mean upon this operation.)

4.3 Matrix

Definition 4.1 (Inverse of a matrix)**.** Let A be a $n \times n$ matrix, then A is invertible if $\exists B$ (also a $n \times n$ matrix) s.t.

$$AB = BA = I_n, \text{ where } I_n = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}_{n \times n}$$

Facts 4.2. If A is a square matrix, then $AB = I_n \implies BA = I_n$.

Idea of the proof. Bijective and therefore invertible.

- $\exists$ left inverse $\iff$ injective;
- $\exists$ right inverse $\iff$ surjective;
- $BA = I_n \implies A$ is injective $\implies \dim(\ker(A)) = 0$.
- The codomain $V = \mathbb{R}^n \implies \dim(V) = \dim(\text{im}A) + \dim(\ker(A))$ where "im$A$" is the image of A.

$\square$

Definition 4.3 (Orthogonal Matrix)**.** An orthogonal matrix is an $n \times n$ matrix A s.t. $AA^T = A^TA = I_n$, where $A^{-1} = A^T$. It has the following properties:

(1) Let $A = [u_1, \cdots, u_n]$ where u_i is an $n \times 1$ (column) vector. Then, each column vector has norm 1, and for $i \neq j$, u_i & u_j are orthogonal. That is:

$$u_i^Tu_j =< u_i, u_j >= 0, \quad \forall i \neq j \text{ and } ||u_i|| = \sqrt{< u_i, u_i >} = 1$$

(2) Let $A = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ where v_i is a $1 \times n$ (row) vector. Similarly, we have:

$$v_iv_j^T =< v_i, v_j >= 0, \quad \forall i \neq j \text{ and } ||v_i|| = \sqrt{< v_i, v_i >} = 1$$

Definition 4.4 (Determinant)**.** For $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, its determinant $\det(A)$ denotes the area enclosed by the two column vectors, when letting them shooting out from origin.

Proposition 4.5. *A is invertible if and only if* $\operatorname{rank}(A) = n$*, or* $\det(A) := |A| \neq 0$*.*

Exam does require calculating the determinant of a 3×3 matrix.

Remark 4.6 (Operation of matrixes as transformation, and Eigenvalues and Eigenvectors)**.** A matrix can realize any one of the following operations (casted upon vectors of same length), as a linear transformation:

- rotate
- shrink/extend
- reflection
- shifting (moving around)

The "eigenvector", if exists, specifies one of the directions along which the matrix does only the job of shrinking/extending. On this very direction, imposing the linear transformation on a vector will shrink/extend the vector by a factor of λ, the eigenvalue.

Definition 4.7 (Eigenvalue and Eigenvectors)**.** For a $n \times n$ matrix, if there exists $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ s.t.

$$Av = \lambda v$$

then v is called a right-eigenvector of A, associated with eigenvalue λ.

If $v^T A = \lambda v^T$, then v is called the left eigenvector of A associated with eigenvalue λ. Note that, by definition, for the λ and v, we can also write $A^T v = \lambda v$.

Cookbook 4.8 (Cookbook procedure to deriver λ and v)**.** Upon simply algebra, we see:

$$Av - \lambda v = 0 \iff Av - \lambda I_n v = (A - \lambda I_n)v = 0$$

Now, if $A - \lambda I_n$ is invertible (?? due to the construct of matrix??), then we are hopeless to find a valid/meaningful eigenvector.

Then, we look at the more interesting case where $A - \lambda I_n$ is not invertible, that is:

$$\det(A - \lambda I_n) = 0$$

To solve for λ and v:

Step 1: Find all λ's s.t. $\det(A - \lambda I_n) = 0$;

Step 2: For each eigenvalue λ, find all the vectors v s.t.:

$$(A - \lambda I_n)v = 0$$

Remark 4.9. When solving for λ of matrix A, we are literally solving a polynomial of degree n. (Remember we have: A is an $n \times n$ matrix.) Threrfore:

- When considering $\lambda \in \mathbb{R}$, we may well have m distinct eigenvalues s.t. $m \leq n$
  1. It is possible that λ_i has multiplicity > 1;
  2. It is also likely that there is not so much real roots available to the polynomial.
- For an eigenvalue λ with multiplicity k, it has $j \leq$k linearly independent eigenvalues. (Again, this peculiarity arose from solving polynomials of degree n, as well as the corresponding traits of matrix A).

Orthogonal projections

Definition 4.10. Let U be a vector space, and V be a linear subspace of U. The orthogonal projection of a vector $u \in U$ to V is a vector $P_v(u)$ s.t.

