HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
FACULTY OF APPLIED MATHEMATICS AND INFORMATICS
ADVANCED TRAINING PROGRAM

Lecture on

ALGEBRA

Assoc. Prof. Dr. Nguyen Thieu Huy

Hanoi 2008

Preface

This *Lecture on Algebra* is written for students of Advanced Training Programs of Mechatronics (from California State University –CSU Chico) and Material Science (from University of Illinois- UIUC). When preparing the manuscript of this lecture, we have to combine the two syllabuses of two courses on Algebra of the two programs (Math 031 of CSU Chico and Math 225 of UIUC). There are some differences between the two syllabuses, e.g., there is no module of algebraic structures and complex numbers in Math 225, and no module of orthogonal projections and least square approximations in Math 031, etc. Therefore, for sake of completeness, this lecture provides all the modules of knowledge which are given in both syllabuses. Students will be introduced to the theory and applications of matrices and systems of linear equations, vector spaces, linear transformations, eigenvalue problems, Euclidean spaces, orthogonal projections and least square approximations, as they arise, for instance, from electrical networks, frameworks in mechanics, processes in statistics and linear models, systems of linear differential equations and so on. The lecture is organized in such a way that the students can comprehend the most useful knowledge of linear algebra and its applications to engineering problems.

We would like to thank Prof. Tran Viet Dung for his careful reading of the manuscript. His comments and remarks lead to better appearance of this lecture. We also thank Dr. Nguyen Huu Tien, Dr. Tran Xuan Tiep and all the lecturers of Faculty of Applied Mathematics and Informatics for their inspiration and support during the preparation of the lecture.

Hanoi, October 20, 2008

Assoc. Prof. Dr. Nguyen Thieu Huy

Contents

Chapter 1: Sets

I. Concepts and Basic Operations

1.1. Concepts of sets: A set is a collection of objects or things. The objects or things in the set are called elements (or member) of the set.

Examples:

- A set of students in a class.

- A set of countries in ASEAN group, then Vietnam is in this set, but China is not.

- The set of real numbers, denoted by **R.**

1.2. Basic notations: Let E be a set. If x is an element of E, then we denote by $x \in E$ (pronounce: x belongs to E). If x is not an element of E, then we write $x \notin E$.

We use the following notations:

$\exists$: "there exists"

$\exists!$: "there exists a unique"

$\forall$: " for each" or "for all"

$\Rightarrow$: "implies"

$\Leftrightarrow$: "is equivalent to" or "if and only if"

1.3. Description of a set: Traditionally, we use upper case letters A, B, C and set braces to denote a set. There are several ways to describe a set.

a) **Roster notation (or listing notation):** We list all the elements of a set in a couple of braces; e.g., A = {1,2,3,7} or B = {Vietnam, Thailand, Laos, Indonesia, Malaysia, Brunei, Myanmar, Philippines, Cambodia, Singapore}.

b) **Set–builder notation:** This is a notation which lists the rules that determine whether an object is an element of the set.

Example: The set of real solutions of the inequality $x^2 \leq 2$ is

$G = \{x \mid x \in \mathbf{R} \text{ and } -\sqrt{2} \leq x \leq \sqrt{2}\} = [-\sqrt{2}, \sqrt{2}]$

The notation "|" means "such that".

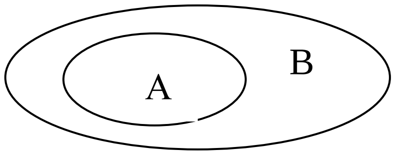
c) **Venn diagram:** Some times we use a closed figure on the plan to indicate a set. This is called Venn diagram.

1.4 Subsets, empty set and two equal sets:

a) Subsets: The set A is called a subset of a set B if from $x \in A$ it follows that $x \in B$. We then denote by $A \subset B$ to indicate that A is a subset of B.

By logical expression: $A \subset B \Leftrightarrow (x \in A \Rightarrow x \in B)$

By Venn diagram:



b) Empty set: We accept that, there is a set that has no element, such a set is called an empty set (or void set) denoted by $\varnothing$.

Note: For every set A, we have that $\varnothing \subset A$.

c) Two equal sets: Let A, B be two set. We say that A equals B, denoted by A = B, if $A \subset B$ and $B \subset A$. This can be written in logical expression by

$A = B \Leftrightarrow (x \in A \Leftrightarrow x \in B)$

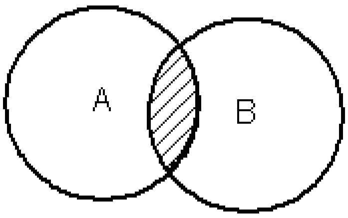
1.5. Intersection: Let A, B be two sets. Then the intersection of A and B, denoted by $A \cap B$, is given by:

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

This means that

$x \in A \cap B \Leftrightarrow (x \in A \text{ and } x \in B)$.

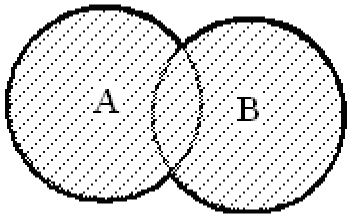
By Venn diagram:



1.6. Union: Let A, B be two sets, the union of A and B, denoted by $A \cup B$, and given by $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$. This means that

$x \in A \cup B \Leftrightarrow (x \in A \text{ or } x \in B)$.

By Venn diagram:



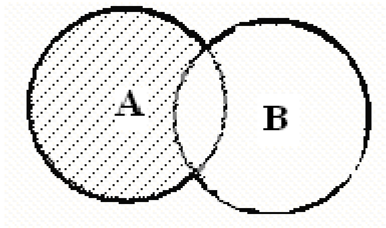
1.7. Subtraction: Let A, B be two sets: The subtraction of A and B, denoted by A\B (or A–B), is given by

$A \backslash B = \{x \mid x \in A \text{ and } x \notin B\}$

This means that:

$x \in A \backslash B \Leftrightarrow (x \in A \text{ and } x \notin B)$.

By Venn diagram:



1.8. Complement of a set:

Let A and X be two sets such that $A \subset X$. The complement of A in X, denoted by $C_X A$ (or A' when X is clearly understood), is given by

$C_X A = X \setminus A = \{x \mid x \in X \text{ and } x \notin A)\}$

$= \{x \mid x \notin A\}$ (when X is clearly understood)

Examples: Consider $X = \mathbf{R}$; $A = [0,3] = \{x \mid x \in \mathbf{R} \text{ and } 0 \le x \le 3\}$

$B = [-1, 2] = \{x | x \in \mathbf{R} \text{ and } -1 \le x \le 2\}$.

Then,

1. $A \cap B = \{x \in \mathbf{R} \mid 0 \le x \le 3 \text{ and } -1 \le x \le 2\} =$

$= \{x \in \mathbf{R} \mid 0 \le x \le -1\} = [0, -1]$

2. $A \cup B = \{x \in \mathbf{R} \mid 0 \le x \le 3 \text{ or } -1 \le x \le 2\}$

$= \{x \in \mathbf{R} \mid -1 \le x \le 3\} = [-1,3]$

3. $A \setminus B = \{x \in \mathbf{R} \mid 0 \le x \le 3 \text{ and } x \notin [-1,2]\}$

$= \{x \in \mathbf{R} \mid 2 \le x \le 3\} = [2,3]$

4. $A' = \mathbf{R} \setminus A = \{x \in \mathbf{R} \mid x < 0 \text{ or } x > 3\}$

II. Set equalities

Let A, B, C be sets. The following set equalities are often used in many problems related to set theory.

1. $A \cup B = B \cup A$; $A \cap B = B \cap A$ (Commutative law)

2. $(A \cup B) \cup C = A \cup (B \cup C)$; $(A \cap B) \cap C = A \cap (B \cap C)$ (Associative law)

3. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$; $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributive law)

4. $A \setminus B = A \cap B'$, where $B' = C_X B$ with a set X containing both A and B.

PROOF: Since the proofs of these equalities are relatively simple, we prove only one equality (3), the other ones are left as exercises.

To prove (3), We use the logical expression of the equal sets.

$$x \in A \cup (B \cap C) \Leftrightarrow \left[\begin{array}{l} x \in A \\ x \in B \cap C \end{array}\right.$$

$$\Leftrightarrow \left[\begin{array}{l} x \in A \\ \begin{cases} x \in B \\ x \in C \end{cases} \end{array}\right. \Leftrightarrow \begin{cases} \left[\begin{array}{l} x \in A \\ x \in B \end{array}\right. \\ \left[\begin{array}{l} x \in A \\ x \in C \end{array}\right. \end{cases}$$

$$\Leftrightarrow \begin{cases} x \in A \cup B \\ x \in A \cup C \end{cases}$$

$$\Leftrightarrow x \in (A \cup B) \cap (A \cup C)$$

This equivalence yields that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

The proofs of other equalities are left for the readers as exercises.

III. Cartesian products

3.1. Definition:

1. Let A, B be two sets. The Cartesian product of A and B, denoted by $A \times B$, is given by

$$A \times B = \{(x,y) \mid (x \in A) \text{ and } (y \in B)\}.$$

2. Let $A_1, A_2 \ldots A_n$ be given sets. The Cartesian Product of $A_1, A_2 \ldots A_n$, denoted by $A_1 \times A_2 \times \ldots \times A_n$, is given by $A_1 \times A_2 \times \ldots . A_n = \{(x_1, x_2 \ldots .. x_n) \mid x_i \in A_i = 1,2 \ldots ., n\}$

In case, $A_1 = A_2 = \ldots = A_n = A$, we denote

$$A_1 \times A_2 \times \ldots \times A_n = A \times A \times A \times \ldots \times A = A^n.$$

3.2. Equality of elements in a Cartesian product:

1. Let $A \times B$ be the Cartesian Product of the given sets A and B. Then, two elements (a, b) and (c, d) of $A \times B$ are equal if and only if $a = c$ and $b=d$.

In other words, $(a, b) = (c, d) \Leftrightarrow \begin{cases} a = c \\ b = d \end{cases}$

2. Let $A_1 \times A_2 \times \ldots \times A_n$ be the Cartesian product of given sets $A_1, \ldots A_n$.

Then, for $(x_1, x_2 \ldots x_n)$ and $(y_1, y_2 \ldots y_n)$ in $A_1 \times A_2 \times \ldots . \times A_n$, we have that

$$(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_n) \Leftrightarrow x_i = y_i \ \forall\, i = 1, 2 \ldots ., n.$$

Chapter 2: Mappings

I. Definition and examples

1.1. Definition: Let X, Y be nonempty sets. A *mapping* with domain X and range Y, is an ordered triple (X, Y, f) where f assigns to each $x \in X$ a well-defined $f(x) \in Y$. The statement that (X, Y, f) is a mapping is written by $f: X \to Y$ (or $X \xrightarrow{f} Y$).

Here, "well-defined" means that for each $x \in X$ there corresponds one and only one $f(x) \in Y$.

A mapping is sometimes called a *map* or a *function*.

1.2. Examples:

1. $f: \mathbf{R} \to \mathbf{R}$; $f(x) = \sin x\ \forall x \in \mathbf{R}$, where **R** is the set of real numbers,

2. $f: X \to X$; $f(x) = x\ \forall x \in X$. This is called the identity mapping on the set X, denoted by I_X

3. Let X, Y be nonvoid sets, and $y_0 \in Y$. Then, the assignment $f: X \to Y$; $f(x) = y_0\ \forall x \in X$, is a mapping. This is called a constant mapping.

1.3. Remark: We use the notation

$$f: X \to Y$$
$$x \mapsto f(x)$$

to indicate that f(x) is assigned to x.

1.4. Remark: Two mappings $X \xrightarrow{f} Y$ and $U \xrightarrow{g} V$ are equal if and only if X = U, Y=V, and $f(x) = g(x)\ \forall x \in X$. Then, we write f = g.

II. Compositions

2.1. Definition: Given two mappings: $f: X \to Y$ and $g: Y \to W$ (or shortly, $X \xrightarrow{f} Y \xrightarrow{g} W$), we define the mapping $h: X \to W$ by $h(x) = g(f(x))\ \forall x \in X$. The mapping h is called the composition of g and f, denoted by $h = g \circ f$, that is, $(g \circ f)(x) = g(f(x))\ \forall x \in X$.

2.2. Example: $\mathbf{R} \xrightarrow{f} \mathbf{R}_+ \xrightarrow{g} \mathbf{R}_-$, here $\mathbf{R}_+ = [0, \infty)$ and $\mathbf{R}_- = (-\infty, 0]$.

$f(x) = x^2\ \forall x \in \mathbf{R}$; and $g(x) = -x\ \forall x \in \mathbf{R}_+$. Then, $(g \circ f)(x) = g(f(x)) = -x^2$.

2.3. Remark: In general, $f \circ g \neq g \circ f$.

Example: $\mathbf{R} \xrightarrow{f} \mathbf{R} \xrightarrow{g} \mathbf{R}$; $f(x) = x^2$; $g(x) = 2x + 1\ \forall x \in \mathbf{R}$.

Then $(f_og)(x) = f(g(x)) = (2x+1)^2\ \forall x \in \mathbf{R}$, and $(g_of)(x) = g(f(x)) = 2x^2 + 1\ \forall x \in \mathbf{R}$.

Clearly, $f_og \neq g_of$.

III. Image and Inverse Images

Suppose that $f : X \to Y$ is a mapping.

3.1. Definition: For $S \subset X$, the image of S in a subset of Y, which is defined by

$$f(S) = \{f(s) | s \in S\} = \{y \in Y | \exists s \in S \text{ with } f(s) = y\}$$

Example: $f: \mathbf{R} \to \mathbf{R}$, $f(x) = x^2\ \forall x \in \mathbf{R}$.

$S = [-1,2] \subset \mathbf{R}$; $f(S) = \{f(s) | s \in [-1, 2]\} = \{s^2 | s \in [-1, 2]\} = [0, 4]$.

3.2. Definition: Let $T \subset Y$. Then, the inverse image of T is a subset of X, which is defined by $f^{-1}(T) = \{x \in X | f(x) \in T\}$. So, $x \in f^{-1}(T)$ if and only if $f(x) \in T$.

Example: $f: \mathbf{R} \setminus \{2\} \to \mathbf{R}$; $f(x) = \dfrac{x+1}{x-2}\ \forall x \in \mathbf{R} \setminus \{2\}$.

$S = (-\infty, -1] \subset \mathbf{R}$; $f^{-1}(S) = \{x \in \mathbf{R} \setminus \{2\} | f(x) \leq -1\}$

$= \{x \in \mathbf{R} \setminus \{2\} | \dfrac{x+1}{x-2} \leq -1\} = [-1/2, 2)$.

3.3. Definition: Let $f: X \to Y$ be a mapping. The image of the domain X, f(X), is called the image of f, denoted by Imf. That is to say,

$$\text{Im} f = f(X) = \{f(x) | x \in X\} = \{y \in Y | \exists x \in X \text{ with } f(x) = y\}.$$

3.4. Properties of images and inverse images: Let $f: X \to Y$ be a mapping; let A, B be subsets of X and C, D be subsets of Y. Then, the following properties hold.

1) $f(A \cup B) = f(A) \cup f(B)$

2) $f(A \cap B) \subset f(A) \cap f(B)$

3) $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$

4) $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$

PROOF: We shall prove (1) and (3), the other properties are left as exercises.

(1) : Since $A \subset A \cup B$ and $B \subset A \cup B$, it follows that

$$f(A) \subset f(A \cup B) \text{ and } f(B) \subset f(A \cup B).$$

These inclusions yield $f(A) \cup f(B) \subset f(A\cup B)$.

