Lecture Notes in Economic Mathematics

Weikai Chen

August 28, 2017

Contents

Introduction

These notes are prepared for the math review at Umass Amherst, summer 2017. The review is based on the following textbook:

> Chiang, Alpha and Wainwright, Kevin, *Fundamental Methods of Mathematical Economics*, (Fourth Edition) McGraw Hill, 2005.

This is also the required textbook for the fall semester course ECON 751 (Mathematical Methods in Economics). We will cover the following sections in the first 8 chapters: 3.2-3.5, 4.3-4.6, 5.3-5.6, 6.4, 6.7, 7.4-7.6, 8.4, 8.5. The students enrolled in ECON 751 during the fall will take an exam in the first week of the semester. This exam, containing questions based on the end of chapter exercises for the above sections, will count for 20% of the final grade for ECON 751.

Note: I will follow these lecture notes instead of the textbook. Some topics in linear algebra and real analysis that are useful but not covered in the textbook will be discussed in the lecture. Daily problem set will be assigned.

1 Preliminaries on Sets, Mappings, Relations, and Real Numbers

In these preliminaries we set up the scene by describing some notions regarding sets, mappings and relations that will be used throughout the notes. Also, we will discuss the axioms on real numbers, which is quite fundamental but not covered in the textbook.

1.1 Unions and intersections of sets

For a set A, an element x may be in A, denoted by $x \in A$, or not, by $x \notin A$.We also call a member of A a *point* in A. Let A, B be sets, we call A is a **subset** of B if each member of A is a member of B. That is,

$$A \subseteq B \Leftrightarrow \forall x \in A, x \in B.$$

We may also say A is contained in B or B contains A. Two sets A and B are the same provided they have the same members, denoted by $A = B$. We have

$$A = B \Leftrightarrow A \subseteq B, B \subseteq A.$$

We say A is a **proper subset** of B if $A \subseteq B$ but $A \neq B$.

The **union** of A and B, denoted by $A \cup B$, is the set of all points that belongs either to A or to B; that is

$$A \cup B = \{x | x \in A \text{ or } x \in B\}.$$

The **intersection** of A and B is defined as

$$A \cap B = \{x | x \in A \text{ and } x \in B\}.$$

The **complement** of A in B, denoted by $B \sim A$, is the set of all points in B that is not in A; that is

$$B \sim A = \{x | x \in B, x \notin B\}.$$

Let U be the universe set, then we often refer to $U \sim A$ simply as the complement of A, denoted by A^c.

The set that has no elements is called the **empty-set**, denoted by $\emptyset$, and the set that has a single member is called a **singleton set**. Given a set X, the **power set** of X is the set of all subsets of X, denoted by $\mathcal{P}(X)$ or 2^X.

Exercise 1.1. Let $X = \{1, 2, 3\}$, how many members are there in $\mathcal{P}(X)$? What if $X = \{1, \ldots, n\}$?

Proposition 1.1 (De Morgan's Rule)**.** *The complement of the union is the intersection of the complements, and the complement of the intersection is the union of the completments, that is,*

$$(A \cup B)^c = A^c \cap B^c, \text{ and } (A \cap B)^c = A^c \cup B^c.$$

1.2 Mappings between sets

Given two set A and B, a correspondence that assigns to each member of A a member of B is called by a **mapping** or **function** from A to B, denoted by $f : A \to B$. When B is the set of real numbers, we always use "function". Given the function $f : A \to B$, for any $x \in A$, denote the member of B that is assigned to x by $f(x)$. For any subset $X \subseteq A$, define the **image** of X under f as

$$f(X) = \{b \in B \mid \exists x \in X, f(x) = b\}.$$

We call A the **domain** of f and $f(A)$ the **range** of f. If $f(A) = B$, then the function f is said to be **onto**. If

$$\forall a_1, a_2 \in A, a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2),$$

then the function f is said to be **one-to-one**. A mapping $f : A \to B$ that is both one-to-one and onto is said to be **invertible**. Given an invertible mapping $f : A \to B$, $\forall b \in B$, there exists exactly one element $a \in A$ such that $f(a) = b$, denoted by $f^{-1}(b)$. This assignment defines the mapping $f^{-1} : B \to A$, called the **inverse** of f. If there is an invertible mapping from A to B, then we say A and B are **equipotent**, which,roughly speaking, means that they have the same number of elements.

Given two mappings $f : A \to B$ and $g : C \to D$ for which $f(A) \subseteq C$ then the **composition** $g \circ f : A \to D$ is defined by

$$(g \circ f)(x) = g(f(x)), \forall x \in A.$$

It is easy to see that the composition of invertible mapping is invertible.

1.3 Equivalence Relations*

Given two nonempty sets A and B, the **Cartesian product** of A with B is defined by

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

For example, $\mathbb{R}^2 = \{(x_1, x_2) \mid x_1 \in \mathbb{R}, x_2 \in \mathbb{R}\}$ is the Cartesian product $\mathbb{R} \times \mathbb{R}$. For a nonempty set X, we call a subset $R \subseteq X \times X$ a **relation** on X and write xRx' if $(x, x') \in R$. The relation R is said to be **reflecxive** if

$$\forall x \in X, xRx;$$

the relationship is said to be **symmetric** if

$$xRx' \Rightarrow x'Rx;$$

the relation is said to be **transitive** if

$$xRx', x'Rx'' \Rightarrow xRx''.$$

Exercise 1.2. In the 2-D plane, show the relation " $\geq$ " as a subset of $\mathbb{R}^2$. Is it reflecxive, symmetric, or transitive?

Definition 1.1. A relation R on a nonempty set X is called an **equivalence relation** provided it is reflexive, symmetric, and transitive.

Definition 1.2. A relation R on a nonempty set X is called a **partial ordering** provided it is reflexive, transitive, and

$$\forall x, x' \in X, xRx', x'Rx \Rightarrow x = x'.$$

Remark. Consider the rational preference $\succeq$. It is reflexive, transitive, but not symmetric; therefore it is not a equivalence relation. $x \succeq y, y \succeq x \Rightarrow x \sim y$ but we may have $x \neq y$ in general; therefore it is not a partial ordering.

1.4 The real numbers

Denote the set of real number by $\mathbb{R}$. For any $a, b \in \mathbb{R}$, the sum $a + b$ and the product ab are well-defined. We assume that $\mathbb{R}$, with the sum and product, satisfies three types of axioms: Field Axioms, Positivity Axioms, and Completeness Axioms.

The Field Axioms.

Commutativity of Addition: $\forall a, b \in \mathbb{R}, a + b = b + a$.

Associativity of Addition: $\forall a, b, c \in \mathbb{R}, (a + b) + c = a + (b + c)$.

The Additive Identity: $\forall a \in \mathbb{R}, a + 0 = a$.

The Additive Adverse: $\forall a \in \mathbb{R}, \exists b \in \mathbb{R}, a + b = 0$. Denote b by $-a$.

Commutativity of Multiplication: $\forall a, b \in \mathbb{R}, ab = ba$.

Associativity of Multiplication: $\forall a, b, c \in \mathbb{R}, (ab)c = a(bc)$.