(i) $P_v(u) \in V$;

(ii) $(u - P_v(u)) \cdot v = 0$, $\forall v \in V$.

Or, **equivalently**, $P_v(u) \coloneqq \arg\min_{v \in V} ||u - v||$.

Proposition 4.11. If V is a linear subspace of linear space U, then $\forall u \in U$, $\exists \hat{u} \in V$ and $\varepsilon \in V^{\text{orthogonal space w.r.t } U}$ s.t.

$$u = \hat{u} + \varepsilon$$

where $\varepsilon \cdot v = 0$, $\forall v \in V$.

5 Lecture 5: (Riemann) Integral

Remark 5.1 (Overview of types of integrations)**.** Riemann Integration $\rightarrow$ Riemann-Stieltjes Integration $\rightarrow$ Lebesgue Integration $\rightarrow$ Lebesgue-Stieltjes Integration.

Definition 5.2 (Partitions in $\mathbb{R}$.)**.** A partition P on $[a, b]$ is a finite set of points $x_0, x_1, \ldots, x_n$ s.t.

$$x_0 = a \leq x_2 \leq \ldots \leq x_{n-1} \leq x_n = b$$

Denote $\Delta x_i = x_i - x_{i-1}$, $\forall i > 0$.

Definition 5.3 (Given partition P, M_i and m_i; $U(P, f)$ and $L(P, f)$; $\overline{\int_a^b} f dx$ and $\underline{\int_a^b} f dx$)**.** Assume first that f is bounded and consider a given partition P, write:

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

Note: the ultimate goal is to establish the notion of integral by refining the partitions.

Given f and a certain partition P, we define:

$$U(P, f) = \sum_{i=1}^{n} M_i \Delta x_i \quad \text{and} \quad L(P, f) = \sum_{i=1}^{n} m_i \Delta x_i$$

Definition 5.4 (Riemann Integral)**.** f is Riemann integrable if

$$\overline{\int_a^b} f dx = \underline{\int_a^b} f dx$$

Then, we can write $\int_a^b f = \overline{\int_a^b} f dx = \underline{\int_a^b} f dx$.

Remark 5.5. When dealing with probability, Riemann integral is not general enough.

Definition 5.6 (Riemann-Stieltjes integral)**.** Given a partition $P = \{x_0, \ldots, x_n\}$ s.t. $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$, then define $\Delta\alpha_i = \alpha(x_i) - \alpha(x_{i-1})$. Equivalently, we write:

$$U(P, f, \alpha) = \sum_{i=1}^{n} M_i \Delta\alpha_i \quad \text{and} \quad L(P, f, \alpha) = \sum_{i=1}^{n} m_i \Delta\alpha_i$$

Accordingly, upper and lower integral is given as:

$$\overline{\int_a^b} f d\alpha = \inf_P U(P, f, \alpha) \quad \text{and} \quad \underline{\int_a^b} f d\alpha = \sup_P L(P, f, \alpha)$$

If $\overline{\int_a^b} f d\alpha = \underline{\int_a^b} f d\alpha$, then we say f is **Riemann-Stieltjes integrable with respect to** α **over** $[a, b]$. Then, we write: $\int_a^b f d\alpha = \overline{\int_a^b} f dx = \underline{\int_a^b} f dx$.

Proposition 5.7 (Compare Riemann to Riemann-Stieltjes)**.** When α is differentiable, Riemann integral is the same as Riemann-Stieltjes integral in the following sense:

$$\int_a^b f d\alpha = \int_a^b (f\alpha')\, dx$$

Example 7. Five examples on evaluating R-S integral were introduced in lecture. It is on page 15-17 on my notes. [Don't really want to bother with these compuational examples.] The take aways are:

- R-S integral is introduced to enhance handling of probabilities, as often the times, random variables may have a discrete distribution.
- When evaluating the R-S integral with respect to α, function values are used at the points at which α is not differentiable. Integrals are taken piece-wisely across the intervals truncated by these points of discontinuity of α.

Proposition 5.8.

- If f is continuous over $[a, b]$, then $\int_a^b f d\alpha$ exists.
- If f is bounded on $[a, b]$, and has finitely many points of discontinuity, **and** α is continuous at every point at which f is discontinuous, then $\int_a^b f d\alpha$ exists.

- If f is monotone and α is continuous, then $\int_a^b f d\alpha$ exists.