Conversely, take any $y \in f(A\cup B)$. Then, by definition of an Image, we have that, there exists an $x \in A \cup B$, such that $y = f(x)$. But, this implies that $y = f(x) \in f(A)$ (if $x \in A$) or $y=f(x) \in f(B)$ (if $x \in B$). Hence, $y \in f(A) \cup f(B)$. This yields that $f(A\cup B) \subset f(A) \cup f(B)$. Therefore, $f(A\cup B) = f(A) \cup f(B)$.

(3): $x\in f^{-1}(C\cup D) \Leftrightarrow f(x) \in C\cup D \Leftrightarrow (f(x) \in C \text{ or } f(x) \in D)$

$\Leftrightarrow (x \in f^{-1}(C) \text{ or } x \in f^{-1}(D)) \Leftrightarrow x \in f^{-1}(C)\cup f^{-1}(D)$

Hence, $f^{-1}(C\cup D) = f^{-1}(D)) = f^{-1}(C)\cup f^{-1}(D)$.

IV. Injective, Surjective, Bijective, and Inverse Mappings

4.1. Definition: Let $f: X \to Y$ be a mapping.

a. The mapping is called surjective (or onto) if Imf = Y, or equivalently,

$$\forall y\in Y, \exists x\in X \text{ such that } f(x) = y.$$

b. The mapping is called injective (or one–to–one) if the following condition holds:

$$\text{For } x_1,x_2\in X \text{ if } f(x_1) = f(x_2), \text{ then } x_1 = x_2.$$

This condition is equivalent to:

$$\text{For } x_1,x_2\in X \text{ if } x_1\neq x_2, \text{ then } f(x_1) \neq f(x_2).$$

c. The mapping is called bijective if it is surjective and injective.

Examples:

1. $\mathbf{R} \xrightarrow{f} \mathbf{R}$; $f(x) = \sin x\ \forall x \in \mathbf{R}$.

This mapping is not injective since $f(0) = f(2\pi) = 0$. It is also not surjective, because, $f(\mathbf{R}) = \text{Imf} = [-1, 1] \neq \mathbf{R}$

2. $f: \mathbf{R} \to [-1,1]$, $f(x) = \sin x\ \forall x\in\mathbf{R}$. This mapping is surjective but not injective.

3. $f: \left[-\frac{\pi}{2},\frac{\pi}{2}\right] \to \mathbf{R}$; $f(x) = \sin x\ \forall x\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. This mapping is injective but not surjective.

4. $f: \left[-\frac{\pi}{2},\frac{\pi}{2}\right] \to [-1,1]$; $f(x) = \sin x \forall x\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. This mapping is bijective.

4.2. Definition: Let f: $X \to Y$ be a bijective mapping. Then, the mapping g: $Y \to X$ satisfying $g_o f = I_X$ and $f_o g = I_Y$ is called the inverse mapping of f, denoted by $g = f^{-1}$.

For a bijective mapping f: $X \to Y$ we now show that there is a unique mapping g: $Y \to X$ satisfying $g_o f = I_X$ and $f_o g = I_Y$.

In fact, since f is bijective we can define an assignment g : $Y \to X$ by g(y) = x if f(x) = y. This gives us a mapping. Clearly, $g(f(x)) = x\ \forall x \in X$ and $f(g(y)) = y\ \forall y \in Y$. Therefore, $g_o f = I_X$ and $f_o g = I_Y$.

The above g is unique is the sense that, if h: $Y \to X$ is another mapping satisfying $h_o f = I_X$ and $f_o h = I_Y$, then $h(f(x)) = x = g(f(x))\ \forall x \in X$. Then, $\forall y \in Y$, by the bijectiveness of f, $\exists!\ x \in X$ such that $f(x) = y \Rightarrow h(y) = h(f(x)) = g(f(x)) = g(y)$. This means that h = g.

Examples:

1. f: $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to [-1,1]$; $f(x) = \sin x\ \forall x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

This mapping is bijective. The inverse mapping $f^{-1}: [-1,1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is denoted by $f^{-1} = \arcsin$, that is to say, $f^{-1}(x) = \arcsin x\ \forall x \in [-1,1]$. We also can write:

$$\arcsin(\sin x) = x\ \forall x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right];\ \text{and } \sin(\arcsin x) = x\ \forall x \in [-1,1]$$

2. f: $\mathbf{R} \to (0,\infty)$; $f(x) = e^x\ \forall x \in \mathbf{R}$.

The inverse mapping is $f^{-1}: (0, \infty) \to \mathbf{R}$, $f^{-1}(x) = \ln x\ \forall x \in (0,\infty)$. To see this, take $(f_o f^{-1})(x) = e^{\ln x} = x\ \forall x \in (0,\infty)$; and $(f^{-1}{}_o f)(x) = \ln e^x = x\ \forall x \in \mathbf{R}$.

Chapter 3: Algebraic Structures and Complex Numbers

I. Groups

1.1. Definition: Suppose that G is non empty set and φ: GxG → G is a mapping. Then, φ is called a binary operation; and we will write φ(a,b) = a∗b for each (a,b) ∈ GxG.

Examples:

1) Consider G = **R**; "∗" = "+" (the usual addition in **R)** is a binary operation defined by

+: **R** x **R** → **R**

(a,b) ⟼ a + b

2) Take G = **R**; "∗" = "•" (the usual multiplication in **R**) is a binary operation defined by

•: **R** x **R** → **R**

(a,b) ⟼ a • b

3. Take G = {f: X → X | f is a mapping}:= Hom (X) for X ≠ ∅.

The composition operation "o" is a binary operation defined by:

o: Hom(X) x Hom(X) → Hom(X)

(f,g) ⟼ fog

1.2. Definition:

a. A couple (G, ∗), where G is a nonempty set and ∗ is a binary operation, is called an algebraic structure.

b. Consider the algebraic structure (G, ∗) we will say that

(b1) ∗ is associative if (a∗b) ∗c = a∗(b∗c) ∀a, b, and c in G

(b2) ∗ is commutative if a∗b = b∗a ∀a, b ∈G

(b3) an element e ∈ G is the neutral element of G if

e∗a = a∗e = a ∀a∈G

Examples:

1. Consider (**R**,+), then "+" is associative and commutative; and 0 is a neutral element.

2. Consider (**R**, •), then "•" is associative and commutative and 1 is an neutral element.

3. Consider (Hom(X), o), then "o" is associative but not commutative; and the identity mapping I_X is an neutral element.

1.3. **Remark:** If a binary operation is written as +, then the neutral element will be denoted by 0_G (or 0 if G is clearly understood) and called the null element.

If a binary operation is written as ·, then the neutral element will be denoted by 1_G (or 1) and called the identity element.

1.4. Lemma: Consider an algebraic structure (G, ∗). Then, if there is a neutral element e ∈G, this neutral element is unique.

PROOF: Let e' be another neutral element. Then, e = e∗e' because e' is a neutral element and e' = e∗e' because e is a neutral element of G. Therefore e = e'.

1.5. Definition: The algebraic structure (G, ∗) is called a group if the following conditions are satisfied:

1. ∗ is associative
2. There is a neutral element e∈G
3. ∀a∈ G, ∃a'∈ G such that a∗a' = a'∗a = e

Remark: Consider a group (G, ∗).

a. If ∗ is written as +, then (G,+) is called an additive group.

b. If ∗ is written as •, then (G, •) is called a multiplicative group.

c. If ∗ is commutative, then (G, ∗) is called abelian group (or commutative group).

d. For a ∈ G, the element a'∈G such that a∗a' = a'∗a=e, will be called the opposition of a, denoted by

$a' = a^{-1}$, called inverse of a, if ∗ is written as • (multiplicative group)

a' = - a, called negative of a, if ∗ is written as + (additive group)

Examples:

1. (**R**, +) is abelian additive group.

2. (**R\{0}**, •) is abel multiplicative group.

3. Let X be nonempty set; End(X) = {f: X → X | f is bijective}.

Then, (End(X), o) is noncommutative group with the neutral element is I_X, where o is the composition operation.

1.6. Proposition:

Let (G, ∗) be a group. Then, the following assertions hold.

1. For a ∈G, the inverse a^{-1} of a is unique.

2. For a, b, c ∈ G we have that

a∗c = b∗c ⇒ a = b

c∗a = c∗b ⇒ a = b

(Cancellation law in Group)

3. For a, x, b ∈ G, the equation a∗x = b has a unique solution x = a^{-1}∗b.

Also, the equation x∗a = b has a unique solution x = b∗a^{-1}

PROOF:

1. Let a' be another inverse of a. Then, a'∗a = e. It follows that

(a'∗a) ∗a^{-1} = a'∗ (a∗a^{-1}) = a'∗e = a'.

2. a∗c = a∗b ⇒ a^{-1}∗ (a∗c) a^{-1}∗ (a∗b) ⇒ (a^{-1}∗a) ∗c = (a^{-1}∗a) ∗b ⇒ e∗c = e∗b ⇒ c = b.

Similarly, c∗a = b∗a ⇒ c = b.

The proof of (3) is left as an exercise.

II. Rings

2.1. Definition: Consider triple (V, +, •) where V is a nonempty set; + and • are binary operations on V. The triple (V, +, •) is called a ring if the following properties are satisfied:

(V, +) is a commutative group

Operation "•" is associative

$\forall$ a,b,c $\in$ V we have that (a + b)•c = a•c + b•c, and c•(a + b) = c•a + c•b

V has identity element 1_V corresponding to operation "•" , and we call 1_V the multiplication identity.

If, in addition, the multiplicative operation is commutative then the ring (V, +, •) is called a commutative ring.

2.2. Example: (**R**, +, •) with the usual additive and multiplicative operations, is a commutative ring.

2.3. Definition: We say that the ring is trivial if it contains only one element, V = $\{O_V\}$.

Remark: If V is a nontrivial ring, then $1_V \neq O_V$.

2.4. Proposition: Let (V, +, •) be a ring. Then, the following equalities hold.

1. $a.O_V = O_V.a = O_V$

2. a • (b – c) = a•b – a•c, where b – c is denoted for b + (-c)

3. (b–c) •a = b•a – c•a

III. Fields

3.1. Definition: A triple (V, +, •) is called a field if (V, +, •) is a commutative, nontrivial ring such that, if $a \in$ V and $a \neq O_V$ then a has a multiplicative inverse $a^{-1} \in$ V.

Detailedly, (V, +, •) is a field if and only if the following conditions hold:

(V, +) is a commutative group,

the multiplicative operation is associative and commutative,

$\forall$a,b,c$\in$V we have that (a + b)•c = a•c + a•b,

there is multiplicative identity $1_V \neq O_V$; and if a$\in$V, $a \neq O_V$, then $\exists a^{-1} \in$V, a^{-1}•a = 1_V.

3.2. Examples: (**R**, +, •); (**Q**, +, •) are fields.

IV. The field of complex numbers

Equations without real solutions, such as $x^2 + 1 = 0$ or $x^2 - 10x + 40 = 0$, were observed early in history and led to the introduction of complex numbers.

4.1. Construction of the field of complex numbers: On the set $\mathbf{R}^2$, we consider additive and multiplicative operations defined by

(a,b) + (c,d) = (a + c, b + d)

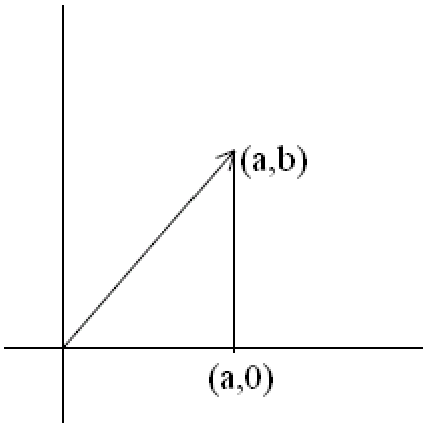
(a,b) • (c,d) = (ac - bd, ad + bc)

Then, ($\mathbf{R}^2$, +, •) is a field. Indeed,

1) ($\mathbf{R}^2$, +, •) is obviously a commutative, nontrivial ring with null element (0, 0) and identity element (1,0) ≠ (0,0).

2) Let now (a,b) ≠ (0,0), we see that the inverse of (a,b) is (c,d) = $\left(\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2}\right)$ since (a,b) • $\left(\frac{a}{a^2+b^2}, -\frac{b}{a^2+b^2}\right)$ = (1,0).

We can present $\mathbf{R}^2$ in the plane



We remark that if two elements (a,0), (c,0) belong to horizontal axis, then their sum (a,0) + (c,0) = (a + c, 0) and their product (a,0)•(c,0) = (ac, 0) are still belong to the horizontal axis, and the addition and multiplication are operated as the addition and multiplication in the set of real numbers. This allows to identify each element on the horizontal axis with a real number, that is (a,0) = a ∈ **R**.

Now, consider i = (0,1). Then, i^2 = i.i = (0, 1). (0, 1) = (-1, 0) = -1. With this notation, we can write: for (a,b)∈ $\mathbf{R}^2$

(a,b) = (a,0) • (1,0) + (b,0) • (0,1) = a + bi

We set **C** = {a + bi | a,b ∈**R** and $i^2 = -1$} and call **C** the set of complex numbers. It follows from above construction that (**C**, +, •) is a field which is called the **field of complex numbers**.

The additive and multiplicative operations on **C** can be reformulated as.

(a+bi) + (c+di) = (a+c) + (b+d)i

(a+bi) • (c+di) = ac + bdi^2 + (ad + bc)i = (ac – bd) + (ad + bc)i

(Because i^2=-1).

Therefore, the calculation is done by usual ways as in **R** with the attention that i^2 = -1.

The representation of a complex number z ∈ **C** as z = a + bi for a,b∈**R** and i^2 = -1, is called the canonical form (algebraic form) of a complex number z.

4.2. Imaginary and real parts: Consider the field of complex numbers **C**. For z∈ **C**, in canonical form, z can be written as

$$z = a + b\mathrm{i}, \text{ where } a, b \in \mathbf{R} \text{ and } \mathrm{i}^2 = -1.$$

In this form, the real number *a* is called the real part of z; and the real number *b* is called the imaginary part. We denote by *a* = Rez and b = Imz. Also, in this form, two complex numbers $z_1 = a_1 + b_1i$ and $z_2 = a_2 + b_2i$ are equal if and only if $a_1 = a_2$; $b_1 = b_2$, that is, $\mathrm{Re}z_1=\mathrm{Re}z_2$ and $\mathrm{Im}z_1 = \mathrm{Im}z_2$.

4.3. Subtraction and division in canonical forms:

1) Subtraction: For $z_1 = a_1 + b_1i$ and $z_2 = a_2 + b_2i$, we have

$$z_1 – z_2 = a_1 – a_2 + (b_1 – b_2)i.$$

Example: 2 + 4i – (3 + 2i) = -1 + 2i.

2) Division: By definition, $\frac{z_1}{z_2} = z_1(z_2^{-1})$ ($z_2 \neq 0$).