The Multiplication Identity: $\forall a \in \mathbb{R}, 1a = a$

The Multiplication Adverse: $\forall a \neq 0, \exists b \in \mathbb{R}, ab = 1$. Denote b by a^{-1}.

The Distributive Property: $\forall a, b, c \in \mathbb{R}, a(b + c) = ab + ac)$.

The Nontriviality Assumption: $1 \neq 0$.

Remark. Any set with the operations sum and product that satisfies these axioms is called a **field**. You can check that with the conventional sum and product, the set of the rational number $\mathbb{Q}$ is a field, while the set of integers $\mathbb{N}$ is not.

The Positivity Axioms.

There is a set of real numbers, denoted by P, called the **positive numbers**, satisfying the following properties:

P1 If a and b are positive, then ab and $a + b$ are positive.

P2 For a real number a, exactly one of the following is true: a is positive, $-a$ is positive, $a = 0$.

Define an ordering of the real number by $a > b$ when $a - b$ is positive and so on, and the **interval**s

$$(a, b) = \{x \mid a < x < b\}; \qquad [a, b) = \{x \mid a \leq x < b\}$$
$$(a, b] = \{x \mid a < x \leq b\}; \qquad [a, b] = \{x \mid a \leq x \leq b\}$$

Remark. Also, the rational number $\mathbb{Q}$ satisfies the positivity axioms, which means that we need other property to distinguish $\mathbb{R}$ from $\mathbb{Q}$.

With the order defined aboved, we define the **absolute value** of a real number x by

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

Then we have the following inequality, called the **Triangle Inequality**: $\forall a, b \in \mathbb{R}$, $|a+b| \leq |a| + |b|$.

The Completeness Axioms.

A nonempty set $E \subseteq \mathbb{R}$ is said to be **bounded above** if $\exists b \in \mathbb{R}, \forall x \in E, x \leq b$: the number b is called an **upper bound** for E. The smallest upper bound, if exists, is called the **least upper bound**, or the **supremum** of E, denoted by $\sup E$.

Exercise 1.3. Similarly, define the set E that is **bounded below**, the **lower bound** and the **greatest lower bound** or **infimum**, $\inf E$.

For any set that is bounded above, the axiom of completeness asserts that its supremum exists.

The Completeness Axiom Let $E \subseteq \mathbb{R}$ be bounded above, its least upper bound, or supremum exists.

Example 1.1. Consider $E = [0, 2)$. Then 2 is a upper bound. For any upper bound b, we have $b \geq 2$: otherwise, we have $b < \frac{b+2}{2} \in E$, contradicted. Therefore, we 2 is the least upper bound, or $\sup E = 2$.

Remark. Consider the set $E_1 = \{x \in \mathbb{Q} \mid x^2 < 2\}$ and $E_2 = \{x \in \mathbb{R} \mid x^2 < 2\}$. Both sets are bounded above. It can be show that in $\sup E_2 = \sqrt{2}$, but $\sqrt{2} \notin \mathbb{Q}$. No least upper bound for E_1. In other words, $\mathbb{Q}$ does not satisfied the completeness axiom.

A taste of $\epsilon - \delta$ language*

Let $E \subseteq \mathbb{R}$ be a nonempty and bounded above. Then if $y = \sup E$ then $\forall \epsilon > 0, \exists x \in E$ such that

$$y - \epsilon < x < y.$$

Exercise 1.4. Write the counterpart for infimum.

2 Basic Linear Algebra

Linear algebra studies the **linear transformation** on **vector spaces**, which can be represented by **matrix**. We will discuss what these terms mean.

2.1 Vector Spaces

Given a set V of objects, called **vectors**, suppose there are two operations: **vector addition** such that the sum of the vectors $\mathbf{x}$ and $\mathbf{y}$ is a vector denoted by $\mathbf{x}+\mathbf{y}$, and **scalar multiplication** such that the product of the scalar (i.e., real number) c and the vector $\mathbf{x}$ is a vector denoted by $c\mathbf{x}$.

The set V, together with these two operations, is called a **vector space** (or **linear space**) if the following properties hold for all vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all scalars c, d:

(1) $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$.

(2) $\mathbf{x}+(\mathbf{y}+\mathbf{z}) = (\mathbf{x}+\mathbf{y})+\mathbf{z}$.

(3) There is a unique vector $\mathbf{0}$ such that $\mathbf{0}+\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$.

(4) $\mathbf{x}+(-1)\mathbf{x} = \mathbf{0}$.

(5) $1\mathbf{x} = \mathbf{x}$.

(6) $c(d\mathbf{x}) = (cd)\mathbf{x}$.

(7) $(c+d)\mathbf{x} = c\mathbf{x}+d\mathbf{y}$.

(8) $c(\mathbf{x}+\mathbf{y}) = c\mathbf{x}+c\mathbf{y}$.

One example of a vector space is the set $\mathbb{R}^n$ with component-wise addition and multiplication by scalars. That is, if $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$, then

$$\begin{aligned} \mathbf{x}+\mathbf{y} &= (x_1+y_1, \ldots, x_n+y_n), \\ c\mathbf{x} &= (cx_1, \ldots, cx_n). \end{aligned}$$

If V is a vector space, then a subset W of V is called a **linear subspace** (or simply, a **subspace**) of V if for all $\mathbf{x}, \mathbf{y} \in W$ and every scalar c, we have

$$\mathbf{x}+\mathbf{y} \in W, c\mathbf{x} \in W.$$

In this case, W itself satisfies properties (1)-(8) if we use the operations that W inherits from V, so that W is a vector space in its own right.

Exercise 2.1. Give two examples of subspaces of $\mathbb{R}^2$.

Let V be a vector space. A set $\mathbf{a}_1, \ldots, \mathbf{a}_m$ of vectors in V is said to **span** V, denoted by $V = span\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$, if $\forall \mathbf{x} \in V$, there exists at least one m-tuple of scalars $c_1, \ldots, c_m$ such that

$$\mathbf{x} = c_1\mathbf{a}_1 + \ldots + c_m\mathbf{a}_m.$$

In this case, we say that $\mathbf{x}$ can be written as a **linear combination** of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

Example 2.1. Let $V = \mathbb{R}^2$, the set of vectors $\mathbf{a}_1 = (1, 0), \mathbf{a}_2 = (1, 1)$ span V.

The set $\mathbf{a}_1, \ldots, \mathbf{a}_m$ of vectors is said to be **independent** if to each $\mathbf{x} \in V$ there corresponds at most one $m-$tuple of scalars $c_1, \ldots, c_m$ such that

$$\mathbf{x} = c_1\mathbf{a}_1 + \ldots + c_m\mathbf{a}_m.$$

Proposition 2.1. *The set $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ is independent provided*

$$\mathbf{0} = d_1\mathbf{a}_1 + \ldots + d_m\mathbf{a}_m \Longrightarrow d_1 = d_2 = \ldots = d_m = 0.$$

If the st of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$ both span V and is independent, it is said to be a **basis** for V.