Proposition 5.9.
- If f_1 and f_2 are R-S integrable with respect to α, then

$$\int_a^b (f_1 + f_2) d\alpha = \int_a^b f_1 d\alpha + \int_a^b f_2 d\alpha$$

For $c \in \mathbb{R}$,

$$\int_a^b (cf) d\alpha = c \int_a^b f d\alpha$$
$$\int_a^b f d(c\alpha) = c \int_b^a f d\alpha$$

- If $f_1 \leq f_2$ $\forall x$ and f_1, f_2 are R-S integrable, then $\int_a^b f_1 d\alpha \leq \int_a^b f_2 d\alpha$
- If f_1, f_2 are R-S with respect to α over $[a, b]$ & $[b, c]$, then

$$\int_a^c f d\alpha = \int_a^b f d\alpha + \int_b^c f d\alpha$$

- If f is R-S integrable with respect to α_1 and α_2, then

$$\int_a^b f d(\alpha_1 + \alpha_2) = \int_a^b f d\alpha_1 + \int_a^b f d\alpha_2$$

- If f, g are R-S integrable with respect to α, then $f \cdot g$ is also R-S integrable with respect to α.

Example 8. A function that is not R-S integrable:

$$f(x) = \begin{cases} 1, & \text{if } x \in [0,1] \setminus \mathbb{Q} \\ 0, & \text{if } x \in [0,1] \bigcap \mathbb{Q} \end{cases}$$

$U(P, f) = 1$ and $L(p, f) = 0$. (Doubt if this is a valid demonstration.)

Theorem 5.10 (Fundamental Theorem of Calculus)**.** f is integrable over $[a, b]$, F is differentiable over $[a, b]$ and $F'(x) = f(x)$, then

$$F(b) - F(a) = \int_a^b f(x) dx$$

Remark 5.11 (Application of Fundamental Theorem of Calculus)**. Integration by parts :**

Let F, G be differentiable over $[a, b]$, with $F' = f, G' = g$ and f, g both integrable.

$$\int_a^b F(x)g(x)dx = F(x)G(x)|_a^b - \int_a^b f(x)G(x)dx$$

Proof.

$$(F(x)G(x))' = f(x)G(x) + F(x)g(x)$$

□

Integrating both parts from a to b yields the desired equality.

Theorem 5.12 (Fubini Theorem)**.** Consider a double integral $\int_X \int_Y f(x,y)dydx$, if either $\int_X \int_Y |f(x,y)|dydx$ or $\int_Y \int_X |f(x,y)|dxdy$ exists, **and** $f(x,y)$ is integrable fixing either x or y, then

$$\int_X \int_Y f(x,y)dydx = \int_Y \int_X f(x,y)dxdy$$

6 Lecture 6: Metric space and more

Definition 6.1 (Metric)**.** A metric d on a set X is a function from $X \times X$ to $\mathbb{R}_+$ s.t.

- $\forall (x,y) \in X \times X$, $d(x,y) \geq 0$;
- $\forall (x,y) \in X \times X$, $d(x,y) = 0$ if and only if $x = y$;
- $\forall (x,y) \in X \times X$, $d(x,y) = d(y,x)$;
- For arbitrary $x, y, z \in X$, $d(x,z) \leq d(x,y) + d(y,z)$ (Triangular inequality).

Definition 6.2. Metric space is a set X together with its metric, denoted as (X, d).

Proposition 6.3 (Cauchy-Schiwanz ineuqality)**.** For arbitrary inner product of $a, b \in \mathbb{R}^n$,

$$< a, b >= \sum_{i=1}^{n} a_i b_i \leq \sqrt{\left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)}$$

Example 9 (Metric spaces in $\mathbb{R}^n$)**.** For $X = \mathbb{R}^n$ and $d_p(x,y) := (\sum_{i=1}^{n}(|x_i - y_i|^p))^{\frac{1}{p}}$ with $p \geq 1$, (X, d_p) is a metric space.

Example 10 (Metric space for X = a set of continuous fucntions)**.** Let $C[0,1]$ denote continuous functions defined on $[0,1]$ interval. Then, $\forall f, g \in C[0,1]$, define:

$$d_p(f,g) = \left[\int_0^1 |f(x) - g(x)|^p dx\right]^{\frac{1}{p}}$$

Remark 6.4. Angle of "vectors" are not defined in metric space. They only apply to inner product space.