For $z_2 = a_2 + b_2i$, we have $z_2^{-1} = \frac{a_2}{a_2^2 + b_2^2} - \frac{b_2}{a_2^2 + b_2^2} i$. Therefore,

$\frac{a_1 + b_1 i}{a_2 + b_2} = (a_1 + b_1 i).\left(\frac{a_2}{a_2^2 + b_2^2} - \frac{b_2 i}{a_2^2 + b_2^2}\right)$. We also have the following

practical rule: To compute $\frac{z_1}{z_2} = \frac{a_1 + b_1 i}{a_2 + b_2 i}$ we multiply both denominator and numerator by $a_2 – b_2i$, then simplify the result. Hence,

$$\frac{a_1+b_1i}{a_2+b_2i}=\frac{a_1+b_1i}{a_2+b_2i}.\frac{a_2-b_2i}{a_2-b_2i}=\frac{a_1a_2+b_1b_2+(a_2b_1-a_1b_2)i}{a_2^2+b_2^2}$$

Example: $\frac{2-7i}{8+3i}=\frac{2-7i}{8+3i}.\frac{8-3i}{8-3i}=\frac{-5-62i}{73}=\frac{-5}{73}-\frac{62}{73}i$

4.4. Complex plane: Complex numbers admit two natural geometric interpretations.

First, we may identify the complex number x + yi with the point (x,y) in the plane (see Fig.4.2). In this interpretation, each real number a, or x+ 0.*i*, is identified with the point (x,0) on the horizontal axis, which is therefore called the real axis. A number 0 + y*i*, or just y*i*, is called a pure imaginary number and is associated with the point (0,y) on the vertical axis. This axis is called the imaginary axis. Because of this correspondence between complex numbers and points in the plane, we often refer to the xy-plane as the complex plane.

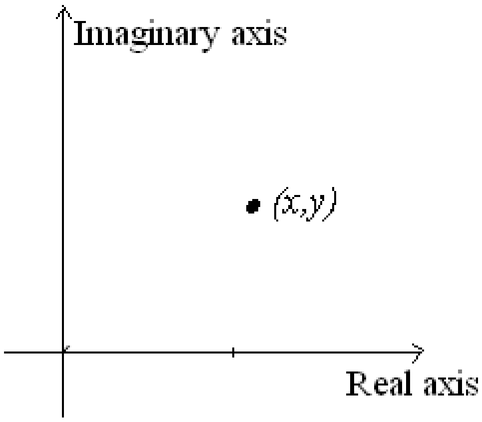


Figure 4.2

When complex numbers were first noticed (in solving polynomial equations), mathematicians were suspicious of them, and even the great eighteen–century Swiss mathematician Leonhard Euler, who used them in calculations with unparalleled proficiency, did not recognize then as "legitimate" numbers. It was the nineteenth–century German mathematician Carl Friedrich Gauss who fully appreciated their geometric significance and used his standing in the scientific community to promote their legitimacy to other mathematicians and natural philosophers.

The second geometric interpretation of complex numbers is in terms of vectors. The complex numbers z = x + yi may be thought of as the vector x $\vec{i}$ +y $\vec{j}$ in the plane, which may in turn be represented as an arrow from the origin to the point (x,y), as in Fig.4.3.

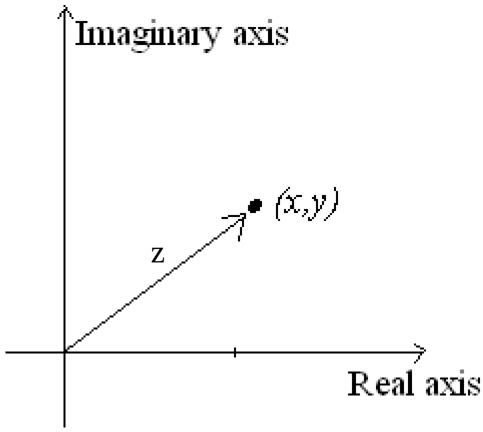


Fig. 4.3. Complex numbers as vectors in the plane

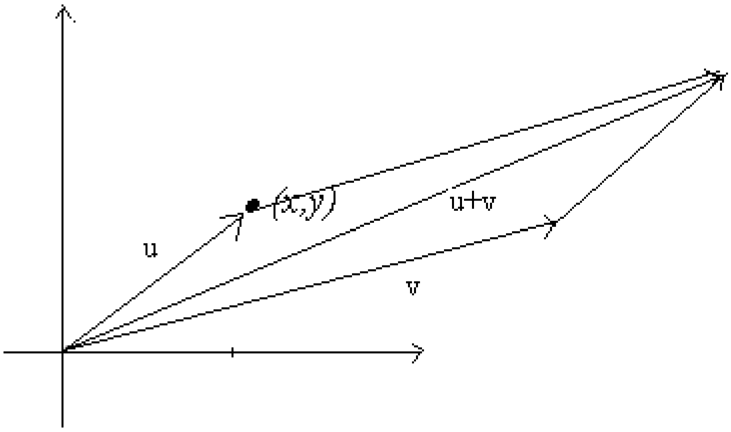


Fig.4.4. Parallelogram law for addition of complex numbers

The first component of this vector is Rez, and the second component is Imz. In this interpretation, the definition of addition of complex numbers is equivalent to the parallelogram law for vector addition, since we add two vectors by adding the respective component (see Fig.4.4).

4.5. Complex conjugate: Let z = x +*i*y be a complex number then the complex conjugate $\bar{z}$ of z is defined by $\bar{z}$ = x – *i*y.

It follows immediately from definition that

Rez = x = (z + $\bar{z}$)/2; and Imz = y = (z - $\bar{z}$)/2

We list here some properties related to conjugation, which are easy to prove.

1. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ ∀z_1, z_2 in **C**

2. $\overline{z_1 . z_2} = \overline{z_1} . \overline{z_2}$ ∀z_1, z_2 in **C**

3. $\overline{\left(\frac{z_1}{z_2}\right)} = \frac{\overline{z_1}}{\overline{z_2}}$ $\forall z_1, z_2$ in **C**

4. if $\alpha \in$ **R** and $z \in$ C, then $\overline{\alpha.z} = \alpha.\overline{z}$

5. For $z \in$ **C** we have that, $z \in$ **R** if and only if $\overline{z} = z$.

4.6. Modulus of complex numbers: For $z = x + iy \in$ C we define $|z| = \sqrt{x^2 + y^2}$, and call it modulus of z. So, the modulus $|z|$ is precisely the length of the vector which represents z in the complex plane.

$z = x + i.y = \overrightarrow{OM}$

$|z| = |\overrightarrow{OM}| = \sqrt{x^2 + y^2}$

4.7. Polar (trigonometric) forms of complex numbers:

The canonical forms of complex numbers are easily used to add, subtract, multiply or divide complex numbers. To do more complicated operations on complex numbers such as taking to the powers or roots, we need the following form of complex numbers.

Let we start by employ the polar coordination: For $z = x + iy$

$0 \neq z = x + iy = \overrightarrow{OM} = (x,y)$. Then we can put $\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}$

where
$$r = |z| = \sqrt{x^2 + y^2} \qquad \text{(I)}$$

and θ is angle between $\overrightarrow{OM}$ and the real axis, that is, the angle θ is defined by

$$\begin{cases} \cos\theta = \dfrac{x}{\sqrt{x^2 + y^2}} \\ \sin\theta = \dfrac{y}{\sqrt{x^2 + y^2}} \end{cases} \qquad \text{(II)}$$

The equalities (I) and (II) define the unique couple (r, θ) with $0 \leq \theta < 2\pi$ such that $\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}$. From this representation we have that

$$z = x + iy = \sqrt{x^2+y^2}\left(\frac{x}{\sqrt{x^2+y^2}} + i\frac{y}{\sqrt{x^2+y^2}}\right)$$

$= |z|(\cos\theta + i\sin\theta)$.

Putting $|z| = r$ we obtain

$$z = r(\cos\theta + i\sin\theta).$$

This form of a complex number z is called polar (or trigonometric) form of complex numbers, where $r = |z|$ is the modulus of z; and the angle θ is called the argument of z, denoted by $\theta = \arg(z)$.

Examples:

1) $z = 1 + i = \sqrt{2}\left(\frac{1}{\sqrt{2}} + i\frac{1}{\sqrt{2}}\right) = \sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$

2) $z = 3 - 3\sqrt{3}i = 6\left(\frac{1}{2} - \frac{\sqrt{3}}{2}i\right) = 6\left(\cos\left(-\frac{\pi}{3}\right) + i\sin\left(-\frac{\pi}{3}\right)\right)$

Remark: Two complex number in polar forms $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$; $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$ are equal if and only if

$$\begin{cases} r_1 = r_2 \\ \theta_1 = \theta_2 + 2k\pi \end{cases} \quad \forall k \in \mathbf{Z}.$$

4.8. Multiplication, division in polar forms:

We now consider the multiplication and division of complex numbers represented in polar forms.

1) Multiplication: Let $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$; $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$

Then, $z_1.z_2 = r_1r_2[\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 + i(\cos\theta_1\sin\theta_2 + \cos\theta_2\sin\theta_1)]$. Therefore,

$$z_1.z_2 = r_1r_2[\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)]. \qquad (4.1)$$

It follows that $|z_1.z_2| = |z_1|.|z_2|$ and $\arg(z_1.z_2) = \arg(z_1) + \arg(z_2)$.

2) Division: Take $z = \frac{z_1}{z_2} \Leftrightarrow z_1 = z.z_2 \Rightarrow |z_1| = |z|.|z_2|$ (for $z_2 \neq 0$)

$\Leftrightarrow |z| = \dfrac{|z_1|}{|z_2|}$

Moreover, $\arg(z_1) = \arg z + \arg z_2 \Leftrightarrow \arg(z) = \arg(z_1) - \arg(z_2)$.

Therefore, we obtain that, for $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2) \neq 0$.

We have that $$\frac{z_1}{z_2} = z = \frac{r_1}{r_2}[\cos(\theta_1-\theta_2) + i\sin(\theta_1-\theta_2)]. \tag{4.2}$$

Example: $z_1 = -2 + 2i$; $z_2 = 3i$.

We first write $z_1 = 2\sqrt{2}\left(\cos\frac{3\pi}{4} + i\sin\frac{5\pi}{5}\right); z_2 = 3\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right)$

Therefore, $z_1.z_2 = 6\sqrt{2}\left(\cos\frac{5\pi}{4} + i\sin\frac{5\pi}{4}\right)$

$$\frac{z_1}{z_2} = \frac{2\sqrt{2}}{3}\left(\cos\frac{\pi}{4} + i\sin\frac{5\pi}{4}\right)$$

3) Integer power: for $z = r(\cos\theta + i\sin\theta)$, by equation (4.1), we see that

$$z^2 = r^2(\cos 2\theta + i\sin 2\theta).$$

By induction, we easily obtain that $z^n = r^n(\cos n\theta + i\sin n\theta)$ for all $n \in \mathbf{N}$.

Now, take $z^{-1} = \frac{1}{z} = \frac{1}{r}\left(\cos(-\theta) + i\sin(-\theta)\right) = r^{-1}\left(\cos(-\theta) + i\sin(-\theta)\right)$.

Therefore $z^{-2} = \left(z^{-1}\right)^2 = r^{-2}\left(\cos(-2\theta) + i\sin(-2\theta)\right)$.

Again, by induction, we obtain: $z^{-n} = r^{-n}(\cos(-n\theta) + i\sin(-n\theta))\ \forall n \in \mathbf{N}$.

This yields that $z^n = r^n(\cos n\theta + i\sin n\theta)$ for all $n \in \mathbf{Z}$.

A special case of this formula is the formula of de Moivre:

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta \qquad \forall n \in \mathbf{N}.$$

Which is useful for expressing $\cos n\theta$ and $\sin n\theta$ in terms of $\cos\theta$ and $\sin\theta$.

4.9. Roots: Given $z \in \mathbf{C}$, and $n \in \mathbf{N}^* = \mathbf{N} \setminus \{0\}$, we want to find all $w \in \mathbf{C}$ such that $w^n = z$.

To find such w, we use the polar forms. First, we write $z = r(\cos\theta + i\sin\theta)$ and $w = \rho(\cos\varphi + i\sin\varphi)$. We now determine ρ and φ. Using relation $w^n = z$, we obtain that

$\rho^n(\cos n\varphi + i\sin n\varphi) = r(\cos\theta + i\sin\theta)$

$$\Leftrightarrow \begin{cases} \rho^n = r \\ n\varphi = \theta + 2k\pi; k \in Z \end{cases} \Leftrightarrow \begin{cases} \rho = \sqrt[n]{r} \text{ (real positive root of r)} \\ \varphi = \dfrac{\theta + 2k\pi}{n}; k \in Z \end{cases}$$

Note that there are only n distinct values of w, corresponding to k = 0, 1…, n-1. Therefore, w is one of the following values

$$\left\{ \sqrt[n]{r}\left(\cos\frac{\theta + 2k\pi}{n} + i\sin\frac{\theta + 2k\pi}{n} \right) \middle| k = 0,1\ldots,n-1 \right\}$$

For z = r(cosθ + isinθ) we denote by

$$\sqrt[n]{z} = \left\{ \sqrt[n]{r}\left(\cos\frac{\theta + 2k\pi}{n} + i\sin\frac{\theta + 2k\pi}{n} \right) \middle| \; k = 0,1,2\ldots n-1 \right\}$$

and call $\sqrt[n]{z}$ the set of all n^{th} roots of complex number z.

For each w ∈ C such that $w^n = z$, we call w an n^{th} root of z and write w ∈ $\sqrt[n]{z}$.

Examples:

1. $\sqrt[3]{1} = \sqrt[3]{1(\cos 0 + i\sin 0)}$ (complex roots)

$$= \left\{ 1\left(\cos\frac{2k\pi}{3} + i\sin\frac{2k\pi}{3} \right) \middle| \; k = 0,1,2 \right\}$$

$$= \left\{ 1; -\frac{1}{2} + \frac{\sqrt{3}}{2}i; -\frac{1}{2} - \frac{\sqrt{3}}{2}i \right\}$$

2. Square roots: For Z = r(cosθ+isinθ) ∈ C, we compute the complex square roots

$$\sqrt{z} = \sqrt{r(\cos\theta + i\sin\theta)} =$$

$$= \left\{ \sqrt{r}\left(\cos\frac{\theta + 2k\pi}{2} + i\sin\frac{\theta + 2k\pi}{2} \right) \middle| \; k = 0,1 \right\}$$

$= \left\{ \sqrt{r}\left(\cos\frac{\theta}{2} + i\sin\frac{\theta}{2} \right); -\sqrt{r}\left(\cos\frac{\theta}{2} + i\sin\frac{\theta}{2} \right) \right\}$. Therefore, for z ≠ 0; the set $\sqrt{z}$ contains two opposite values {w, -w}. Shortly, we write $\sqrt{z} = \pm w$ (for $w^2 = z$).

Also, we have the practical formula, for z = x + iy,

$$\sqrt{z} = \pm\left[\sqrt{\frac{1}{2}(|z| + x)} + \left((sign\,y).i\sqrt{\frac{1}{2}(|z| - x)}\right)\right] \qquad (4.3)$$

where sign$y$ = $\begin{cases} 1 & if \quad y \geq 0 \\ -1 & if \quad y < 0 \end{cases}$.

3. Complex quadratic equations: $az^2 + bz + c = 0$; $a,b,c \in \mathbf{C}$; $a \neq 0$.

By the same way as in the real-coefficient case, we obtain the solutions $z_{1,2} = \frac{-b \pm w}{2a}$ where $w^2 = \Delta = b^2 - 4ac \in \mathbf{C}$.

Concretely, taking $z^2 - (5+i)z + 8 + i = 0$, we have that $\Delta = -8+6i$. By formula (4.3), we obtain that

$$\sqrt{\Delta} = \pm w = \pm(1 + 3i)$$

Then, the solutions are $z_{1,2} = \frac{5 + i \pm (1 + 3i)}{2}$ or $z_1 = 3 + 2i$; $z_2 = 2 - i$.

We finish this chapter by introduce (without proof) the Fundamental Theorem of Algebra.