Theorem 2.2. *Suppose V has a basis consisting of m vectors. Then any set of vectors that span V has at least m vectors, and any set of vectors that is independent has at most m vectors. In particular, any basis for V has exactly m vectors.*

If V has a basis consisting of m vectors, we say m is the **dimension** of V, denoted by $\dim V$. We make the convention that the vector space consisting of the zero vector alone has dimension zero. It is easy to see that $\mathbb{R}^n$ has dimension n. The following set of vectors is called the standard basis for $\mathbf{R}^n$:

$$\begin{aligned}
\mathbf{e}_1 &= (1, 0, 0, \ldots, 0),\\
\mathbf{e}_2 &= (0, 1, 0, \ldots, 0),\\
&\ldots\\
\mathbf{e}_n &= (0, 0, 0, \ldots, 1).
\end{aligned}$$

The vector space $\mathbb{R}^n$ has many other bases, but any basis of $\mathbb{R}^n$ must consist of precisely n vectors.

Because $\mathbb{R}^n$ has a finite basis, so does every subspace of $\mathbb{R}^n$. This fact is a consequence of the following theorem:

Theorem 2.3. *Let V be a vector space of dimension m. If W is a linear subspace of V (different from V), then W has dimension less than m. Furthermore, any basis $\mathbf{a}_1, \ldots, \boldsymbol{a}_k$ for W may be extended to a basis*

$$\mathbf{a_1}, \ldots, \mathbf{a}_k, \boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_m$$

for V.

2.2 Inner Products

If V is a vector space, an **inner product** on V is a function assigning, to each pair of $\mathbf{x}, \mathbf{y}$ of vectors of V,a real number denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$, such that the following properties hold for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all scalars c:

(1) $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.

(2) $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$.

(3) $\langle c\mathbf{x}, \mathbf{y} \rangle = c\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, c\mathbf{y} \rangle$.

(4) $\langle \mathbf{x}, \mathbf{x} \rangle > 0$ if $\mathbf{x} \neq \mathbf{0}$.

A vector space V together with an inner product on V is called an **inner product space**.

A given vector space may have many different inner products. One particularly useful inner product on $\mathbb{R}^n$ is defined as follows: for all $\mathbf{x} = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$, define

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \ldots + x_n y_n.$$

This is the inner product we commonly use. It is also called the **dot product** and denoted by $\mathbf{x} \cdot \mathbf{y}$ or $\mathbf{x}'\mathbf{y}$.

Exercise 2.2. Verify the the inner product defined above satisfies the properties (1)-(4).

If V is an inner product space, one can defines the **length** (or **norm**) of a vector of V by

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{\frac{1}{2}}.$$

You can verify that the norm function has the following properties:

(1) $\|\mathbf{x}\| > 0$ if $\mathbf{x} \neq \mathbf{0}$.

(2) $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$.

(3) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The last inequality is called the **triangle inequality**, which can be proved by the **Cauchy-Schwarz inequality**

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

The proof of Cauchy-Schwarz Inequality. For any $a \in \mathbb{R}$, we have $\langle a\mathbf{x} + \mathbf{y}, a\mathbf{x} + \mathbf{y} \rangle \geq 0$. That is

$$\|\mathbf{x}\|^2 a^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle a + \|\mathbf{y}\|^2 \geq 0, \forall a \in \mathbb{R}$$

Then we have

$$\Delta = 4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\|\mathbf{x}\|^2\|\mathbf{y}\|^2 \leq 0 \Rightarrow \langle \mathbf{x}, \mathbf{y} \rangle \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

□

The proof of Triangle Inequality.

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \\ &\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \\ &= (\|\mathbf{x} + \mathbf{y}\|)^2 \end{aligned}$$

□

Remark. Any function from V to real number that satisfies properties (1)-(3) just listed is called a **norm** on V. The length function derived from an inner product is one example of a norm, but there are other norms that are not derived from inner products. On $\mathbb{R}^n$, for example, one has not only the familiar norm derived from the dot product, which is called the **euclidean norm**, but one has also the **sup norm**, defined by

$$|\mathbf{x}| = \max\{|x_1|, \ldots, |x_n|\}.$$

2.3 Matrices

A **matrix** is a rectangular array of numbers.

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

is called an $m \times n$ matrix, sometimes denoted by $(a_{ij})_{m\times n}$. Let $B = (b_{ij})_{m\times n}$, define the sum of A and B by

$$\begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

and the product of A and real number c by

$$cA = \begin{bmatrix} ca_{11} & \dots & ca_{1n} \\ \vdots & \ddots & \vdots \\ ca_{m1} & \dots & ca_{mn} \end{bmatrix}$$

With these operation, the set of all $m \times n$ matrices is a vector space, denoted by $\mathcal{M}(m, n)$.

The set of matrices has an additional operation called **matrix multiplication**. For $A_{m\times n}$ and $B_{n\times p}$, the product AB is defined to be the matrix $C = (c_{ij})_{m\times p}$ where

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Define the identical matrix $I_k = (\delta_{ij})_{k\times k}$ where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Then it has the form

$$I_k = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

You can verify that the matrix multiplication satisfies the following properties:

(1) $A(BC) = (AB)C$

(2) $A(B + C) = AB + AC$

(3) $(A + B)C = AC + BC$

(4) $(cA)B = c(AB) = A(cB)$

(5) For any $A_{m\times n}$,$I_m A = A$, and $AI_n = A$.

Given a matrix $A = (a_{ij})_{m\times n}$, define its **transpose** by $D = (d_{ij})_{n\times m}$ with $d_{ij} = a_{ji}$,denoted by A^{tr} (or A' or A^t). Then we have

(1) $(A^{tr})^{tr} = A$.

(2) $(A + B)^{tr} = A^{tr} + B^{tr}$.

(3) $(AC)^{tr} = C^{tr}A^{tr}$.

2.4 Linear Transformations

Let V and W be two vector spaces, a mapping $T : V \to W$ is called a **linear transformation** if it satisfies: $\forall \mathbf{x}, \mathbf{y} \in V, \forall c \in \mathbb{R}$:

(1) $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$

(2) $T(c\mathbf{x}) = cT(\mathbf{x})$.

The following theorem asserts that a linear transformation is uniquely determined by the values on basis elements.

Theorem 2.4. *Let V be a vector spaces with basis $\mathbf{a}_1, \ldots, \mathbf{a}_m$. Let W be a vector space. Given any m vectors $\mathbf{b}_1, \ldots, \mathbf{b}_m$, there is exactly one linear transformation $T : V \to W$ such that*

$$T(\mathbf{a}_i) = \mathbf{b}_i, \forall i = 1, \ldots, m$$

Proof. First we show the existence by constructing a function $T : V \to W$ in the following way. Let

$$T(\mathbf{a}_i) = \mathbf{b}_i, \forall i = 1, \ldots, m$$

and for any $\mathbf{x} = \sum_{i=1}^n c_i \mathbf{a_i} \in V$, let

$$T(\mathbf{x}) = \sum_i^n c_i T(\mathbf{a}_i) = \sum_i^n c_i \mathbf{b}_i.$$

Then $T : V \to W$ is a linear transformation.

Next we show that the linear transformation is unique. Suppose there are two linear transformation $T : V \to W$ and $S : V \to W$ such that

$$T(\mathbf{a}_i) = S(\mathbf{a}_i) = \mathbf{b}_i, \forall i = 1, \dots, m$$

the for any $\mathbf{x} = \sum_{i=1}^n c_i \mathbf{a_i} \in V$,

$$T(\mathbf{x}) = \sum_i^n c_i T(\mathbf{a}_i) = \sum_i^n c_i \mathbf{b}_i = \sum_i^n c_i S(\mathbf{b}_i) = S(\mathbf{x}).$$

□

The matrix representation of a linear transformation.