6.1 Topology

Definition 6.1 (ε-neighborhood of a point)**.** Let (X, d) be a metric space. Then, $\forall x \in X$, $\forall \varepsilon > 0$, the ε-neighborhood of x is:

$$B_\varepsilon(x) = \{\forall y \in X : d(x - y) < \varepsilon\}$$

Definition 6.2 (Open)**.** A subset S of X is open if $\forall x \in S$, $\exists \varepsilon > 0$ s.t. $B_\varepsilon \subseteq S$.

Definition 6.3 (Close)**.** A subset S of X is closed if $X \setminus S$ is open.

Note that,

- "not open" is not the converse of "closed". Consider $[a, b)$, which is not open nor closed.
- Open/closed is defined **w.r.t. the underlying background set**. Also note that, compactness is "background-free".
- Example: let $X = (-\infty, 0)$, then $[-1, 0)$ is closed in X. (Clearly, it is not closed in $\mathbb{R}$.)

Definition 6.4 (Limit point)**.** x is a limit point of S if $\forall \varepsilon > 0$,

$$B_\varepsilon(x) \bigcap (S \setminus \{x\}) \neq \emptyset$$

Note that, x is not necessarily an element of S.

- "Isolated points" are not limit points.
- Points at the "edge" are limit points.
- Points in the interior of the set are limit points.

Proposition 6.5 (Convergent sequence to limit point)**.** If x is a limit point of $S \subseteq X$, where X is a metric space with d; then $\exists$ a sequence $\{x_n\}$ s.t.

- $x_n \in S$, $\forall n \in \mathbb{N}$;
- $\forall N \in \mathbb{N}$, $\exists n, m \geq N$ s.t. $x_n \neq x_m$
- and $x_n \to x$.

$x_n \in S$

Definition 6.6 (Boundary point)**.** A point $x \in X$ is called a boundary point of $S \subseteq X$ if

$$\forall \varepsilon > 0, \;\; B_\varepsilon(x) \bigcap S \neq \emptyset \text{ and } B_\varepsilon(x) \bigcap (X \setminus S) \neq \emptyset$$

Definition 6.7 (Interior point)**.** A point $x \in S$ is an interior point of $S \subseteq X$ if

$$\exists \varepsilon > 0, \;\; \text{s.t. } B_\varepsilon(x) \subseteq S$$

The set of interior points of S is denoted as $int(S)$.

Definition 6.8 (Isolated point)**.** A point $x \in S$ is called an isolated point of $S \subseteq X$ if

$$\exists \varepsilon > 0, \quad \text{s.t. } B_\varepsilon(x) \bigcap S = \{x\}$$

Definition 6.9 (Closure)**.** The closure of $S \subseteq X$ is the smallest cloest set that contains S. It is denoted as $cl(S)$.

Proposition 6.10 (open and closed)**.** S is open if and only if $S = int(S)$; S is closed if and only if $S = cl(S)$.

Definition 6.11 (Boundary of a set)**.** The boundary of S is $bd(S) = cl(S) \setminus int(S)$.

6.2 Sequence in Metric space

Definition 6.1. $\{a_n\}$ is a sequence in metric space (X, d). We say that $\{a_n\}$ converges to $a \in X$ if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ s.t.

$$a_n \in B_\varepsilon(A), \quad \forall n \geq N$$

Proposition 6.2. A convergent sequence in (X, d) has a unique limit.

Proof. Suppose $a_n \to a$, $a_n \to b$ and $a \neq b$. Then, by definition of limit of sequences:

$$\exists N_1 \text{ s.t. } d(a_n a) < \varepsilon = \frac{d(a,b)}{2}$$

$$\exists N_2 \text{ s.t. } d(a_n b) < \varepsilon = \frac{d(a,b)}{2}$$

Now, by triangular inequality of metrics, we have:

$$d(a,b) \leq d(a, a_n) + d(a_n, b) < d(a,b)$$

This inequality violates the assumption that $a \neq b$. We hereby reached a contradiction. □

Proposition 6.3. Every convergent sequence in a metric space is bounded. (That is to say: $\exists M \in \mathbb{R}_+$ s.t. $\forall x, y \in \{a_n\}$, $d(x,y) < M$; alternatively, $\exists M \in \mathbb{R}_+$ s.t. $B_M(a) \supseteq S$.)

Proof. By definition of convergent sequence, take $\varepsilon = 1$, $\exists N$ s.t. $\forall n \geq N$: $d(a_n, a) < 1$. Therefore, take:

$$M = \max\{d(a, a_1), \ldots, d(a, a_N), 1\}$$

Then, we see that $a_n \in B_M(a)$, $\forall n \in \mathbb{N}$. □

Proposition 6.4. $S \subseteq X$ is closed if **every convergent sequence** `in` S has a limit in S.