4.10. Fundamental Theorem of Algebra: Consider the polynomial equation of degree n in **C**:

$$a_nx^n + a_{n-1}x^{n-1} + \ldots + a_1x + a_0 = 0;\ a_i \in \mathbf{C}\ \forall i = 0, 1, \ldots n.\ (a_n \neq 0) \qquad (4.4)$$

Then, Eq (4.4) has n solutions in **C**. This means that, there are complex numbers x_1, $x_2 \ldots x_n$, such that the left-hand side of (4.4) can be factorized by

$$a_nx^n + a_{n-1}x^{n-1} + \ldots + a_0 = a_n(x - x_1)(x - x_2)\ldots(x - x_n).$$

Chapter 4: Matrices

I. Basic concepts

Let **K** be a field (say, **K** = **R** or **C**).

1.1. Definition: A matrix is a rectangular array of numbers (of a field **K**) enclosed in brackets. These numbers are called entries or elements of the matrix

Examples 1: $\begin{bmatrix} 2 & 0.4 & 8 \\ 5 & -32 & 0 \end{bmatrix}$; $\begin{bmatrix} 6 \\ 1 \end{bmatrix}$; $\begin{bmatrix} 1 & 5 & 4 \end{bmatrix}$; $\begin{bmatrix} 2 & 3 \\ 1 & 8 \end{bmatrix}$.

Note that we sometimes use the brackets (.) to indicate matrices.

The notion of matrices comes from variety of applications. We list here some of them

Sales figures: Let a store have products I, II, III. Then, the numbers of sales of each product per day can be represented by a matrix

$$\begin{matrix} \\ I \\ II \\ III \end{matrix}\begin{bmatrix} Monday & Tuesday & Wednesday & Thursday & Friday \\ 10 & 20 & 15 & 3 & 4 \\ 0 & 12 & 7 & 3 & 5 \\ 0 & 9 & 6 & 8 & 9 \end{bmatrix}$$

- Systems of equations: Consider the system of equations

$$\begin{cases} 5x - 10y + z = 2 \\ 6x - 3y - 2z = 0 \\ 2x + y - 4z = 0 \end{cases}$$

Then the coefficients can be represented by a matrix

$$\begin{bmatrix} 5 & -10 & 1 \\ 6 & -3 & -2 \\ 2 & 1 & 4 \end{bmatrix}$$

We will return to this type of coefficient matrices later.

1.2. Notations: Usually, we denote matrices by capital letter A, B, C or by writing the general entry, thus

A = $[a_{jk}]$, so on…

By an mxn matrix we mean a matrix with m rows (called row vectors) and n columns (called column vectors). Thus, an mxn matrix A is of the form:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

Example 2: On the example 1 above we have the 2x3; 2x1; 1x3 and 2x2 matrices, respectively.

In the double subscript notation for the entries, the first subscript is the row, and the second subscript is the column in which the given entry stands. Then, a_{23} is the entry in row 2 and column 3.

If m = n, we call A an n-square matrix. Then, its diagonal containing the entries a_{11}, a_{22},…, a_{nn} is called the main diagonal (or principal diagonal) of A.

1.3. Vectors: A vector is a matrix that has only one row – then we call it a row vector, or only one column - then we call it a column vector. In both case, we call its entries the components. Thus,

$A = [a_1\ a_2 \dots a_n]$ – row vector

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$ - column vector.

1.4. Transposition: The transposition A^T of an mxn matrix A = $[a_{jk}]$ is the nxm matrix that has the first row of A as its first column, the second row of A as its second column,…, and the m^{th} row of A as its m^{th} column. Thus,

for $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$ we have that $A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$

Example 3: $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 0 & 7 \end{bmatrix}; A^T = \begin{bmatrix} 1 & 4 \\ 2 & 0 \\ 3 & 7 \end{bmatrix}$

II. Matrix addition, scalar multiplication

2.1. Definition:

1. Two matrices are said to have the same size if they are both mxn.

2. For two matrices A = $[a_{jk}]$; B = $[b_{jk}]$ we say A = B if they have the same size and the corresponding entries are equal, that is, $a_{11} = b_{11}$; $a_{12} = b_{12}$, and so on…

2.2. Definition: Let A, B be two matrices having the same sizes. Then their sum, written A + B, is obtained by adding the corresponding entries. (Note: Matrices of different sizes can not be added)

Example: For A = $\begin{bmatrix} 2 & 3 & -1 \\ 1 & 4 & 2 \end{bmatrix}; B = \begin{bmatrix} 2 & 0 & 5 \\ 7 & 1 & 4 \end{bmatrix}$

A + B = $\begin{bmatrix} 2+2 & 3+0 & (-1)+5 \\ 1+7 & 4+1 & 2+4 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 \\ 8 & 5 & 6 \end{bmatrix}$

2.3. Definition: The product of an mxn matrix A = $[a_{jk}]$ and a number *c* (or scalar *c*), written *c*A, is the mxn matrix *c*A = $[ca_{jk}]$ obtained by multiplying each entry in A by *c.*

Here, (-1)A is simply written –A and is called negative of A; (-k)A is written – kA, also A +(-B) is written A – B and is called the difference of A and B.

Example:

For A = $\begin{bmatrix} 2 & 5 \\ -1 & 4 \\ 3 & -7 \end{bmatrix}$ we have 2A = $\begin{bmatrix} 4 & 10 \\ -2 & 8 \\ 6 & 14 \end{bmatrix}$; and - A = $\begin{bmatrix} -2 & -5 \\ 1 & -4 \\ -3 & 7 \end{bmatrix}$;

0A = $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

2.3. Definition: An mxn zero matrix is an mxn matrix with all entries zero – it is denoted by O.

Denoted by $M_{mxn}(\mathbf{R})$ the set of all mxn matrices with the entries being real numbers. The following properties are easily to prove.

2.4. Properties:

1. ($M_{mxn}(\mathbf{R})$, +) is a commutative group. Detailedly, the matrix addition has the following properties.

$(A+B) + C = A + (B+C)$

$A + B = B + A$

$A + O = O + A = A$

$A + (-A) = O$ (written $A – A = O$)

2. For the scalar multiplication we have that (α, β are numbers)

$\alpha(A + B) = \alpha A + \alpha B$

$(\alpha+\beta)A = \alpha A + \beta A$

$(\alpha\beta)A = \alpha(\beta A)$ (written $\alpha\beta A$)

$1A = A$.

3. Transposition: $(A+B)^T = A^T+B^T$

$(\alpha A)^T=\alpha A^T$.

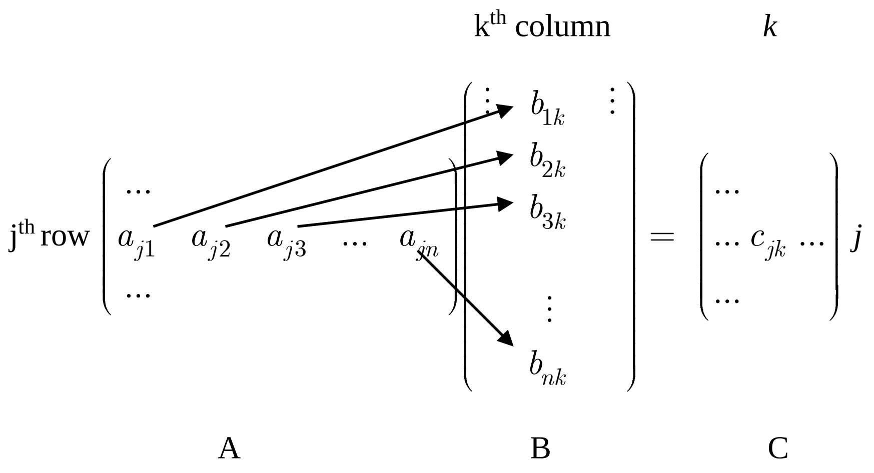
III. Matrix multiplications

3.1. Definition: Let $A = [a_{jk}]$ be an mxn matrix, and $B = [b_{jk}]$ be an nxp matrix. Then, the product C = A.B (in this order) is an mxp matrix defined by

$C = [c_{jk}]$, with the entries:

$c_{jk} = a_{j1}b_{1k} + a_{j2}b_{2k} + \ldots + a_{jn}b_{nk} = \sum_{l=1}^{n} a_{jl}b_{lk}$ where j = 1, 2,…, m; k = 1,2,…,p.

That is, multiply each entry in j^{th} row of A by the corresponding entry in the k^{th} column of B and then add these n products. Briefly, "multiplication of rows into columns". We can illustrate by the figure



Note: AB is defined only if the number of columns of A is equal to the number of rows of B.

Example: $\begin{bmatrix}1 & 4\\ 2 & 2\\ 3 & 0\end{bmatrix}\begin{bmatrix}2 & 5\\ 1 & -1\end{bmatrix} = \begin{bmatrix}1\times 2+4\times 1 & 1\times 5+4\times(-1)\\ 2\times 2+2\times 1 & 2\times 5+2\times(-1)\\ 3\times 2+0\times(-1) & 3\times 5+0\times(-1)\end{bmatrix} =$

$= \begin{bmatrix}6 & 1\\ 6 & 8\\ 6 & 15\end{bmatrix}$.

Exchanging the order, then $\begin{bmatrix}2 & 5\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 4\\ 2 & 2\\ 3 & 0\end{bmatrix}$ is not defined

Remarks:

1) The matrix multiplication is not commutative, that is, AB ≠BA in general.

Examples:

A = $\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$; B = $\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$, then

AB = **O** and BA = $\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$. Clearly, AB≠BA.

2) The above example also shows that, AB = **O** does not imply A = **O** or B = **O**.

3.2 Properties: Let A, B, C be matrices and k be a number.

a) (kA)B = k(AB) = A(kB) (written k AB)

b) A(BC) = (AB)C (written ABC)

c) (A+B).C = AC + BC

d) C (A+B) = CA+CB

provided, A,B and C are matrices such that the expression on the left are defined.

IV. Special matrices

4.1. Triangular matrices: A square matrix whose entries above the main diagonal are all zero is called a lower triangular matrix. Meanwhile, an upper triangular matrix is a square matrix whose entries below the main diagonal are all zero.

Example: $A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 4 \\ 0 & 0 & 3 \end{bmatrix}$ - Upper triangular matrix

$B = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 7 & 0 \\ 2 & 0 & 0 \end{bmatrix}$ - Lower triangular matrix

4.2. Diagonal matrices: A square matrix whose entries above and below the main diagonal are all zero, that is $a_{jk} = 0\ \forall j \neq k$ is called a diagonal matrix.

Example: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$

4.3. Unit matrix: A unit matrix is the diagonal matrix whose entries on the main diagonal are all equal to 1. We denote the unit matrix by I_n (or I) where the subscript n indicates the size nxn of the unit matrix.

Example: $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$; $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Remarks: 1) Let A $\in$ M_{mxn} (R) – set of all mxn matrix whose entries are real numbers. Then,

$$AI_n = A = I_mA.$$

2) Denote by $M_n(\mathbf{R}) = M_{nxn}(\mathbf{R})$, then, $(M_n(\mathbf{R}), +, \bullet)$ is a noncommutative ring where + and $\bullet$ are the matrix addition and multiplication, respectively.

4.4. Symmetric and antisymmetric matrices:

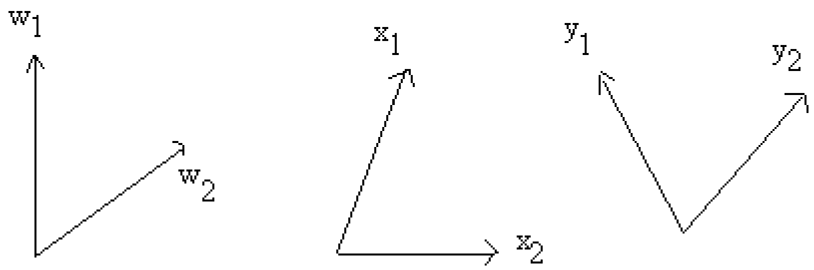
A square matrix A is called symmetric if A^T=A, and it is called anti-symmetric (or skew-symmetric) if A^T = -A.

4.5. Transposition of matrix multiplication:

$(AB)^T = B^TA^T$ provided AB is defined.

4.6. A motivation of matrix multiplication:

Consider the transformations (e.g. rotations, translations…)



The first transformation is defined by $\begin{cases} x_1 = a_{11}w_1 + a_{12}w_2 \\ x_2 = a_{21}w_1 + a_{22}w_2 \end{cases}$ (I)

or, in matrix form $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = A \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$.

The second transformation is defined by $\begin{cases} y_1 = b_{11}x_1 + b_{12}x_2 \\ y_2 = b_{21}x_1 + b_{22}x_2 \end{cases}$ (II)

or, in matrix form $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = B \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

To compute the formula for the composition of these two transformations we substitute (I) in to (II) and obtain that

$$\begin{cases} y_1 = c_{11}w_1 + c_{12}w_2 \\ y_2 = c_{21}w_1 + c_{22}w_2 \end{cases} \text{ where } \begin{cases} c_{11} = b_{11}a_{11} + b_{12}a_{21} \\ c_{12} = b_{11}a_{12} + b_{12}a_{22} \\ c_{21} = b_{21}a_{11} + b_{22}a_{21} \\ c_{22} = b_{21}a_{12} + b_{22}a_{22} \end{cases}.$$

This yields that: $C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = BA$; and $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = BA\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

However, if we use the matrix multiplication, we obtain immediately that

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = B\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = B.A\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

Therefore, the matrix multiplication allows to simplify the calculations related to the composition of the transformations.

V. Systems of Linear Equations

We now consider one important application of matrix theory. That is, application to systems of linear equations. Let us start by some basic concepts of systems of linear equations.

5.1. Definition: A *system of* m *linear equations* in n unknowns $x_1, x_2, \ldots, x_n$ is a set of equations of the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases} \tag{5.1}$$

Where, the a_{jk}; $1 \le j \le m$, $1 \le k \le n$, are given numbers, which are called the coefficients of the system. The b_i, $1 \le i \le m$, are also given numbers. Note that the system (5.1) is also called a *linear system of equations*.

If b_i, $1 \le i \le m$, are all zero, then the system (5.1) is called a homogeneous system. If at least one b_k is not zero, then (5.1) in called a nonhomogeneous system.

A solution of (5.1) is a set of numbers $x_1, x_2\dots,x_n$ that satisfy all the m equations of (5.1). A solution vector of (5.1) is a column vector X = $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ whose components constitute a solution of (5.1). If the system (5.1) is homogeneous, it has at least one trivial solution $x_1 = 0, x_2 = 0,...,x_n = 0$.

5.2. Coefficient matrix and augmented matrix:

We write the system (5.1) in the matrix form: AX = B,

where A = [a_{jk}] is called the coefficient matrix; X = $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and B = $\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ are column vectors.

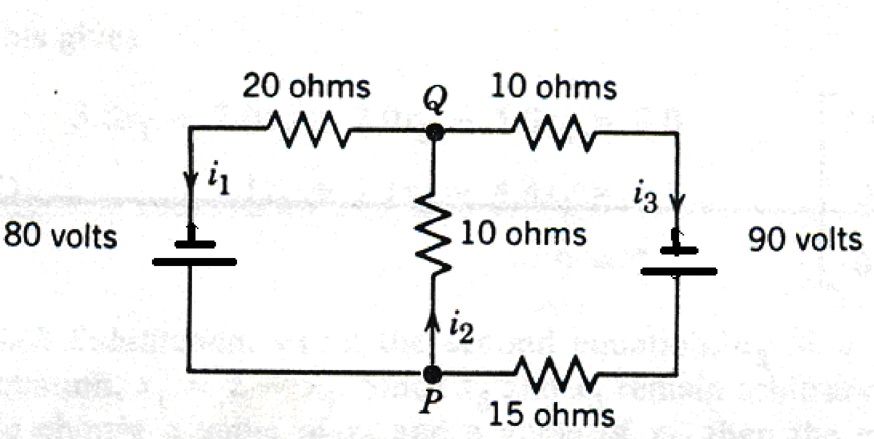
The matrix $\tilde{A} = [A \vdots B]$ is called the augmented matrix of the system (5.1). $\tilde{A}$ is obtained by augmenting A by the column B. We note that $\tilde{A}$ determines system (5.1) completely, because it contains all the given numbers appearing in (5.1).