Any linear transformation from one finite dimensional mapping to another can be represented by matrices.

Let's first look at the special case. Let $A \in \mathcal{M}(m, n)$, the function $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T(\mathbf{x}) = A\mathbf{x}, \quad \forall \mathbf{x} \in \mathbb{R}^n$$

is a linear transformation. You can check it using the properties of matrix multiplications.

On the other hand, for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, there exists some matrix $A \in \mathcal{M}(m, n)$ such that $T(\mathbf{x}) = A\mathbf{x}, \forall \mathbf{x} \in \mathbb{R}^n$.

Proof. Let $\mathbf{b}^i = T(\mathbf{e}_i), \forall i = 1, \dots, n$ where $\mathbf{e}_1, \dots, \mathbf{e}_n$ is the standard basis for $\mathbb{R}^n$, and construct a matrix $A = [\mathbf{b}^1 \dots \mathbf{b}^n] \in \mathcal{M}(m, n)$. Then the mapping $S(\mathbf{x}) = A\mathbf{x}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ with

$$S(\mathbf{e}_i) = \mathbf{b}^i, \forall i = 1, \dots, n.$$

By Theorem 2.4, we have $T(\mathbf{x}) = S(\mathbf{x}) = A\mathbf{x}$. □

In general, let $T : V \to W$ be linear, $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be a basis for the vector space V, and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be a basis for W. For any $\mathbf{x} \in V$, there are unique numbers $x_1, \ldots, x_m$ such that

$$\mathbf{x} = x_1\mathbf{v}_1 + \ldots + x_m\mathbf{v}_m.$$

Let $\mathbf{y} = T(\mathbf{x}) \in W$, then $\mathbf{y}$ can be expressed as a unique linear combination

$$\mathbf{y} = y_1\mathbf{w}_1 + \ldots + y_n\mathbf{w}_n.$$

Also, $T(\mathbf{v}_j) \in W$, there are unique numbers $a_{1j}, \ldots, a_{nj}$ such that

$$T(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + \ldots + a_{nj}\mathbf{w}_n.$$

Then

$$\begin{aligned}\sum_{i=1}^{n} y_i\mathbf{w}_i = T(\mathbf{x}) = T(\sum_{j=1}^{m} x_j\mathbf{v}_j) &= \sum_{j=1}^{m} x_j T(\mathbf{v}_j) \\ = \sum_{j=1}^{m} x_j \sum_{i=1}^{n} a_{ij}\mathbf{w}_i &= \sum_{i=1}^{n}(\sum_{j=1}^{m} a_{ij}x_j)\mathbf{w}_i\end{aligned}$$

Therefore, $y_i = \sum_{j=1}^{m} a_{ij}x_j, \forall i$. Denote $\mathbf{x} = (x_1, \ldots, x_m)$, $\mathbf{y} = (y_1, \ldots, y_n)$, $A = (a_{ij}) \in \mathcal{M}(n, m)$. Then

$$\mathbf{y} = T(\mathbf{x}) = A\mathbf{x}.$$

Therefore, the matrix A represents the linear transformation $T : V \to W$. Furthermore, it can be seen that the matrix A depends on the bases we used for V and W. In general, the representing matrix would be different if we choose a different basis for V or W.

Example 2.2. The identity mapping $id : V \to V$ defined by

$$id(\mathbf{x}) = \mathbf{x}, \forall \mathbf{x} \in V$$

is a linear transformation. It can be represented by the identity matrix I_k where k is the dimension of V, no matter which basis is used.

Composition and inverse of linear transformation.

Let V, W, U be three vector spaces with dimension m, n and p, and $T : V \to W$ and $S : W \to U$ be two linear transformation, you can show that the composite $S \circ T : V \to U$ is also linear.

Given the bases $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$, and $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, suppose T and S are represented by matrices $A_{n\times m}$ and $B_{p\times n}$. Then the composite $S \circ T$ is represented by the matrix multiplication $BA = (c_{ij})_{p\times m}$.

Proof. We have $S \circ T : V \to U$ is linear. Suppose that it is represented by the $p \times m$ matrix C. We want to show that $C = BA$. Since

$$T(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + \ldots + a_{nj}\mathbf{w}_n,$$

$$S(\mathbf{w}_k) = b_{1k}\mathbf{u}_1 + \ldots + b_{pk}\mathbf{u}_p.$$

we have

$$\begin{aligned} \sum_{i=1}^{p} c_{ij}\mathbf{u}_i = S(T(\mathbf{v}_j)) &= S(\sum_{k=1}^{n} a_{kj}\mathbf{w}_k) = \sum_{k=1}^{n} a_{kj}S(\mathbf{w}_k) \\ &= \sum_{k=1}^{n} a_{kj}\sum_{i=1}^{p} b_{ik}\mathbf{u}_i = \sum_{i=1}^{p}(\sum_{k=1}^{n} b_{ik}a_{kj})\mathbf{u}_i \end{aligned}$$

Therefore, $c_{ij} = \sum_{k=1}^{n} b_{ik}a_{kj}, \forall i, j$. Then $C = BA$. $\square$

For any invertible linear transformation $T : V \to W$, its inverse mapping $T^{-1} : W \to V$ is also linear. Also, in the following exercises you may show that if $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is basis for V, then $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ is basis for W provided that T is invertible.

Exercise 2.3. Show that $T^{-1} : W \to V$ is linear.

Exercise 2.4. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is basis for V, $T : V \to W$ is invertible. Show that $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ span the vector space W.

Exercise 2.5. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is basis for V, $T : V \to W$ is invertible. Show that $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ is independent.

Then, W must have the same dimension as V. Therefore, T can be represented by a $n \times n$ matrix, denoted by$A_{n\times n}$. Suppose T^{-1} is represented by $B_{n\times n}$. By the definition of inverse mapping, we have

$$T \circ T^{-1} = id : W \to W$$
$$T^{-1} \circ T = id : V \to V$$

Then by the previous result for composite mapping and the identity mapping on Example 2.2, we have

$$AB = I_n = BA.$$

For any matrix $A_{n\times n}$, if there exist another matrix $B_{n\times n}$ such that $AB = BA = I_n$, then the matrix$A$ is said to be **invertible**, and B is the **inverse matrix** of A, denoted by A^{-1}. With this definition, we have T^{-1} is represented by A^{-1} provided that $T : V \to V$, represented by A, is invertible.

Furthermore, the invertible linear transformation is also called a **linear isomorphism**, and two vector spaces are **isomorphic** if there exists an linear isomorphism between them.

For any vector space V with basis $\mathbf{v}_1, \ldots, \mathbf{v}_m$, there exists an invertible mapping $T : V \to \mathbb{R}^m$ such that $T(\mathbf{v}_i) = \mathbf{e}_i, i = 1, \ldots, m$. Therefore, any vector spaces with dimension m are isomorphic to $\mathbb{R}^m$, which means that they the same structure as the Euclidean space with the same dimension.