Proof. Pending. (Page 27 of personal note.) □

6.3 Functions in Metric Space

Definition 6.1 (continuity)**.** (X, d_X), (Y, d_Y) are two metric spaces. $f : X \to Y$ is continuous at $x \in X$ if $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t.

$$d_X(x', x) < \delta \implies d_Y(f(x'), f(x)) < \varepsilon$$

Example: under discrete metric, $d_X(a, b) = \begin{cases} 0, & \text{if } x = y \\ 1, & \text{if } x \neq y \end{cases}$, arbitrary f is continuous. (For all $\varepsilon > 0$, let $\delta < 1$, only the original point x is left.)

Definition 6.2 (connected metric space)**.** A metric space is connected if $\nexists$ nonempty open sets A, B s.t.

$$A \bigcup B = X, \quad \text{yet } A \bigcap B = \emptyset$$

Note that, by construction, since $A \bigcup B = X$, using definition of closed sets:

- A, B open $\implies$ they are closed;
- A, B closed $\implies$ they are open;

Proposition 6.3. A metric space (X, d) is connected if and only if the only pair of sets that are both closed and open are $\emptyset$ and X.

Proof. If x is connected, then $\nexists$ nonempty A, B s.t. A, B, are open, and $A \bigcup B = X$, $A \bigcap B = \emptyset$. This implies that:

$$B = X \setminus A \quad \text{and } A = X \setminus B$$

By definition, B and A are closed. Therefore, the condition that characterize "connectedness" of metric space stipulates explicitly that there is no nonempty sets that are both closed and open.

To prove the other direction, we choose to prove the contrapositive here. That is: assume that X is not connected, then there exists nonempty A and B s.t. $A \bigcup B = X$ and $A \bigcup B = \emptyset$. This conditional ($P \implies Q$) holds since Q is a rephrasing of P by definition of connected metric space. □

Proposition 6.4. (X, d_X) and (Y, d_Y) are two metric spaces and $f : X \to Y$ is a continuous function. Then, if $S \subseteq X$ is connected, then $f(X)$ is also connected.

Proof. This is a constructive proof, depending on two lemmas:

- Lemma 1: if f is continuous on X and $T \subseteq Y$ is open, then $f^{-1}(T)$ is open in X;
- Lemma 2: If A is open in X, and $S \subseteq X$, then $A \bigcap S$ is open in S.

□

7 Lecture 7: skipped

8 Lecture 8: compact set and optimization

Definition 8.1. Let (X, d) be a metric space and $S \subseteq X$ be a collection of subsets of X. Then, $\mathcal{A}$ is said to cover S if $S \subseteq \bigcup \mathcal{A}$. (Here, we write: $\bigcup \mathcal{A} = \bigcup\limits_{A \in \mathcal{A}} A$)

Definition 8.2. If $\mathcal{A}$ covers S, and $\forall A \in \mathcal{A}$ is open, then $\mathcal{A}$ is an open cover of S.

Definition 8.3 (Compact metric space)**.** A metric space X is said to be compact if every open cover of X has a finite subset of open sets that covers X.

Example 11. For $X = [0, 1)$ and $\mathcal{A} = \{[0, 1 - \frac{1}{n})\}$. Then, $\mathcal{A}$ is a open cover of A. ($\forall x \in [0, 1]$, $x \in [0, 1 - \frac{1}{n})$ for n large enough.)

Therefore, we see that the definition for compact set is violated. (For the specific open cover we fund.)

Example 12 (Metric matters)**.** For a set A being uncountable or countable infinite, under discrete metric $d(x, y) = \begin{cases} 1, & \text{if } x \neq y \\ 0, & \text{if } x = y \end{cases}$, then for the following open cover

$$\mathcal{A} = \{\{x\} : x \in X\}$$

We can never find a finite subcover.

Definition 8.4 (Compact sets (in metric spaces))**.** $S \subseteq X$ is compact if $\forall$ open cover of S has a finite subcover.

Proposition 8.5. If $S \subseteq X$ is compact (X being a metric space), then $\forall \{a_n\}$ in S has a convergent subsequence whose limit is in S.

Proof. Suppose that $\{a_n\}$ is a sequence in S s.t. $\nexists$ a subsequence that converges to $x \in S$. That is, for arbitrary point in S, $B_\varepsilon(x)$ shall only contain finitely many members of the sequence. (Otherwise, this would be the limit of a convergent subsequence.)