VI. Gauss Elimination Method

We now study a fundamental method to solve system (5.1) using operations on its augmented matrix. This method is called Gauss elimination method. We first consider the following example from electric circuits.

6.1. Examples:

Example 1: Consider the electric circuit



Label the currents as shown in the above figure, and choose directions arbitrarily. We use the following Kirchhoffs laws to derive equations for the circuit:

+ Kirchhoff's current law (KCL) at any node of a circuit, the sum of the inflowing currents equals the sum of the outflowing currents.

+ Kirchhoff's voltage law (KVL). In any closed loop, the sum of all voltage drops equals the impressed electromotive force.

Applying KCL and KVL to above circuit we have that

Node P: $i_1 - i_2 + i_3 = 0$

Node Q: $- i_1 + i_2 - i_3 = 0$

Right loop: $10i_2 + 25i_3 = 90$

Left loop: $20i_1 + 10i_2 = 80$

Putting now $x_1 = i_1$; $x_2 = i_2$; $x_3 = i_3$ we obtain the linear system of equations

$$\begin{cases} x_1 - x_2 + x_3 = 0 \\ -x_1 + x_2 - x_3 = 0 \\ 10x_2 + 25x_3 = 90 \\ 20x_1 + 10x_2 = 80 \end{cases} \tag{6.1}$$

This system is so simple that we could almost solve it by inspection. This is not the point. The point is to perform a systematic method – the Gauss elimination – which will work in general, also for large systems. It is a reduction to "triangular form" (or, precisely, echelon form-see Definition 6.2 below) from which we shall then readily obtain the values of the unknowns by "back substitution".

We write the system and its augmented matrix side by side:

Equations

Pivot → $\boxed{x_1}$ $-x_2 + x_3 = 0$

Eliminate → $\boxed{-x_1}$ $+x_2 - x_3 = 0$

$10x_2+25x_3 = 90$

$20x_1$ $+10x_2 = 80$

Augmented matrix:

$$\tilde{A} = \begin{bmatrix} 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{bmatrix}$$

First step: Elimination of x_1

Call the first equation the pivot equation and its x_1 – term the pivot in this step, and use this equation to eliminate x_1 (get rid of x_1) in other equations. For this, do these operations.

Add the pivot equation to the second equation;

Subtract 20 times the pivot equation from the fourth equation.

This corresponds to row operations on the augmented matrix, which we indicate behind the new matrix in (6.2). The result is

$$\begin{cases} x_1 - x_2 + x_3 = 0 \\ 0 = 0 \\ 10x_2 + 25x_3 = 90 \\ 30x_2 - 20x_3 = 80 \end{cases} \quad \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \end{bmatrix} \begin{matrix} \\ \text{Row2} + \text{Row1} \to \text{Row2} \\ \text{Row4} - 20\times\text{Row1} \to \text{Row4} \\ \\ \end{matrix} \tag{6.2}$$

Second step: Elimination of x_2

The first equation, which has just served as pivot equation, remains untouched. We want to take the (new) second equation as the next pivot equation. Since it contain no x_2-term (needed as the next pivot, in fact, it is 0 = 0) – first we have to change the order of equations (and corresponding rows of the new matrix) to get a nonzero pivot. We put the second equation (0 = 0) at the end and move the third and the fourth equations one place up. We get

$x_1 - x_2 + x_3 = 0$

Pivot → $\boxed{10x_2}$ $+25x_3 = 90$

Eliminate → $\boxed{30x_2}$ $-20x_3 = 80$

$0 = 0$

Corresponding to

$$\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

To eliminate x_2, do

Subtract 3 times the pivot equation from the third equation, the result is

$x_1 - x_2 + x_3 = 0$

$10x_2 + 25x_3 = 90$

$-95x_3 = -190$

$0 = 0$

$$\begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 0 & -95 & -190 \\ 0 & 0 & 0 & 0 \end{bmatrix} Row3 - 3\times Row2 \to Row3 \qquad (6.3)$$

Back substitution: Determination of x_3, x_2, x_1.

Working backward from the last equation to the first equation the solution of this "triangular system" (6.3), we can now readily find x_3, then x_2 and then x_1:

$-95x_3 = -190$ | $i_3 = x_3 = 2$ (amperes)

$10x_2 + 25x_3 = 90$ $\Rightarrow$ $i_2 = x_2 = 4$ (amperes)

$x_1 - x_2 + x_3 = 0$ | $i_1 = x_1 = 2$ (amperes)

Note: A system (5.1) is called overdetermined if it has more equations than unknowns, as in system (6.1), determined if m = n, and underdetermined if (5.1) has fewer equations than unknowns.

Example 2: Gauss elimination for an underdetermined system. Consider

$$\begin{cases} 3x_1 + 2x_2 + 2x_3 - 5x_4 = 8 \\ 6x_1 + 15x_2 + 15x_3 - 54x_4 = 27 \\ 12x_1 - 3x_2 - 3x_3 + 24x_4 = 21 \end{cases}$$

The augmented matrix is: $\begin{bmatrix} 3 & 2 & 2 & -5 & 8 \\ 6 & 15 & 15 & -54 & 27 \\ 12 & -3 & -3 & 24 & 21 \end{bmatrix}$

1st step: elimination of x_1

$$\begin{bmatrix} 3 & 2 & 2 & -5 & 8 \\ 0 & 11 & 11 & -44 & 11 \\ 0 & -11 & -11 & 44 & -11 \end{bmatrix} \begin{matrix} Row2 - 2\times Row1 \to Row2 \\ Row3 - 4\times Row1 \to Row3 \end{matrix}$$

2nd step: Elimination of x_2

$$\begin{bmatrix} 3 & 2 & 2 & -5 & 8 \\ 0 & 11 & 11 & -44 & 11 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{Row3 – Row1} \to \text{Row3}$$

Back substitution: Writing the matrix into the system we obtain

$3x_1 + 2x_2 + \;\; 2x_3 - 5x_4 \;\; = 8$

$11x_2 + 11x_3 - 44x_4 = 11$

We can divide both sides of the second equation to obtain equivalent system:

$3x_1 + 2x_2 + \;\; 2x_3 - 5x_4 \;\; = 8$

$x_2 + \;\; x_3 - \; 4x_4 \;\; = 1$

From the second equation, $x_2 = 1 - x_3 + 4x_4$. From this and the first equation, we obtain that $x_1 = 2 - x_4$. Since x_3, x_4 remain arbitrary, we have infinitely many solutions, if we choose a value of x_3 and a value of x_4, then the corresponding values of x_1 and x_2 are uniquely determined.

Example 3: What will happen if we apply the Gauss elimination to a linear system that has no solution? The answer is that in this case the method will show this fact by producing a contradiction – for instance, consider

$$\begin{cases} 3x_1 + 2x_2 + x_3 = 3 \\ 2x_1 + x_2 + x_3 = 0 \\ 6x_1 + 2x_2 + 4x_3 = 6 \end{cases}$$

The augmented matrix is $\begin{bmatrix} 3 & 2 & 1 & 3 \\ 2 & 1 & 1 & 0 \\ 6 & 2 & 4 & 6 \end{bmatrix}$

$$\to \begin{bmatrix} 3 & 2 & 1 & 3 \\ 0 & \frac{-1}{3} & \frac{1}{3} & -2 \\ 0 & -2 & 2 & 0 \end{bmatrix} \to \begin{bmatrix} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{bmatrix}$$

The last row correspond to the last equation which is 0 = 12. This is a contradiction yielding that the system has no solution.

The form of the system and of the matrix in the last step of Gauss elimination is called the echelon form. Precisely, we have the following definition.

6.2. Definition: A matrix is of echelon form if it satisfies the following conditions:

i) All the zero rows, if any, are on the bottom of the matrix

ii) In the nonzero row, each leading nonzero entry is to the right of the leading nonzero entry in the preceding row.

Correspondingly, A system is called an echelon system if its augmented matrix is an echelon matrix.

Example: A = $\begin{bmatrix} 1 & -2 & 3 & -4 & 5 \\ 0 & 2 & 1 & 5 & 7 \\ 0 & 0 & 0 & 7 & 8 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ is an echelon matrix

6.3. Note on Gauss elimination: At the end of the Gauss elimination (before the back substitution) the reduced system will have the echelon form:

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\
\tilde{a}_{2j_2}x_{j_2} + \dots + \tilde{a}_{2n}x_n &= \tilde{b}_2 \\
\ddots \\
\tilde{a}_{rj_r}x_{j_r} + \dots + \tilde{a}_{rn}x_n &= \tilde{b}_r \\
0 &= \tilde{b}_{r+1} \\
0 &= 0 \\
&\vdots \\
0 &= 0
\end{aligned}$$

where $r \le m$; $1 < j_2 < \dots < j_r$ and $a_{11} \neq 0$; $\tilde{a}_{2j_2} \neq 0, \dots, \tilde{a}_{rj_r} \neq 0$.

From this, there are three possibilities in which the system has

a) no solution if $r < m$ and the number $\tilde{b}_{r+1}$ is not zero (see example 3)

b) precisely one solution if $r = n$ and $\tilde{b}_{r+1}$, if present, is zero (see example 1)

c) infinitely many solution if $r < n$ and $\tilde{b}_{r+1}$, if present, is zero.

Then, the solutions are obtained as follows:

+) First, determine the so–called free variables which are the unknowns that are not leading in any equations (i.e, x_k is free variable $\Leftrightarrow x_k \notin \{x_1, x_{j_2} \dots, x_{j_r}\}$

+) Then, assign arbitrary values for free variables and compute the remain unknowns $x_1, x_{j_2}, \dots, x_{j_r}$ by back substitution (see example 2).

6.4. Elementary row operations:

To justify the Gauss elimination as a method of solving linear systems, we first introduce the two related concepts.

Elementary operations for equations:

1. Interchange of two equations
2. Multiplication of an equation by a nonzero constant
3. Addition of a constant multiple of one equation to another equation.

To these correspond the following

Elementary row operations of matrices:

1. Interchange of two rows (denoted by $R_i \leftrightarrow R_j$)
2. Multiplication of a row by a nonzero constant: ($kR_i \rightarrow R_i$)
3. Addition of a constant multiple of one row to another row

$$(R_i + kR_j \rightarrow R_i)$$

So, the Gauss elimination consists of these operations for pivoting and getting zero.

6.5. Definition: A system of linear equations S_1 is said to be row equivalent to a system of linear equations S_2 if S_1 can be obtained from S_2 by (finitely many) elementary row operations.

Clearly, the system produced by the Gauss elimination at the end is row equivalent to the original system to be solved. Hence, the desired justification of the Gauss elimination as a solution method now follows from the subsequent theorem, which implies that the Gauss elimination yields all solutions of the original system.

6.6. Theorem: Row equivalent systems of linear equations have the same sets of solutions.

PROOF: The interchange of two equations does not alter the solution set. Neither does the multiplication of the new equation a nonzero constant c, because multiplication of the new equation by $1/c$ produces the original equation. Similarly for the addition of an equation αE_i to an equation E_j, since by adding $-\alpha E_i$ to the equation resulting from the addition we get back the original equation.

Chapter 5: Vector spaces

I. Basic concepts

1.1. Definition: Let **K** be a field (e.g., **K** = **R** or **C**), V be nonempty set. We endow two operations as follows.

Vector addition +: $V \times V \to V$

$(u, v) \mapsto u + v$

Scalar multiplication : $\mathbf{K} \times V \to V$

$(\lambda, v) \mapsto \lambda v$

Then, V is called a vector space over **K** if the following axioms hold.

1) $(u + v) + w = u + (v + w)$ for all $u, v, w \in V$

2) $u + v = v + u$ for all $u, v \in V$

3) there exists a null vector, denoted by $O \in V$, such that $u + O = u$ for all $u, v \in V$.

4) for each $u \in V$, there is a unique vector in V denoted by $-u$ such that $u + (-u) = O$ (written: $u - u = O$)

5) $\lambda(u+v) = \lambda u + \lambda v$ for all $\lambda \in \mathbf{K}$; $u, v \in V$

6) $(\lambda+\mu) u = \lambda u + \mu u$ for all $\lambda, \mu \in \mathbf{K}$; $u \in V$

7) $\lambda(\mu u) = (\lambda\mu)u$ for all $\lambda, \mu \in \mathbf{K}$; $u \in V$

8) $1.u = u$ for all $u \in V$ where 1 is the identity element of **K**.

Remarks:

1. Elements of V are called vectors.

2. The axioms (1)– (4) say that (V, +) is a commutative group.

1.2. Examples:

1. Consider $\mathbf{R}^3 = \{(x, y, z) | x, y, z \in \mathbf{R}\}$ with the vector addition and scalar multiplication defined as usual:

$(x_1, y_1, z_1) + (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$

$\lambda(x, y, z) = (\lambda x, \lambda y, \lambda z)$; $\lambda \in \mathbf{R}$.

Thinking of $\mathbf{R}^3$ as the coordinators of vectors in the usual space, then the axioms (1) – (8) are obvious as we already knew in high schools. Then $\mathbf{R}^3$ is a vector space over $\mathbf{R}$ with the null vector is O = (0,0,0).

2. Similarly, consider $\mathbf{R}^n = \{(x_1, x_2,\dots,x_n) | x_i \in \mathbf{R}\ \forall i = 1, 2\dots n\}$ with the vector addition and scalar multiplication defined as

$(x_1, x_2,\dots x_n) + (y_1, y_2,\dots y_n) = (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$

$\lambda(x_1, x_2,..., x_n) = (\lambda x_1, \lambda x_2...., \lambda x_n)$; $\lambda \in \mathbf{R}$. It is easy to check that all the axioms (1)-(8) hold true. Then $\mathbf{R}^n$ is a vector space over $\mathbf{R}$. The null vector is O = (0,0,…,0).

3. Let M_{mxn} (**R**) be the set of all mxn matrices with real entries. We consider the matrix addition and scalar multiplication defined as in Chapter 4.II. Then, the properties 2.4 in Chapter 4 show that M_{mxn} (**R**) is a vector space over **R**. The null vector is the zero matrix O.

4. Let P[x] be the set of all polynomials with real coefficients. That is, $P[x] = \{a_0 + a_1x + \dots + a_nx^n | a_0, a_1,\dots,a_n \in \mathbf{R}; n = 0, 1, 2\dots\}$. Then P[x] in a vector space over **R** with respect to the usual operations of addition of polynomials and multiplication of a polynomial by a real number.

1.3. Properties: Let V be a vector space over **K**, $\lambda, \mu \in \mathbf{K}$, $x, y \in V$. Then, the following assertion hold:

1. $\lambda x = O \Leftrightarrow \left[\begin{array}{l}\lambda = 0\\ x = O\end{array}\right.$

2. $(\lambda-\mu)x = \lambda x - \mu x$

3. $\lambda(x - y) = \lambda x - \mu y$

4. $(-\lambda)x = - \lambda x$

PROOF:

1. "$\Leftarrow$": Let $\lambda = 0$, then $0x = (0+0)x = 0x + 0x$

Using cancellation law in group (V, +) we obtain: 0x = O.

Let x = O, then $\lambda O = \lambda(O +O) = \lambda O + \lambda O$. Again, by cancellation law, we have that $\lambda O = O$.