2.5 The Four Fundamental Subspaces.

Let $T : V \to W$ is linear. The range of T is defined by

$$\{T(\mathbf{x}) = A\mathbf{x} \in W \mid \forall \mathbf{x} \in V\} \subseteq V$$

as we discussed in Section 1.2. Also, we can define the **kernel** of T, denoted by $\ker T$, as the set of vectors that has $\mathbf{0}_W$ as image. That is

$$\ker T = \{\mathbf{x} \in V \mid T(\mathbf{x}) = \mathbf{0}_W\} \subseteq V.$$

It is easy to check that the range of T is a subspace of W, and that the kernel of T is a subspace of V, which is also called the **null space** of T.

Exercise 2.6. Show that $\mathbf{0} \in \ker T$, and that $\ker T$ is a subspace of V.

Define the **rank** of the T by the dimension of the range of T, and the **nullity** of T by the dimension of the null space of T. Then we have the following result called the **rank-nullity theorem**.

Theorem 2.5 (Rank-Nullity Theorem)**.** *Let $T : V \to W$ be a linear transformation, then*

$$\operatorname{rank} T + \operatorname{nullity} T = \dim V$$

Sketch of the proof. Suppose $\dim V = n$, and $\operatorname{nullity} T = k$, we want to show that $\operatorname{rank} T = n - k$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be a basis for the null space. Then it can be extended to a basis for V, $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$. Then it is sufficient to show that $T(\mathbf{v}_{k+1}), \ldots, T(\mathbf{v}_n)$ is a basis for the range of T. □

For any matrix $A_{m\times n}$, rewrite

$$A = [\mathbf{a}^1 \ldots \mathbf{a}^n] = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$$

where $\mathbf{a}_i$ is the ith row of A and $\mathbf{a}^j$ is the jth column of A. Define the **column space** by $span\{\mathbf{a}^1, \ldots, \mathbf{a}^n\}$, and call its dimension the **column rank** of A. Then, it can be seen that the column space of the matrix A is the same as the range of the linear transformation T by

$$T(\mathbf{x}) = A\mathbf{x} = [\mathbf{a}^1 \ldots \mathbf{a}^m] \begin{bmatrix} x_1 \\ \ldots \\ x_m \end{bmatrix} = x_1\mathbf{a}^1 + \ldots + x_m\mathbf{a}^m.$$

Therefore, the column rank of a matrix A is equal to the rank of the linear transformation $T(\mathbf{x}) = A\mathbf{x}$.

Define the row space as $span\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ and call its dimension the **row rank** of A. The row space of A is the column space of A^{tr}, the range of the mapping $S(\mathbf{y}) = A^{tr}\mathbf{y} : \mathbb{R}^m \to \mathbb{R}^n$.

One fundamental result is that for any matrix the row rank is equal to the column rank, which is called the **rank of the matrix**. Therefore, it is the same as the rank of the corresponding linear transformation. It is clear that the row rank of $A_{m\times n}$ is less than or equal to m and the column rank is less than or equal to n. Therefore, we have

$$\operatorname{rank} A \leq \min\{m, n\}.$$

If the equality holds, the matrix is said to be **full rank**.

For a square matrix $A_{n\times n}$, it is full rank when $\operatorname{rank} A = n$ and $\operatorname{nullity} A = 0$. Therefore, it is invertible (check it).

Remark. For any matrix $A_{m\times n}$, the *null spaces* and the *column spaces* of A and A^{tr} are called the **four fundamental subspaces**.

Linear equation system and linear transformation.

The system of linear equations with n variables and m equations can be expressed in matrix form,

$$A\mathbf{x} = \mathbf{b} \tag{1}$$

where $A \in \mathcal{M}(m, n)$. Let $T(\mathbf{x}) = A\mathbf{x}$. Consider the null space of T. We have any member in the null space is a solution to the homogeneous system

$$A\mathbf{x} = \mathbf{0} \tag{2}$$

Since $\mathbf{0}$ is always in the null space, the equation (2) at least has one solution $\mathbf{x} = \mathbf{0}$. If the nullity of T is positive, then there are infinite number of solutions to (2). If the nullity of T is 0, then $\mathbf{x} = \mathbf{0}$ is the unique solution to (2). Next, consider the range of T, or the column space of A. If $\mathbf{b}$ is not in the column space, then there is no solution to (1). If $\mathbf{b}$ belongs to it, then the equation (1) has at least one solution, say $\mathbf{x}_b$. Then all the vectors in the null space plus $\mathbf{x}_b$ also solve the equation (1).

If $A_{n\times n}$ is a square and full rank matrix, that is $\operatorname{rank} A = n$ and $\operatorname{nullity} A = 0$, and A is invertible, therefore, the (1) has an unique solution for any $\mathbf{b}$,

$$\mathbf{x}_b = A^{-1}\mathbf{b}.$$

2.6 Determinant and Matrix Inversion.

The determinant is a function that assigns, to each square matrix A, a number called the **determinant** of A and denoted $\det A$ or $|A|$, if it satisfies the following properties:

(1) $\det A$ is *multilinear* in the column vectors of matrix A. That is,

$$\det[\ldots, \mathbf{b}^i + \mathbf{c}^i, \ldots,] = \det[\ldots, \mathbf{b}^i, \ldots] + \det[\ldots, \mathbf{c}^i, \ldots]$$

$$\det[\ldots, \lambda\mathbf{a}^i, \ldots] = \lambda \det[\ldots, \mathbf{a}^i, \ldots] \forall \lambda \in \mathbb{R}$$

(2) *Alternating*:

$$\det[\ldots, \mathbf{a}^i, \mathbf{a}^{i+1}, \ldots] = -\det[\ldots, \mathbf{a}^{i+1}, \mathbf{a}^i, \ldots]$$

(3) For any identity matrix I_n, $\det I_n = 1$.

It can be shown that there exists exactly one function satisfies all these properties. Therefore, the determinant of any square matrix is well-defined. There are different ways to define this function. Here we will use induction and the so called **Laplace expansion** formula.

- For $n = 1$, $A = [a]$, $\det A = a$.
- For $n = 2$, $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $\det A = a_{11}a_{22} - a_{12}a_{21}$.
- For $n \geq 3$, define the $(i,j)-$**cofactor** of A by

  $$C_{ij} = (-1)^{i+j} \det A_{ij}$$

  where A_{ij} is the the determinant of the $(n-1) \times (n-1)$ matrix that results from deleting the i-th row and the j-th column of A, called the $(i,j)-$**minor** of A . Then

  $$\det A = a_{i1}C_{i1} + \ldots + a_{in}C_{in}.$$

You can check that the function defined above satisfies all the three properties of the determinant.

Example 2.3. For $n = 3$

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

We have the following properties of the determinant:

(1) $\det A = \det A^{tr}$.

(2) $\det AB = \det A \det B$.

(3) $\det[\ldots, \mathbf{a}^i, \ldots, \mathbf{a}^j \ldots] = 0$, when $\mathbf{a}^i = \mathbf{a}^j, i \neq j$.

(4) $\det A \neq 0$ if and only if A is invertible.