Then, for any point $x \in X$, we can construct an open cover. By compactness, the set X shall have a finite subcoering. This contradicts the assumption. (According to the assumption, ruling out these finite members of the sequence that are covered, the rest shall go somewhere.) $\square$

Remark:

1. We only stated one direction, but this is a two way result: "if and only if" relation. We only show $\implies$ direction.

2. This result also says: compactness implies *sequential compactness*. [This is a comment]

Proposition 8.6 (Extreme Value Theorem)**.** If $f : X \to \mathbb{R}$ is continuous and X is a compact metric space, then $\exists x_1, x_2 \in X$, s..t. $f(x_1) = \sup\limits_{x\in X} f(x)$, $f(x_2) = \inf\limits_{x\in X} f(x)$ where f is bounded.

Proof. Suppose first that f is bounded (to be proved later). it suffices to show that $\exists x_1 \in X$ s.t. $f(x_1) = \sup\limits_{x\in X} f(x)$.

Let $M = \sup\limits_{x\in X} f(x)$, by definition, $\forall \varepsilon > 0$,

$$\exists y_1 \in X \text{ s.t. } f(y_1) \in (M - \varepsilon, M)$$

Then, we can find a sequence $\{y_n\}$ s.t. $\lim_{n\to\infty} f(y_n) = M$.

Since $\{y_n\}$ is defined on a compact metric space, then $\exists \{y_{n_k}\}_{k=1}^{\infty}$ s.t. $\lim\limits_{k\to\infty} y_{n_k} = y \in X$. Now, consider such sequence y_{n_k} and $f(y_{n_k})$:

$$\lim_{k\to\infty} f(y_{n_k}) = M \text{ by } y_n \to M$$

$\square$

Idea of proof for boundedness of f. $\forall n \in \mathbb{N}$, we find y_n s.t. $f(y_n) > n$ but $y_{n_k} \to y \in X$. So, $f(y) = \lim f(y_{n_k}) = +\infty$, contradiction to the fact that $f(y) \in \mathbb{R}$. $\square$

Proposition 8.7. X, Y are metric spaces and $f : X \to Y$ is continuous if $S \subseteq X$ is compact then $f(S)$ is compact.

Proof. Suppose $\mathcal{A}$ is an open cover of $f(S)$, then we want to show that there exists a finite subcover of $f(S)$.

Since $\forall A \in \mathcal{A}$ is open, therefore, by continuity of f, $f^{-1}(A) \subseteq X$ is open. [**The preimage of an open set is open.**]. Then, we can write:

$$S \subseteq \bigcup_{A\in\mathcal{A}} f^{-1}(A)$$

(If not, $\exists x \in S$ s.t. $x \notin f^{-1}(A)$, $\forall A \in \mathcal{A}$. That is, $f(x) \notin A$, $\forall A \in \mathcal{A}$. However, since $x \in S$, $f(x) \in f(S)$ and $\mathcal{A}$ covers $f(S)$. This is a contradiction.)

Now, because S is compact, we can find a finite subcover, $\{f^{-1}(A_i)\}_{i=1}^{n} \subseteq \mathcal{A}$, that covers S. Then,

$$f(S) \subseteq \bigcup_{i=1}^{n} A_i \quad \Longrightarrow \quad f(S) \text{ is compact.}$$

$\square$

Proposition 8.8. A compact subset $S \subseteq X$ is closed and bounded if (X, d) is a metric space.

Proof.

- Boundedness: $\{B_1(x) : x \in S\}$ is an open cover of S;

  S is compact, thus, for $\{B_1(x_i\}_{i=1}^n$ is an open subcover of S. Then, $\forall (x, y) \in S \implies x \in B_1(x_i)$ and $y \in B_1(x_j)$. To reveal boundless, find the furtherst distance:

  $$d(x, y) \leq d(x_1, x_i) + d(x_i, x_j) + d(x_j, y)$$

  This is bounded by a finite number, since the term in the middle is bounded, yet the rest are bounded by 1 in terms of the distance apart.

- Closed (call the definition using convergent sequence);

  1. $\exists x \in X \setminus S$ s.t. $\forall B_\varepsilon(x)$, $B_\varepsilon(x) \bigcup S \neq \emptyset$.
  2. Find a sequence $\{x_n\}$ in S s.t. $x_n \to x$.
  3. We thereby find a sequence whose any subsequence converges to $x \notin S$, contradicting the previous result.

□

References

[Rudin(1976)] Walter Rudin. Principles of mathematical analysis. *New York: McGraw-Hill,[1976]*, 1976.