"$\Rightarrow$": Let $\lambda x = O$ and $\lambda \neq 0$.

Then $\exists\lambda^{-1}$; multiplying λ^{-1} we obtain that $\lambda^{-1}(\lambda x) = \lambda^{-1}O = O \Rightarrow (\lambda^{-1}\lambda)x = 0 \Rightarrow 1x = O \Rightarrow x = O$.

2. $\lambda x = (\lambda-\mu+\mu)x = (\lambda-\mu)x + \mu x$

Therefore, adding both sides to $-\mu x$ we obtain that $\lambda x - \mu x = (\lambda-\mu)x$

The assertions (3), (4) can be proved by the similar ways.

II. Subspaces

2.1. Definition: Let W be a subset of a vector space V over **K**. Then, W is called a subspace of V if and only if W is itself a vector space over **K** with respect to the operations of vector addition and scalar multiplication on V.

The following theorem provides a simpler criterion for a subset W of V to be a subspace of V.

2.2. Theorem: Let W be a subset of a vector space V over **K**. Then W is a subspace of V if and only if the following conditions hold.

1. $O\in W$ (where O is the null element of V)

2. W is closed under the vector addition, that is $\forall u,v\in W \Rightarrow u + v \in W$

3. W is closed under the scalar multiplication, that is $\forall u \in W, \forall\lambda \in \mathbf{K} \Rightarrow \lambda u \in W$

Proof: "$\Rightarrow$" let $W \subset V$ be a subspace of V. Since W is a vector space over **K**, the conditions (2) and (3) are clearly true. Since (W,+) is a group, $w \neq \emptyset$. Therefore, $\exists x \in W$.

Then, $0x = O \in W$.

"$\Leftarrow$": This implication is obvious: Since V is already a vector space over **K** and $W \subset V$, to prove that W is a vector space over **K** what we need is the fact that the vector addition and scalar multiplication are also the operation on W, and the null element belong to W. They are precisely (1); (2) and (3).

2.3. Corollary: Let W be a subset of a vector space V over **K**. Then, W is a subspace of V if and only if the following conditions hold:

(i) $O\in W$

(ii) for all a, $b \in \mathbf{K}$ and all $\mathbf{u}$, $\mathbf{v} \in W$ we have that $a\mathbf{u}+b\mathbf{v}\in W$.

Examples: 1. Let O be a null element of a vector space V is then {O} is a subspace of V.

2. Let V = $\mathbf{R}^3$; M = {x, y, 0)⏐x,y∈**R**} is a subspace of $\mathbf{R}^3$, because, (0,0,0) ∈M and for $(x_1,y_1 0)$ and $(x_2,y_2,0)$ in M we have that $\lambda(x_1, y_1,0) + \mu(x_2,y_2,0) = (\lambda x_1 + \mu x_2, \lambda y_1 + \mu_2, 0) \in M\ \forall\lambda\in\mathbf{R}$.

3. Let V = $M_{nx1}(\mathbf{R})$; and A $\in M_{mxn}(\mathbf{R})$. Consider M = {X $\in M_{nx1}(\mathbf{R})$⏐AX = 0}. For X_1, $X_2 \in M$, it follows that $A(\lambda X_1 + \mu X_2) = \lambda AX_1 + \mu AX_2 = O + O = O$. Therefore, $\lambda X_1 + \mu X_2 \in M$. Hence M is a subspace of $M_{nx1}(\mathbf{R})$.

III. Linear combinations, linear spans

3.1 Definition: Let V be a vector space over **K** and let $v_1, v_2\ldots,v_n \in V$. Then, any vector in V of the form $\alpha_1 v_1 + \alpha_2 v_2 + \ldots+\alpha_n v_n$ for $\alpha_1,\alpha_2\ldots\alpha_n \in\mathbf{R}$, is called a linear combination of $v_1, v_2\ldots,v_n$. The set of all such linear combinations is denoted by

Span{$v_1, v_2\ldots v_n$} and is called the linear span of $v_1, v_2\ldots v_n$.

That is, Span {$v_1, v_2\ldots v_n$} = {$\alpha_1 v_1 + \alpha_2 v_2 + \ldots.+ \alpha_n v_n$⏐$\alpha_i \in\mathbf{K}\ \forall i = 1,2,\ldots n$}

3.2. Theorem: Let S be a subset of a vector space V. Then, the following assertions hold:

i) The Span S is a subspace of V which contains S.

ii) If W is a subspace of V containing S, then Span S ⊂ W.

3.3. Definition: Given a vector space V, the set of vectors {$u_1, u_2\ldots.u_r$} are said to span V if V = Span {$u_1, u_2.., u_n$}. Also, if the set {$u_1, u_2\ldots.u_r$} spans V, then we call it a spanning set of V.

Examples:

1. S = {(1, 0, 0); (0, 1, 0)}⊂ $\mathbf{R}^3$.

Then, span S = {x(1, 0, 0) + y(0, 1, 0)⏐x,y ∈**R**} = {(x,y,0)⏐x,y ∈**R**}.

2. S = {(1,0,0), (0,1,0), (0,0,1)} ⊂$\mathbf{R}^3$. Then,

Span S = {x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1)⏐x,y,z ∈**R**}

= {(x, y, t)⏐x, y, z ∈**R**} = $\mathbf{R}^3$

Hence, S = {(1, 0, 0) ; (0, 1, 0) ; (0, 0, 1)} is a spanning set of $\mathbf{R}^3$

IV. Linear dependence and independence

4.1. Definition: Let V be a vector space over **K**. The vectors $v_1, v_2 \ldots v_m \in V$ are said to be linearly dependent if there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$ belonging to **K,** not all zero**,** such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_m v_m = O. \qquad (4.1)$$

Otherwise, the vectors $v_1, v_2 \ldots, v_m$ are said to be linear independent.

We observe that (4.1) always holds if $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$. If (4.1) holds only in this case, that is,

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_m v_m = O \Rightarrow \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0,$$

then the vectors $v_1, v_2 \ldots v_m$ are linearly independent.

If (4.1) also holds when one of $\alpha_1, \alpha_2, \ldots, \alpha_m$ is not zero, then the vectors $v_1, v_2 \ldots v_m$ are linearly dependent.

If the vectors $v_1, v_2 \ldots v_m$ are linearly independent; we say that the set $\{v_1, v_2 \ldots v_m\}$ is linearly independent. Otherwise, the set $\{v_1, v_2 \ldots v_m\}$ is said to be linearly dependent.

Examples:

1) u = (1,-1,0); v=(1, 3, - 1); w = (5, 3, -2) are linearly dependent since 3u + 2v – w = (0, 0, 0). The first two vectors u and v are linearly independent since:

$\lambda_1 u + \lambda_2 v = (0, 0, 0) \Rightarrow (\lambda_1 + \lambda_2, -\lambda_1 + 3\lambda_2, -\lambda_2) = (0, 0, 0)$

$\Rightarrow \lambda_1 = \lambda_2 = 0.$

2) u = (6, 2, 3, 4); v = (0, 5, - 3, 1); w = (0, 0, 7, - 2) are linearly independent since

$x(6, 2, 3, 4) + y(0, 5, -3, 1) + z(0, 0, 7, -2) = (0, 0, 0)$

$\Rightarrow (6x – 2x + 5y; 3x – 3y + 7z; 4x + y – 2z) = (0, 0, 0)$

$$\Rightarrow \begin{cases} 6x = 0 \\ 2x + 5y = 0 \\ 4x + y - 2z = 0 \end{cases} \Rightarrow x = y = z = 0$$

4.2. Remarks: (1) If the set of vectors $S = \{v_1 \ldots, v_m\}$ contains O, say $v_1 = O$, then S is linearly dependent. Indeed, we have that $1.v_1 + 0.v_2 + \ldots + 0.v_m = 1.O + O + \ldots + O = O$

(2) Let $v \in V$. Then v is linear independent if and only if $v \neq O$.

(3) If S = {v_1,…, v_m} is linear independent, then every subset T of S is also linear independent. In fact, let T = {v_{k1}, v_{k2},…v_{kn}} where {k_1, k_2,…k_n} ⊂ {1, 2…m}. Take a linear combination $\alpha_{k1}v_{k1} + \ldots + \alpha_{kn}v_{kn}$=O. Then, we have that

$$\alpha_{k1}v_{k1} + \ldots + \alpha_{kn}v_{kn} + \sum_{i\in\{1,2..,m\}\backslash\{k_1,k_2..k_n\}} 0.v_i = 0$$

This yields that $\alpha_{k1} = \alpha_{k2} = \ldots = \alpha_{kn} = 0$ since S in linearly independent. Therefore, T is linearly independent.

Alternatively, if S contains a linearly dependent subset, then S is also linearly dependent.

(4) S = {v_1; … v_m} is linearly dependent if and only if there exists a vector $v_k \in S$ such that v_k is a linear combination of the rest vectors of S.

PROOF. "⇒" since S is linearly dependent, there are scalars λ_1, λ_2, … λ_m, not al zero, say $\lambda_k \neq 0$, such that.

$$\sum_{i=1}^{m}\lambda_i v_i = 0 \Rightarrow \lambda_k v_k = \sum_{k\neq i=1}^{m}(-\lambda_i)v_i$$

$$\Rightarrow v_k = \sum_{k\neq i=1}^{m}\left(-\frac{\lambda_i}{\lambda_k}\right)v_i$$

"⇐" If there is $v_k \in$ S such that $v_k = \sum_{k\neq i=1}^{m}\alpha_i v_i$, then $v_k - \sum_{k\neq i=1}^{m}\alpha_i v_i = 0$. Therefore, there exist α_1, $\alpha_k = 1$, α_{k+1}, … α_m not all zero (since $\alpha_k = 1$) such that

$$-\sum_{k\neq i=1}^{m}\alpha_i v_i + v_k = 0.$$

This means that S is linearly dependent.

(5) If S ={v_1,…v_m} linearly independent, and x is a linear combination of S, then this combination is unique in the sense that, if x = $\sum_{k=1}^{m}\alpha_i v_i = \sum_{i=1}^{m}\alpha'_i v_i$, then $\alpha_i = \alpha_i'$ ∀i = 1,2,…,m.

(6) If S = {v_1,…v_m} is linearly independent, and y ∈ V such that {v_1, v_2…v_m, y} is linearly dependent then y is a linearly combination of S.

4.3. Theorem: Suppose the set S = $\{v_1 \ldots v_m\}$ of nonzero vectors ($m \geq 2$). Then, S is linearly dependent if and only if one of its vector is a linear combination of the preceding vectors. That is, there exists a k > 1 such that $v_k = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_{k-1} v_{k-1}$

PROOF: "⇒" since $\{v_1, v_2 \ldots v_m\}$ are linearly dependent, we have that there exist scalars $a_1, a_2 \ldots, a_m$, not all zero, such that $\sum_{i=1}^{m} a_i v_i = 0$. Let k be the largest integer such that $a_k \neq 0$. Then if k > 1, $v_k = -\frac{a_1}{a_k} v_1 \ldots - \frac{a_{k-1}}{a_k} v_{k-1}$

If $k = 1 \Rightarrow v_1 = 0$ since $a_2 = a_3 = \ldots = a_m = 0$. This is a contradiction because $v_1 \neq 0$.

"⇐": This implication follows from Remark 4.2 (4).

An immediate consequence of this theorem is the following.

4.4. Corollary: The nonzero rows of an echelon matrix are linearly independent.

Example:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 2 & -4 \\ 0 & 0 & 1 & 2 & 5 & 2 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ \\ \end{matrix}$$

Then u_1, u_2, u_3, u_4 are linearly independent because we can not express any vector u_k ($k \geq 2$) as a linear combination of the preceding vectors.

V. Bases and dimension

5.1. Definition: A set S = $\{v_1, v_2 \ldots v_n\}$ in a vector space V is called a basis of V if the following two conditions hold.

(1) S is linearly independent

(2) S is a spanning set of V.

The following proposition gives a characterization of a basis of a vector space

5.2. Proposition: Let V be a vector space over **K**, and S = $\{v_1 \ldots, v_n\}$ be a subset of V. Then, the following assertions are equivalent:

i) S is a basis of V

ii) $\forall u \in V$, u can be uniquely written as a linear combination of S.

PROOF: (i) $\Rightarrow$ (ii). Since S is a Spanning set of V, we have that $\forall u \in V$, u can be written as a linear combination of S. The uniqueness of such an expression follows from the linear independence of S.

(ii) $\Rightarrow$ (i): The assertion (ii) implies that span S = V. Let now $\lambda_1,\ldots,\lambda_n \in \mathbf{K}$ such that $\lambda_1 v_1 + \ldots + \lambda_n v_n = O$. Then, since $O \in V$ can be uniquely written as a linear combination of S and $O = 0.v_1 + \ldots + 0.v_2$ we obtain that $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$. This yields that S is linearly independent.

Examples:

1. V = $\mathbf{R}^2$; S = {(1,0); (0,1)} is a basis of $\mathbf{R}^2$ because S is linearly independent and $\forall (x,y) \in \mathbf{R}^2$, (x,y) = x(1,0) + y(0,1) $\in$ Span S.

2. In the same way as above, we can see that S = {(1, 0,…,0); (0, 1, 0,…,0),…(0, 0, …,1)} is a basis of $\mathbf{R}^n$; and S is called usual basis of $\mathbf{R}^n$.

3. $P_n[x] = \{a_0 + a_1x + \ldots + a_nx^n \mid a_0 \ldots a_n \in \mathbf{R}\}$ is the space of polynomials with real coefficients and degrees $\le$ n. Then, S = $\{1, x, \ldots x^n\}$ is a basis of $P_n[x]$.

5.4. Definition: A vector space V is said to be of finite dimension if either V = {O} (trivial vector space) or V has a basis with n elements for some fixed $n \ge 1$.

The following lemma and consequence show that if V is of finite dimension, then the number of vectors in each basis is the same.

5.5. Lemma: Let S = $\{u_1, u_2 \ldots u_r\}$ and T = $\{v_1, v_2, \ldots, v_k\}$ be subsets of vector space V such that T is linearly independent and every vector in T can be written as a linear combination of S. Then $k \le r$.

PROOF: For the purpose of contradiction let $k > r \Rightarrow k \ge r+1$.

Starting from v_1 we have: $v_1 = \lambda_1 u_1 + \lambda_2 u_2 + \ldots + \lambda_r u_r$.

Since $v_1 \ne 0$, it follows that not all $\lambda_1 \ldots, \lambda_r$ are zero. Without loosing of generality we can suppose that $\lambda_1 \ne 0$. Then, $u_1 = \dfrac{1}{\lambda_1} v_1 - \dfrac{\lambda_2}{\lambda_1} u_2 - \ldots - \dfrac{\lambda_r}{\lambda_1} u_r$

For v_2 we have that

$$v_2 = \sum_{i=1}^{r}\mu_i u_i = \mu_1\left(\frac{1}{\lambda_1}v_1 - \sum_{i=2}^{r}\frac{\lambda_i}{\lambda_1}u_i\right) + \sum_{i=2}^{r}\mu_i u_i$$

Therefore, v_2 is a linear combination of $\{v_1, u_2,\dots,u_r\}$. By the similar way as above, we can derive that

v_3 is a linear combination of $\{v_1, v_2, u_3\dots u_r\}$, and so on.

Proceeding in this way, we obtain that

v_{r+1} is a linear combination of $\{v_1, v_2,\dots v_r\}$.

Thus, $\{v_1,v_2\dots v_{r+1}\}$ is linearly dependent; and therefore, T is linearly dependent. This is a contradiction.