Remark. The geometric interpretation of the determinant. Writing $A = [\mathbf{a}^1, \ldots, \mathbf{a}^n] \in \mathcal{M}(n,n)$, we have the absolute value of $\det A$ is the n-dimensional volume of the parallelepiped spanned by the vectors $\mathbf{a}^1, \ldots, \mathbf{a}^n$.

Cramer's rule and a formula for A^{-1}

Let $A_{n\times n} = [\mathbf{a^1}, \ldots, \mathbf{a}^n]$, then the equation system $A\mathbf{x} = \mathbf{b}$ can be rewritten as

$$x_1\mathbf{a}^1 + \ldots + x_n\mathbf{a}^n = \mathbf{b}.$$

Then for each i we have

$$x_i \det[\mathbf{a}^1, \ldots, \mathbf{a}^n] = \det[\mathbf{a}^1, \ldots, \mathbf{b}, \ldots, \mathbf{a}^n]$$

where $\mathbf{b}$ occurs in the ith place. This is the so-called **Cramer's rule**, which can be used to solve the system of linear equations when $\det A \neq 0$.

Proof.

$$\begin{aligned}
\det[\mathbf{a}^1, \ldots, \mathbf{b}, \ldots, \mathbf{a}^n] &= \det[\mathbf{a}^1, \ldots, \sum x_k\mathbf{a}^k, \ldots, \mathbf{a}^n] \\
&= \sum \det[\mathbf{a}^1, \ldots, x_k\mathbf{a}^k, \ldots, \mathbf{a}^n] \\
&= \sum x_k \det[\mathbf{a}^1, \ldots, \mathbf{a}^k, \ldots, \mathbf{a}^n] \\
&= x_i \det[\mathbf{a}^1, \ldots, \mathbf{a}^i, \ldots, \mathbf{a}^n]
\end{aligned}$$

□

Now assume that $\det A \neq 0$ so there exists $B = (b_{ij}) \in \mathcal{M}(n, n)$ such that $AB = I_n$. Write $\mathbf{e_j} \in \mathbb{R}^n$ for the column vector with 1 in the jth place and zeros elsewhere. We have

$$b_{ij} = \frac{\det[\mathbf{a}^1, \ldots, \mathbf{e}_j, \ldots, \mathbf{a}^n]}{\det A}$$

where the $\mathbf{e}_j$ occurs in the ith position. This can be shown by

$$AB = I \Rightarrow A\mathbf{b^j} = \mathbf{e}_j, \forall j$$

and the Cramer's rule. By expanding $\det[\mathbf{a}^1, \ldots, \mathbf{e}_j, \ldots, \mathbf{a}^n]$ down the ith column using the Laplace expansion, we can show that

$$b_{ij} = \frac{(-1)^{i+j}\det(A_{ji})}{\det A}$$

where the A_{ji} is (j, i) minor of A, obtained from A by deleting the jth row and ith column.

Example 2.4. For example, let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and assume $\det A \neq 0$, then we have

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

2.7 Eigenvectors and Eigenvalues*

Let V be a vector space and $T : V \to V$ be linear transformation represented by the square matrix A with respect to some basis. A vector $\mathbf{v} \in V$ is called an **eigenvector** of T is $\mathbf{v} \neq \mathbf{0}$ and $T(\mathbf{v}) = \lambda\mathbf{v}$ for some $\lambda \in \mathbb{R}$. We call $\lambda \in \mathbb{R}$ an **eigenvalue** of T if $T(\mathbf{v}) = \lambda\mathbf{v}$ for some nonzero vector $\mathbf{v} \in V$.

Using the terminologies we defined in section 2.5, we have the following statements are equivalent.

(1) λ is an eigenvalue of T;

(2) The null space of $T - \lambda I$ is not $\{\mathbf{0}\}$;

(3) The linear transformation $T - \lambda I$ is invertible;

(4) $\det(T - \lambda I) = 0$.

Then we define the **characteristic polynomial** by $\det(A - xI_n)$. We denote the characteristic polynomial of the linear transformation T by $\chi_T(x)$, and of a matrix A by $\chi_A(x)$. Then we have λ is an eigenvalue of T if and only if λ is a root of $\chi_T(x)$, that is $\chi_T(\lambda) = 0$.

Example: An Application on Marxian Economics

Consider a input-output matrix $A = (a_{ij}) \in \mathcal{M}(n, n)$ where a_{ij} is the amount of good i required to produce 1 unit of good j, and the labor vector $\mathbf{l} = (l_1, \ldots, l_n)$ where l_j is the amount of labor required to produce 1 unit of good j. Define the value of good j, λ_j as the total amount of labor directly and indirectly required to produce one unit of good j, then we have

$$\Lambda = \Lambda A + \mathbf{l}$$

where $\Lambda = (\lambda_1, \ldots, \lambda_n)$ is the value vector. Therefore, we have

$$\Lambda = (I - A)^{-1}\mathbf{l}$$

if $(I - A)^{-1}$ exists.

Let $\mathbf{p}$ be the row vector of prices, and $\mathbf{b}$ the column vector of real wage, then $\mathbf{p}A + \mathbf{pbl}$ is the row vector of unit cost. Assume the uniform rate of profit is π, then we have

$$\mathbf{p} = (1 + \pi)(\mathbf{p}A + \mathbf{pbl}) = (1 + \pi)\mathbf{p}(A + \mathbf{bl})$$

Therefore, the price $\mathbf{p}$ is a eigenvector of the matrix $M \equiv A + \mathbf{bl}$, and the rate of profit is determined by the corresponding eigenvalue. To make these variables meaningful, all the prices should be nonnegative. This is guaranteed by the so-called **Perron-Frobenius Theorem** when the matrix M is nonnegative. Some famous results in Marxian economics like the *Fundamental Marxian Theorem* and the *Okishio's Theorem* can be derived in this framework.

3 Rudiments of Analysis

In this section, we will review the basic topics of analysis, including the convergence of sequence, the continuity and differentiability of the real-valued function.

3.1 Convergence of Sequences

A **sequence** of real number is a function $x : \mathbb{N}_+ \to \mathbb{R}$ that assigns a real number $x(n)$ to each $n = 1, 2, \ldots$, denoted by $(x_n) = (x_1, x_2, \ldots,)$.

Example 3.1. Let $x_n = \frac{\sin n}{2n+1}$. Then the sequence looks like

$$(\frac{1}{3}\sin 1, \frac{1}{5}\sin 2, \frac{1}{7}\sin 3 \ldots,)$$

Given the sequence (x_n) and (y_n), we have $(x_n + y_n), (-x_n), (x_n y_n)$ and, provided every y_n is non-zero, (x_n/y_n). Also, we may have (cx_n) for any constant c, and $(|x_n|)$.

We want to analyze how the terms x_n of the sequence behavior as n get *arbitrarily large* and specifically whether or not the terms approach *arbitrarily closely* some *limiting value.*

First, we don't care what the values of the first few terms are, or what the first 100 million terms are. That is, the long-run behavior of the terms of a sequence is not is not affected by chopping the first k terms. For any sequence (x_n) and any $k \in \mathbb{N}$, let $y_n = x_{n+k}$ for all n. We call (y_n) a **tail** of (x_n).