5.6. Theorem**:** Let V be a finite–dimensional vector space; $V \neq \{0\}$. Then every basic of V has the same number of elements.

PROOF: Let $S = \{u_1\dots,u_n\}$ and $T = \{v_1\dots v_m\}$ be bases of V. Since T is linearly independent, and every vector of T is a linear combination of S, we have that $m \le n$. Interchanging the roll of T to S and vice versa we obtain $n \le m$. Therefore, m = n.

5.7. Definition: Let V be a vector space of finite dimension. Then:

1) if V = {0}, we say that V is of null–dimension and write dim V = 0

2) if $V \neq \{0\}$ and $S = \{v_1,v_2\dots.v_n\}$ is a basic of V, we say that V is of n–dimension and write dimV = n.

Examples: $\dim(\mathbf{R}^2) = 2$; $\dim(\mathbf{R}^n) = n$; $\dim(P_n[x]) = n+1$.

The following theorem is direct consequence of Lemma 5.5 and Theorem 5.6.

5.8. Theorem: Let V be a vector space of n–dimension then the following assertions hold:

1. Any subset of V containing n+1 or more vectors is linearly dependent.

2. Any linearly independent set of vectors in V with n elements is basis of V.

3. Any spanning set $T = \{v_1, v_2,\dots, v_n\}$ of V (with n elements) is a basis of V.

Also, we have the following theorem which can be proved by the same method.

5.9. Theorem: Let V be a vector space of n –dimension then, the following assertions hold.

(1) If S = {u_1, u_2,…,u_k} be a linearly independent subset of V with k<n then one can extend S by vectors u_{k+1},…,u_n such that {u_1,u_2,…,u_n} is a basis of V.

(2) If T is a spanning set of V, then the maximum linearly independent subset of T is a basis of V.

By "maximum linearly independent subset of T" we mean the linearly independent set of vectors $S \subset T$ such that if any vector is added to S from T we will obtain a linear dependent set of vectors.

VI. Rank of matrices

6.1. Definition: The maximum number of linearly independent row vectors of a matrix A = [a_{jk}] is called the rank of A and is denoted by rank A.

Example:

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 3 & -1 \\ 5 & 3 & -2 \end{pmatrix}$$; rank A = 2 because the first two rows are linearly independent; and the three rows are linearly dependent.

6.2. Theorem. The rank of a matrix equals the maximum number of linearly independent column vectors of A. Hence, A and A^T has the same rank.

PROOF: Let r be the maximum number of linearly independent row vectors of A; and let q be the maximum number of linearly independent column vectors of A. We will prove that $q \le r$. In fact, let $v_{(1)}$, $v_{(2)}$…,$v_{(r)}$ be linearly independent; and all the rest row vectors $u_{(1)}$, $u_{(2)}$…$u_{(s)}$ of A are linear combinations of $v_{(1)}$; $v_{(2)}$…,$v_{(r)}$,

$$u_{(1)} = c_{11}v_{(1)}+c_{12}v_{(2)} + \ldots+c_{1r}v_{(r)}$$

$$u_{(2)} = c_{21}v_{(1)}+c_{22}v_{(2)} + \ldots+ c_{2r}v_{(r)}$$

$$u_{(s)} = c_{s1}v_{(1)}+c_{s2}v_{(2)} + \ldots + c_{sr}v_{(r)}.$$

Writing

$$v_{(1)} = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1n} \end{pmatrix}$$
$$\vdots$$
$$v_{(r)} = \begin{pmatrix} v_{r1} & v_{r2} & \dots & v_{rn} \end{pmatrix}$$
$$u_{(1)} = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \end{pmatrix}$$
$$\vdots$$
$$u_{(s)} = \begin{pmatrix} u_{s1} & u_{s2} & \dots & u_{sn} \end{pmatrix}$$

we have that

$$u_{1k} = c_{11}v_{1k}+c_{12}v_{2k}+\dots+c_{1r}v_{rk}$$

$$u_{2k} = c_{21}v_{1k}+c_{22}v_{2k}+\dots+c_{2r}v_{rk}$$

$$\vdots$$

$$u_{sk} = c_{s1}v_{1k} + c_{s2}v_{2k} +\dots+ c_{sr}v_{rk}$$

for all k = 1, 2, …n.

Therefore,

$$\begin{pmatrix} u_{1k} \\ u_{2k} \\ \vdots \\ u_{sk} \end{pmatrix} = v_{1k}\begin{pmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{s1} \end{pmatrix} + v_{2k}\begin{pmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{s2} \end{pmatrix} + \dots + v_{rk}\begin{pmatrix} c_{12} \\ c_{2r} \\ \vdots \\ c_{sr} \end{pmatrix}$$

For all k = 1, 2…n. This yields that

$$V = \text{Span }\{\text{column vectors of } A\} \subset \text{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ c_{11} \\ \vdots \\ c_{s1} \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ c_{12} \\ \vdots \\ c_{s2} \end{pmatrix}; \dots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ c_{1r} \\ \vdots \\ c_{sr} \end{pmatrix}\right\}$$

Hence, q = dim V ≤ r.

Applying this argument for A^T we derive that r ≤ q, there fore, r = q.

6.3. Definition: The span of row vectors of A is called row space of A. The span of column vectors of A is called column space of A. We denote the row space of A by rowsp(A) and the column space of A by colsp(A).

From Theorem 6.2 we obtain the following corollary.

6.4. Corollary: The row space and the column space of a matrix A have the same dimension which is equal to rank A.

6.5. Remark: The elementary row operations do not alter the rank of a matrix. Indeed, let B be obtained from A after finitely many row operations. Then, each row of B is a linear combination of rows of A. This yields that rowsp(B) $\subset$ rowsp(A). Note that each row operation can be inverted to obtain the original state. Concretely,

+) the inverse of $R_i \leftrightarrow R_j$ is $R_j \leftrightarrow R_i$

+) the inverse of $kR_i \to R_i$ is $\frac{1}{k}R_i \to R_i$, where $k \neq 0$

+) the inverse of $kR_i + R_j \to R_j$ is $R_j - kR_i \to R_j$

Therefore, A can be obtained from B after finitely many row operations. Then,

rowsp (A) $\subset$ rowsp (B). This yields that rowsp (A) = rowsp (B).

The above arguments also show the following theorem.

6.6. Theorem: Row equivalent matrices have the same row spaces and then have the same rank.

Note that, the rank of a echelon matrix equals the number of its non–zero rows (see corollary 4.4). Therefore, in practice, to calculate the rank of a matrix, we use the row operations to deduce it to an echelon form, then count the number of non-zero rows of this echelon form, this number is the rank of A.

Example: A = $\begin{pmatrix} 1 & 2 & 0 & -1 \\ 2 & 6 & -3 & -3 \\ 3 & 10 & -6 & -5 \end{pmatrix} \xrightarrow[R_3 - 3R_1 \to R_3]{R_2 - 2R_1 \to R_2} \begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 2 & -3 & -1 \\ 0 & 4 & -6 & -2 \end{pmatrix}$

$$\xrightarrow{R_3 - 2R_2 \to R_3} \begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 2 & -3 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \Rightarrow rank(A) = 2$$

From the definition of a rank of a matrix, we immediately have the following corollary.

***6.7. Corollary*:** Consider the subset S = {v_1, v_2,…, v_k} ⊂ $\mathbf{R}^n$ the the following assertions hold.

1. S is linearly independent if and only if the kxn matrix with row vectors v_1, v_2,…, v_k has rank *k*.

2. Dim spanS = rank(A), where A is the kxn matrix whose row vectors are v_1, v_2,…, v_k, respectively.

6.8. Definition: Let S be a finite subset of vector space V. Then, the number dim(Span S) is called the rank of S, denoted by rankS.

Example: Consider

S = {(1, 2, 0, - 1); (2, 6, -3, - 3); (3, 10, -6, - 5)}. Put

$$A = \begin{pmatrix} 1 & 2 & 0 & -1 \\ 2 & 6 & -3 & -3 \\ 3 & 10 & -6 & -5 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 2 & -3 & -1 \\ 0 & 4 & -6 & -2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 2 & -3 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Then, rank S = dim span S = rank A =2. Also, a basis of span S can be chosen as

{(0, 2, - 3, -1); (0, 2, -3, - 1)}.

VII. Fundamental theorem of systems of linear equations

7.1. Theorem: Consider the system of linear equations

$$AX = B \tag{7.1}$$

where A = (a_{jk}) is an mxn coefficient matrix; $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is the column vector of unknowns;

and $B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$. Let $\tilde{A}$ = [A⋮B] be the augmented matrix of the system (7.1). Then the system (7.1) has a solution if and only if rank A = rank $\tilde{A}$. Moreover, if rankA = rank $\tilde{A}$ = n, the system has a unique solution; and if rank A = rank $\tilde{A}$ <n, the system has infinitely many

solutions with n – r free variables to which arbitrary values can be assigned. Note that, all the solutions can be obtained by Gauss eliminations.

PROOF: We prove the first assertion of the theorem. The second assertion follows straightforwardly.

"$\Rightarrow$" Denote by C_1; $C_2 \ldots C_n$ the column vectors of A. Since the system (7.1) has a solution, say $x_1, x_2 \ldots x_n$ we obtain that $x_1C_1 + x_2C_2 + \ldots + x_nC_n = B$.

Therefore, $B \in \text{colsp}(A) \Rightarrow \text{colsp}(\tilde{A}) \subset \text{colsp}(A)$. It is obvious that

$\text{colsp}(A) \subset \text{colsp}(\tilde{A})$. Thus, $\text{colsp}(\tilde{A}) = \text{colsp}(A)$. Hence, $\text{rank}A = \text{rank}\tilde{A} = \dim(\text{colsp}(A))$.

"$\Leftarrow$" if $\text{rank}A = \text{rank}\tilde{A}$, then $\text{colsp}(\tilde{A}) = \text{colsp}(A)$. This follows that $B \in \text{colsp}A$. This yields that there exist $x_1, x_2 \ldots x_n$ such that $B = x_1C_1 + x_2C_2 + \ldots + x_nC_n$. This means that system (7.1) has a solution $x_1, x_2 \ldots, x_n$.

In the case of homogeneous linear system, we have the following theorem which is a direct consequence of Theorem 7.1.

7.2. Theorem: **(Solutions of homogeneous systems of linear equations)**

A homogeneous system of linear equations

$$AX = O \qquad (7.2)$$

always has a trivial solution $x_1 = 0; x_2 = 0 \ldots, x_n = 0$. Nontrivial solutions exist if and only if $r = \text{rank}A < n$. The solution vectors of (7.2) form a vector space of dimension n – r; it is a subspace of $M_{nx1}(\mathbf{R})$.

7.3. Definition: Let A be a real matrix of the size mxn ($A \in M_{mxn}(\mathbf{R})$). Then, the vectors space of all solution vectors of the homogeneous system (7.2) is called the null space of the matrix A, denoted by nullA. That is, $\text{null}A = \{X \in M_{nx1}(\mathbf{R}) \mid AX = O\}$, then the dimension of nullA is called nullity of A.

Note that it is easy to check that nullA is a subspace of $M_{nx1}(\mathbf{R})$ because,

+) $O \in \text{null}A$ (since $AO = O$)

+) $\forall\lambda, \mu \in \mathbf{R}$ and $X_1, X_2 \in \text{null } A$, we have that $A(\lambda X_1 + \mu X_2) = \lambda AX_1 + \mu AX_2 = \lambda O + {}$ $+\mu O = 0$. there fore, $\lambda X_1 + \mu X_2 \in \text{null}A$.

Note also that nullity of A equals n – rank A.

We now observe that, if X_1, X_2 are solutions of the nonhomogeneous system AX = B, then $A(X_1 - X_2) = AX_1 - AX_2 = B - B = 0$.

Therefore, $X_1 - X_2$ is a solution of the homogeneous system

$$AX = O.$$

This observation leads to the following theorem.

7.4. Theorem: Let X_0 be a fixed solution of the nonhomogeneous system (7.1). Then, all the solutions of the nonhomogeneous system (7.1) are of the form

$$X = X_0 + X_h$$

where X_h is a solution of the corresponding homogeneous system (7.2).

VIII. Inverse of a matrix

8.1. Definition: Let A be a n-square matrix. Then, A is said to be invertible (or nonsingular) if the exists an nxn matrix B such that $BA = AB = I_n$. Then, B is called the inverse of A; denoted by $B = A^{-1}$.

Remark: If A is invertible, then the inverse of A is unique. Indeed, let B and C be inverses of A. Then, B = BI = BAC = IC = C.

8.2. Theorem (Existence of the inverse):

The inverse A^{-1} of an nxn matrix exists if and only if rankA = n.

Therefore, A is nonsingular $\Leftrightarrow$ rank A = n.

PROOF: Let A be nonsingular. We consider the system

$$AX = H \tag{8.1}$$

It is equivalent to $A^{-1}AX = A^{-1}H \Leftrightarrow IX = A^{-1}H \Leftrightarrow X = A^{-1}H$, therefore, (8.1) has a unique solution $X = A^{-1}H$ for any $H \in M_{nx1}$ (**R**). This yields that Rank A = n.

Conversely, let rankA = n. Then, for any $\mathbf{b} \in M_{nx1}$ (**R**), the system $A\mathbf{x} = \mathbf{b}$ always has a unique solution x; and the Gauss elimination shows that each component x_j of $\mathbf{x}$ is a linear combination of those of $\mathbf{b}$. So that we can write $\mathbf{x} = B\mathbf{b}$ where B depends only on A. Then, $A\mathbf{x} = AB\mathbf{b} = \mathbf{b}$ for any $\mathbf{b} \in M_{nx1}$ (**R**).

Taking now $\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix};\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix};...;\begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix}$ as **b**, from AB**b** = **b** we obtain that AB = I.

Similarly, **x** = B**b** = BA**x** for any x $\in M_{nx1}(\mathbf{R})$. This also yield BA = I. Therefore, AB = BA = I. This means that A is invertible and $A^{-1} = B$.

Remark: Let A, B be n-square matrices. Then, AB = I if and only if BA = I. Therefore, we have to check only one of the two equations AB = I or BA = I to conclude B = A^{-1}, also $A=B^{-1}$.

PROOF: Let AB = I. Obviously, Null $(B^TA^T) \supset$ Null (A^T)

Therefore, nullity of $B^TA^T \geq$ nullity of A^T. Hence,

$n - \text{rank}(B^TA^T) \geq n - \text{rank}(A^T) \Rightarrow \text{rank } A = \text{rank } A^T \geq \text{rank}(B^TA^T) = \text{rank } I = n.$

Thus, rank A = n; and A is invertible.

Now, we have that $B = IB = A^{-1}AB = A^{-1}I = A^{-1}$.

8.3. Determination of the inverse (Gauss – Jordan method):

1) **Gauss – Jordan method** (a variation of Gauss elimination): Consider an nxn matrix with rank A = n. Using A we form n systems:

$AX_{(1)} = \mathbf{e}_{(1)}$; $AX_{(2)} = \mathbf{e}_{(2)} \ldots AX_{(n)} = \mathbf{e}_{(n)}$ where $\mathbf{e}_{(j)}$ has the j[th] component 1 and other components 0. Introducing the nxn matrix $X = [X_{(1)}, X_{(2)}, \ldots X_{(n)}]$ and $I = [\mathbf{e}_{(1)}, \mathbf{e}_{(2)} \ldots \mathbf{e}_{(n)}]$ (unit matrix). We combine the n systems into a single system AX = I, and the n augmented matrix $\tilde{A} = [A \vdots I]$. Now, AX= I implies $X = A^{-1}I = A^{-1}$ and to solve AX = I for X we can apply the Gauss elimination to $\tilde{A} = [A \vdots I]$ to get $[U \vdots H]$, where U is upper triangle since Gauss elimination triangularized systems. The Gauss – Jordan elimination now operates on $[U \vdots H]$, and, by eliminating the entries in U above the main diagonal, reduces it to $[I \vdots K]$ the augmented matrix of the system $IX = A^{-1}$. Hence, we must have $K = A^{-1}$ and can then read off A^{-1} at the end.