Next, we can capture "arbitrarily close to" via ϵ. From the definition of absolute value of real number in Section , we have, for any $x, L \in \mathbb{R}$ and $\epsilon > 0$,

$$|x - L| < \epsilon \Longleftrightarrow L - \epsilon < x < L + \epsilon$$

That is x lies within a distance ϵ of L. To formulate that x is arbitrarily close to L, we allow $\epsilon > 0$ to be very small.

Now we have a precise, formal definition of convergence. Let (x_n) be a sequence of real numbers and let $L \in \mathbb{R}$. Then we say (x_n) **converges to** L, denoted by $\lim_{n\to\infty} x_n = L$ or $x_n \to L, n \to \infty$, if

$$\forall \epsilon > 0 \quad \exists N \in \mathbb{N} \quad \forall n > N \quad |x_n - L| < \epsilon.$$

When $x_n \to L, n \to \infty$, we say L is the **limit** of (x_n).

For any (x_n), if there exists $L \in \mathbb{R}$ such that $x_n \to L$ as $n \to \infty$, we say (x_n) **converges**; otherwise, we say (x_n) **diverges**. For convenience, we define the two specific cases of divergence. We say x_n **tends to infinity** and write $x_n \to \infty$ as $n \to \infty$ if

$$\forall M \in \mathbb{R}, \exists N \in \mathbb{N}, \forall n \geq N, x_n > M$$

Similarly we can define $x_n \to -\infty$ if

$$\forall M \in \mathbb{R}, \exists n \in \mathbb{N}, \forall n > N, x_n < M$$

If a sequence of real numbers converges, we can show that the limit is unique. The proof of this statement is a good example to show the construction of $\epsilon - N$ proofs.

Example 3.2. Let (x_n) be a sequence and suppose that $x_n \to L_1$ and $x_n \to L_2$ as $n \to \infty$. Then $L_1 = L_2$.

Proof. Suppose $L_1 \neq L_2$. Take $\epsilon = \frac{1}{2}|L_1 - L_2|$. Since $x_n \to L_1$,

$$\exists N_1 \in \mathbb{N}, \forall n > N_1, |x_n - L_1| < \epsilon.$$

Similarly, since $x_n \to L_2$, we have

$$\exists N_2 \in \mathbb{N}, \forall n > N_2, |x_n - L_2| < \epsilon.$$

Let $N = \max\{N_1, N_2\}$, then we have $\forall n > N$,

$$|L_1 - L_2| = |L_1 - x_n + x_n - L_2| \leq |L_1 - x_n| + |x_n - L_2| < \epsilon = |L_1 - L_2|$$

contradicted. $\square$

In the following, we have some more exercises to help you master the construction of $\epsilon - N$ proofs.

Exercise 3.1. Let (x_n) be a sequence and suppose that $x_n \to L$ as $n \to \infty$, show that $|x_n| \to |L|$, as $n \to \infty$.

Hints. Show that $||a| - |b|| \leq |a - b|$ by reversing the Triangle Inequality. $\square$

Also, we can show the *preservation of weak inequalities*: assume that $x_n \to a, y_n \to b$ and $x_n \leq y_n, \forall n$, then $a \leq b$.

Proof. Suppose, on the contrary, $a > b$ and let $\epsilon = \frac{1}{2}(b - a)$. Then we have N_1 and N_2 such that

$$\forall n > N_1, |x_n - a| < \epsilon \Rightarrow x_n > a - \epsilon$$
$$\forall n > N_2, |y_n - b| < \epsilon \Rightarrow y_n < b + \epsilon$$

Therefore, take $N = \max\{N_1, N_2\}$, for $n > N$, we have

$$a - \epsilon < x_n \leq y_n < b + \epsilon$$

then $a - b < 2\epsilon = a - b$ contradicted. □

However, it should be noted that the limit does not preserve the strict inequalities. This can be seen from the following example.

Example 3.3. Let $x_n = 0, y_n = \frac{1}{n}$, then we have $x_n < y_n$, but $x_n \to 0, y_n \to 0$.

Exercise 3.2 (The Sandwiching Rule.)**.** Let $(x_n), (y_n), (z_n)$ be three sequences of real numbers. Suppose $x_n \to L$, $y_n \to L$, and $x_n \leq z_n \leq y_n$ for all n, then show that $z_n \to L$, as $n \to \infty$.

As we can see from the definition of convergence, a sequence that is convergent will get "trapped" into the neighbor of limit after chopping the first finite number of elements. Therefore, it should be **bounded** in the sense that

$$\exists M \in \mathbb{R}, \forall n, |x_n| \leq M.$$

Proof. Assume that $x_n \to L$ and take $\epsilon = 1$. Then there exists N such that

$$\forall n > N, |x_n - L| < 1 \Longrightarrow |x_n| \leq |L| + 1$$

Let $M = \max\{x_1, \ldots, x_N, |L| + 1$, then we have $|x_n| \leq M, \forall n$. □

Since any convergent sequence is bounded, then if a sequence is not bounded, it must diverge. However, the sequence that diverges can be bounded, e.g., $y_n = (-1)^n$. In this case, we usually show the divergence of a sequence by constructing subsequences.

A **subsequence** (y_k) of the sequence (x_n) is defined by a *strictly increasing* mapping $f : \mathbb{N} \to \mathbb{N}$, so that

$$y_k = x_{n_k},$$

where $n_k = f(k)$. The strictly increasing mapping f is the rule of choice by which terms of x_n are selected to form the subsequence. It is strictly increasing to maintain the same order and make sure that each term of x_n is selected once. Formally, the subsequence y_k is the composite of this the strictly increasing mapping f and the function $x(n)$, that is $y = x \circ f : \mathbb{N} \to \mathbb{R}$ with $y(k) = x(f(k))$.

Every subsequence of a convergent sequence is convergent. Furthermore, it converges to the same limit of the the sequence. The contrapositive of this statement is usually used to establish the divergence. That is, if a sequence subsequences which converge to different limits, then the sequence does not converge.

Example 3.4. Let

$$x_n = (-1)^n \frac{n^2}{n^2+1}$$

Then $x_{2n} \to 1$ and $x_{2n+1} \to -1$. Therefore, x_n diverges.

The algebra of limits (AOL)

Suppose $x_n \to a, y_n \to b$ as $n \to \infty$ and c is a constant number, then we have

(1) If $x_n = c$, then $a = c$.

(2) $x_n + y_n \to a + b$.

(3) $cx_n \to ca$.

(4) $x_n - y_n \to a - b$.

(5) $x_n y_n \to ab$.

(6) If $b \neq 0$ and $y_n \neq 0$, then $1/y_n \to 1/b$.

(7) If $b \neq 0$ and $y_n \neq 0$, then $x_n/y_n \to a/b$.

Here we only provide the proof of (5), and the rest are left as exercises for the reader.

Proof of (5). Since y_n convergent, it is bounded. So we can find $M \in \mathbb{R}$ such that $|y_n| < M$. For any $\epsilon > 0$, from $x_n \to a$ and $y_n \to b$, we can find N_1 and N_1 such that

$$\forall n \geq N_1, |x_n - a| < \frac{\epsilon}{2M}$$
$$\forall n \geq N_2, |y_n - b| < \frac{\epsilon}{2|a|}$$

Therefore, for all $n > \max\{N_1, N_2\}$

$$\begin{aligned} |x_n y_n - ab| &= |x_n y_n - a y_n + a y_n - ab| \\ &\leq |x_n - a||y_n| + |a||y_n - b_n| \\ &\leq M|x_n - a| + |a||y_n - b_n| \\ &\leq \epsilon \end{aligned}$$

□

A typical application of AOL is given by the following example.

Example 3.5. Let $x_n = \frac{n^2+n+1}{3n^2+4}$. Then

$$x_n = \frac{n^2(1 + \frac{1}{n} + \frac{1}{n^2})}{3n^2(1 + \frac{4}{3n^2})} = \frac{1}{3}(\frac{1 + \frac{1}{n} + \frac{1}{n^2}}{1 + \frac{4}{3n^2}}) \to \frac{1}{3}(\frac{1+0+0}{1+0}) = \frac{1}{3}$$

Monotonic Sequences

Let (x_n) be a sequence of real number, then it is said to be **monotonic increasing** if $x_n \leq x_{n+1}, \forall n$, and **monotonic decreasing** if $x_n \geq x_{n+1}, \forall n$. It is **monotonic** if it is either monotonic increasing or monotonic decreasing. Then we have the following theorem asserts that any monotonic sequence converges if and only if it is bounded.

Theorem 3.1 (Monotonic Sequence Theorem)**.** *Let (x_n) be a sequence of real numbers.*

(1) Assume (x_n) is monotonic increasing. Then (x_n) converges if and only if it is bounded above, i.e., $\exists M, \forall n, |x_n| \leq M$.

(2) *Assume (x_n) is monotonic decreasing. Then (x_n) converges if and only if it is bounded below.*

We prove the first part, recalling the completeness axiom for real number in Section 1.4 and the condition for supremum in terms of $\epsilon - \delta$ language.

Proof. Let $E = \{x_n \mid n \in \mathbb{N}\}$. Then we have E is a nonempty subset of $\mathbb{R}$ and bounded above since (x_n) is bounded above. By the completeness axiom, $\sup S$ exists. Let $L = \sup S$, then $\forall \epsilon > 0$, $\exists x_N \in E$ such that

$$L - \epsilon < x_N \leq L.$$

Therefore, $\forall n > N$,

$$L - \epsilon < x_N \leq x_n \leq L \Longrightarrow |x_n - L| < \epsilon,$$

since x_n is monotonic increasing. Then we have $x_n \to L$.

For the converse we use the fact that any convergent sequence is bounded. $\square$

The Bolzano-Weierstrass Theorem

Since monotonic real subsequences behave so well, the following fact is welcome: any real sequence has a monotonic subsequence. This is the so-called **scenic viewpoint theorem**. Once it is established, we have the famous result called the **Bolzano-Weierstrass Theorem** asserting that any bounded real sequence has convergent subsequence, by applying the monotonic sequence theorem.

Now we first prove the scenic viewpoint theorem. Let (x_n) be a real sequence. Then (x_n) has a monotonic subsequence. We consider the set of "scenic viewpoint" or "peaks", $V = \{k \in \mathbb{N} \mid m > k \Rightarrow x_m < a_k\}$. That is, given any $k \in V$, looking toward infinity from a point at height x_k, no higher point would impede our view. Then V can be infinite or finite. In both cases, we can construct a monotonic subsequence of (x_n) as follows.

Case 1: V is infinite. Then enumerate the elements of V as $n_1 < n_2 < \ldots$. Then x_{n_k} is a subseqence of (x_n) and

$$r > s \Longrightarrow n_r > n_s \Longrightarrow x_{n_r} < x_{n_s}$$

that is, (x_{n_k}) is monotonic decreasing.

Case 2: V is finite. Then $\exists N \in \mathbb{N}$ such that $\forall k \in V, k < N$. Take $m_1 = N$, then $m_1 \notin V$. Therefore, $\exists m_2 > m_1$ with $x_{m_2} \geq x_{m_1}$ since m_1 is not a "peak". Since $m_2 \notin V$, there exists $m_3 > m_2$ such that $x_{m_3} \geq x_{m_2}$. Proceeding in this way, we can generate a monotonic increasing subsequence (x_{m_k}).

Let (x_n) be a bounded real sequence. By the scenic viewpoint theorem, we know that (x_n) has a monotonic subsequence which is also bounded. Then this subsequence converges, by the Monotonic Sequence Theorem 3.1. Then we have

Theorem 3.2 (Bolzano-Weierstrass Theorem)**.** *Let (x_n) be a bounded real sequence. Then (x_n) has a convergent subseqence.*

Cauchy Sequences.

The idea behind the Cauchy sequence is that if the terms of a sequence are ultimately arbitrarily close to one another, then there should be a value to which they converge. We say a sequence of real number is a **Cauchy sequence** if

$$\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall m, n \geq N, |x_m - x_n| < \epsilon$$

This condition is also called the **Cauchy condition**. About Cauchy sequences, we have the following facts.

(1) A Cauchy sequence is bounded.

(2) A convergent sequence is a Cauchy sequence.

(3) If a Cauchy sequence has a convergent subsequence, the Cauchy sequence itself must also converge the limit of its subsequence.

Proof. The proof is left as an exercise, with solution provided below.

(1) Let (x_n) be a Cauchy sequence. Take $\epsilon = 1$, then $exists N$ such that $\forall m, n \geq N$, we have $|x_m - x_n| < 1$. Thus we get $|x_n| \leq |x_N| + 1$ for all $n \geq N$. Then we have $|x_n| \leq M$ where

$$M = \max\{|a_1|, \ldots, |a_{N-1}|, |a_N| + 1\}.$$

(2) Let $x_n \to L$. Then $\forall \epsilon > 0$, $\exists N \in \mathbb{N}$ such that $\forall k \geq N$, we have $|x_k - L| < \frac{\epsilon}{2}$. Then for any $m, n \geq N$, we have

$$|x_m - x_n| = |x_m - L + L - x_n| \leq |x_m - L| + |x_n - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

(3) Let (x_n) be a Cauchy sequence, and $x_{n_k} \to L$. For any $\epsilon > 0$, $\exists N_1$ such that

$$\forall m, n \geq N_1, |x_m - x_n| < \frac{\epsilon}{2}.$$

Since $x_{n_k} \to L$, $\exists N_2$ such that

$$\forall k \geq N_2, |x_{n_k} - L| < \frac{\epsilon}{2}.$$

Then fix k such that $k \geq \max\{N_1, N_2\}$,then $n_k \geq k \geq N_1$. Therefore,

$$\forall n > N, |x_n - L| = |x_n - x_{n_k} + x_{n_k} - L| < |x_n - x_{n_k}| + |x_{n_k} - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2}.$$

□

The second fact show that any convergent sequence is a Cauchy sequence. The converse is true. Since any Cauchy sequence is bounded by the first fact, then it has convergent subsequence by Theorem 3.2. Together with the last fact, we have it converges. Therefore, we have the **Cauchy Convergence Criterion**.

Theorem 3.3 (Cauchy Convergence Criterion)**.** *Let (x_n) be a sequence of real number, then (x_n) is convergent if and only if (x_n) is a Cauchy sequence.*