2) **The algorithm of Gauss – Jordan method to determine** A^{-1}

Step1: Construct the augmented matrix $[A \vdots I]$

Step2: Using row elementary operations on [A$\vdots$I] to reduce A to the unit matrix. Then, the obtained square matrix on the right will be A^{-1}.

Step3: Read off the inverse A^{-1}.

Examples:

1. $A = \begin{bmatrix} 5 & 3 \\ 2 & 1 \end{bmatrix}$

$$[A\vdots I] = \left[\begin{array}{cccc} 5 & 3 & \vdots & 1 & 0 \\ 2 & 1 & \vdots & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccccc} 1 & \frac{3}{5} & \vdots & \frac{1}{5} & 0 \\ 2 & 1 & \vdots & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccccc} 1 & \frac{3}{5} & \vdots & \frac{1}{5} & 0 \\ 0 & -\frac{1}{5} & \vdots & -\frac{2}{5} & 1 \end{array}\right]$$

$$\to \left[\begin{array}{ccccc} 1 & \frac{3}{5} & \vdots & \frac{1}{5} & 0 \\ 0 & 1 & \vdots & 2 & -5 \end{array}\right] \to \left[\begin{array}{ccccc} 1 & 0 & \vdots & -1 & 3 \\ 0 & 1 & \vdots & 2 & -5 \end{array}\right] \Rightarrow A^{-1} = \begin{bmatrix} -1 & 3 \\ 2 & -5 \end{bmatrix}$$

2. $A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{bmatrix}; [A\vdots I] = \left[\begin{array}{ccccccc} 1 & 0 & 2 & \vdots & 1 & 0 & 0 \\ 2 & -1 & 3 & \vdots & 0 & 1 & 0 \\ 4 & 1 & 8 & \vdots & 0 & 0 & 1 \end{array}\right]$

$$\xrightarrow[R_3 - 4R_1 \to R_3]{R_2 - 2R_1 \to R_2} \begin{bmatrix} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & -1 & -1 & -2 & 1 & 0 \\ 0 & 1 & 0 & -4 & 0 & 1 \end{bmatrix} \xrightarrow{R_3 + R_2 \to R_3}$$

$$\left[\begin{array}{ccccccc} 1 & 0 & 2 & \vdots & 1 & 0 & 0 \\ 0 & -1 & -1 & \vdots & -2 & 1 & 0 \\ 0 & 0 & -1 & \vdots & -6 & 1 & 1 \end{array}\right] \xrightarrow[(-1)R_3 \to R_3]{(-1)R_2 \to R_2} \left[\begin{array}{ccccccc} 1 & 0 & 2 & \vdots & 1 & 0 & 0 \\ 0 & 1 & 1 & \vdots & 2 & -1 & 0 \\ 0 & 0 & 1 & \vdots & 6 & -1 & -1 \end{array}\right]$$

$$\xrightarrow[R_1 - 2R_3 \to R_1]{R_2 - R_3 \to R_2} \left[\begin{array}{ccccccc} 1 & 0 & 0 & \vdots & -11 & 2 & 02 \\ 0 & 1 & 0 & \vdots & -4 & 0 & 1 \\ 0 & 0 & 1 & \vdots & 6 & -1 & -1 \end{array}\right]$$

8.4. Remarks:

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

2. If A, B are nonsingular nxn matrices, then AB is a nonsingular and

$$(AB)^{-1} = B^{-1}A^{-1}.$$

3. Denote by GL($\mathbf{R}^n$) the set of all invertible matrices of the size nxn. Then,

(GL($\mathbf{R}^n$), •) is a group. Also, we have the cancellation law: AB = AC; A ∈GL($\mathbf{R}^n$) ⇒B = C.

IX. Determinants

9.1. Definitions: Let A be an nxn square matrix. Then, a determinant of order n is a number associated with A = $[a_{jk}]$, which is written as D = det A = $\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = |A|$

and defined inductively by

for n = 1; A = $[a_{11}]$; D = det A = a_{11}

for n = 2; A = $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$; D = det A = $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ = $a_{11}a_{22} - a_{12}a_{21}$

for n ≥ 2; A = $[a_{jk}]$; D = $(-1)^{1+1}a_{11}M_{11}+(-1)^{1+2}a_{12}M_{12}+...+(-1)^{1+n}a_{1n}M_{1n}$

where M_{1k}; $1 \le k \le n$, is a determinant of order n – 1, namely, the determinant of the submatrix of A, obtained from A by deleting the first row and the k^{th} column.

Example: $\begin{vmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 0 & 1 & 2 \end{vmatrix} = 1\begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} - 1\begin{vmatrix} 1 & 3 \\ 0 & 2 \end{vmatrix} + 2\begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix}$ = -1-2+2 = -1

To compute the determinant of higher order, we need some more subtle properties of determinants.

9.2. Properties:

1. If any two rows of a determinant are interchanged, then the value of the determinant is multiplied by (-1).

2. A determinant having two identical rows is equal to 0.

3. We can compute a determinant by expanding any row, say row j, and the formula is

$$\text{Det A} = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk}$$

where M_{jk} ($a \le k \le n$) is a determinant of order n – 1, namely, the determinant of the submatrix of A, obtained from A by deleting the j^{th} row and the k^{th} column.

Example: $\begin{vmatrix} 1 & 2 & 1 \\ 3 & 0 & 0 \\ 2 & 1 & 4 \end{vmatrix} = -3\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} + 0\begin{vmatrix} 1 & 1 \\ 2 & 4 \end{vmatrix} + 0\begin{vmatrix} 1 & 1 \\ 2 & 4 \end{vmatrix} - 0\begin{vmatrix} 1 & 2 \\ 2 & 1 \end{vmatrix} = -21$

4. If all the entries in a row of a determinant are multiplied by the same factor α, then the value of the new determinant is α times the value of the given determinant.

Example: $\begin{vmatrix} 5 & 10 & 35 \\ 1 & 2 & 21 \\ 1 & 3 & 5 \end{vmatrix} = 5\begin{vmatrix} 1 & 2 & 7 \\ 1 & 2 & 1 \\ 1 & 33 & 5 \end{vmatrix} = 5.(-6) = -30$

5. If corresponding entries in two rows of a determinant are proportional, the value of determinant is zero.

Example: $\begin{vmatrix} 5 & 10 & 35 \\ 2 & 4 & 14 \\ 1 & 3 & 9 \end{vmatrix} = 0$ sine the first two rows are proportional.

6. If each entry in a row is expressed as a binomial, then the determinant can be written as the sum of the two corresponding determinants. Concretely,

$$\begin{vmatrix} a_{11}+a'_{11} & a_{12}+a'_{12} & \cdots & a_{1n}+a'_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} a'_{11} & a'_{12} & \cdots & a'_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

7. The value of a determinant is left unchanged if the entries in a row are altered by adding them any constant multiple of the corresponding entries in any other row.

8. The value of a determinant is not altered if its rows are written as columns in the same order. In other words, det A = det A^T.

Note: From (8) we have that, the properties (1) – (7) still hold if we replace rows by columns, respectively.

9. The determinant of a triangular matrix is the multiplication of all entries in the main diagonal.

10. Det (AB) = det(A).det(B); and thus det(A^{-1}) = $\dfrac{1}{\det A}$

Remark: The properties (1); (4); (7) and (9) allow to use the elementary row operations to deduce the determinant to the triangular form, then multiply the entries on the main diagonal to obtain the value of the determinant.

Example:

$$\begin{vmatrix} 1 & 2 & 7 \\ 5 & 11 & 37 \\ 3 & 5 & 3 \end{vmatrix} \xrightarrow[R_3 - 3R_1 \to R_3]{R_2 - 5R_1 \to R_2} \begin{vmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & -1 & -18 \end{vmatrix} \xrightarrow{R_3 + R_2 \to R_3} \begin{vmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 0 & -16 \end{vmatrix} = -16$$

X. Determinant and inverse of a matrix, Cramer's rule

10.1. Definition: Let A = $[a_{jk}]$ be an nxn matrix, and let M_{jk} be a determinant of order n – 1, obtained from D = det A by deleting the j^{th} row and k^{th} column of A. Then, M_{jk} is called the minor of a_{jk} in D. Also, the number $A_{jk}=(-1)^{j+k} M_{jk}$ is called the cofactor of a_{jk} in D. Then, the matrix C = $[A_{ij}]$ is called the cofactor matrix of A.

10.2 Theorem: The inverse matrix of the nonsingular square matrix A = $[a_{jk}]$ is given by

$$A^{-1} = \frac{1}{\det A} C^T = \frac{1}{\det A}[A_{jk}]^T = \frac{1}{\det A}\begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}$$

where A_{jk} occupies the same place as a_{kj} (not a_{jk}) does in A.

PROOF: Denote by AB = $[g_{kl}]$ where B = $\frac{1}{\det A}[A_{jk}]^T$.

Then, we have that $g_{kl} = \sum_{s=1}^{n} a_{ks}\frac{A_{ls}}{\det A}$.

Clearly, if k = l, $g_{kk} = \sum_{s=1}^{n} (-1)^{k+s} a_{ks}\frac{M_{ks}}{\det A} = \frac{\det A}{\det A} = 1$.

If k ≠l; $g_{kl} = \sum_{J=1}^{n} (-1)^{l+s} a_{ks}\frac{M_{ls}}{\det A} = \frac{\det A_k}{\det A}$

where A_k is obtained from A by replacing l^{th} row by k^{th} row. Since A_k has two identical row, we have that det A_k = 0. Therefore, g_{kl} = 0 if k ≠l. Thus, AB = I. Hence, B = A^{-1}.

10.3. Definition: Let A be an mxn matrix. We call a submatrix of A a matrix obtained from A by omitting some rows or some columns or both.

10.4. Theorem: An mxn matrix A = $[a_{jk}]$ has rank r ≥ 1 if and only if A has an rxr submatrix with nonzero determinant, whereas the determinant of every square submatrix of r + 1 or more rows is zero.

In particular, for an nxn matrix A, we have that rank A = n ⇔ det A ≠0 ⇔ $\exists A^{-1}$(that mean that A is nonsingular).

10.5. Cramer's theorem (solution by determinant):

Consider the linear system AX = B where the coefficient matrix A is of the size nxn.

If D = det A ≠ 0, then the system has precisely one solution X= $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ given by the formula

$$x_1 = \frac{D_1}{D};\ x_2 = \frac{D_2}{D};...;\ x_n = \frac{D_n}{D} \qquad \text{(Cramer's rule)}$$

where D_k is the determinant obtained from D by replacing in D the k^{th} column by the column vector B.

PROOF: Since rank A = rank $\tilde{A}$ = n, we have that the system has a unique solution $(x_1,..., x_n)$. Let $C_1, C_2, ... C_n$ be columns of A, so A = $[C_1\ C_2...C_n]$;

Putting $A_k = [\ C_1...C_{k-1}\ \ B\ C_{k+1}\ ...\ C_n]$ we have that D_k=det A_k

Since $(x_1,x_2,..., x_n)$ is the solution of AX = B, we have that

$$B = \sum_{i=1}^{n} x_i C_i \Rightarrow \det A_k = \sum_{i=1}^{n} x_i \det\left[C_1...C_{k-1}C_iC_{k+1}\right]$$

= x_k det A (note that, if i ≠ k, det ($[C_1...C_{k-1}\ \ C_i\ C_{k+1}\ ...\ C_n]$ = 0)

Therefore, $x_k = \dfrac{\det A_k}{\det A} = \dfrac{D_k}{D}$ for all k=1,2,…,n.

Note: Cramer's rule is not practical in computations but it is of theoretical interest in differential equations and other theories that have engineering applications.

XI. Coordinates in vector spaces

11.1. Definition: Let V be an n–dimensional vector space over **K**, and S = {e_1,..., e_n} be its basis. Then, as in seen in sections IV, V, it is known that any vector v $\in$V can be uniquely expressed as a linear combination of the basis vectors in S, that is to say: there exists a unique (a_1,a_2,...a_n) $\in\mathbf{K}^n$ such that v = $\sum_{i=1}^{n} a_i e_i$.

Then, n scalars a_1, a_2,..., a_n are called coordinates of v with respect to the basis S. Denote by $[v]_S = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ and call it the coordinate vector of v. We also denote by

$(v)_S=(a_1\ a_2\ ...\ a_n)$ and call it the row vector of coordinates of v.

Examples:

1) v = (6,4,3) $\in\mathbf{R}^3$; S = {(1,0,0); (0,1,0); (0,0,1)}- the usual basis of $\mathbf{R}^3$. Then,

$$[v]_S = \begin{pmatrix} 6 \\ 4 \\ 3 \end{pmatrix} \Leftrightarrow v = 6(1,0,0) + 4(0,1,0) + 3(0,0,1)$$

If we take other basis ε = {(1,0,0); (1,1,0); (1,1,1)} then

$[v]_\varepsilon = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$ since v = 2(1,0,0) + 3(1,1,0) + 1 (1,1,1).

2. Let V = $P_2[t]$; S = {1, t-1; $(t-1)^2$} and v = $2t^2 - 5t + 6$. To find $[v]_S$ we let

$[v]_S = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$. Then, v = $2t^2-5t+6= \alpha .1+ \beta(t-1)+ \gamma(t-1)^2$ for all t.

Therefore, we easily obtain α = 3; β = - 1; γ = 2. This yields $[v]_S = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$

11.2. Change of bases:

Definition: Let V be vector space and U = { u_1, u_2..., u_n}; and S = {v_1,v_2,...v_3} be two bases of V. Then, since U is a basis of V, we can express:

$v_1 = a_{11}u_1 + a_{21}u_2+...+a_{n1}u_n$

$v_2 = a_{12}u_1 + a_{22}u_2+...+a_{n2}u_n$

$\vdots$

$v_n = a_{1n}u_1 + a_{2n}u_2+...+a_{nn}u_n$

Then, the matrix P = $\begin{pmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ \vdots & & \\ a_{n1} & a_{n2} \dots & a_{nn} \end{pmatrix}$ is called the change-of-basis matrix from U to S.

Example: Consider $\mathbf{R}^3$ and U = {(1,0,0), (0,1,0);(0,0,1)}- the usual basis;

S = {(1,0,0), (1,1,0);(1,1,1)}- the other basis of $\mathbf{R}^3$. Then, we have the expression

(1,0,0) = 1 (1,0,0) + 0 (0,1,0) + 0 (0,0,1)

(1,1,0) = 1 (1,0,0) + 1 (0,1,0) + 0 (0,0,1)

(1,1,1) = 1 (1,0,0) + 1 (0,1,0) + 1 (0,0,1)

This mean that the change-of-basis matrix from U to S is

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Theorem: Let V be an n–dimensional vector space over **K**; and U, S be bases of V, and P be the change-of-basis matrix from U to S. then,

$$[v]_U = P[v]_S \text{ for all } v \in V.$$

PROOF. Set U = {u_1, u_2,... u_n}, S = {v_1, v_2..., v_n}, P = [a_{jk}]. By the definition of P, we obtain that

$$v = \sum_{i=1}^{n} \lambda_i v_i = \sum_{i=1}^{n} \lambda_i \sum_{k=1}^{n} a_{ki} u_k = \sum_{i=1}^{n} \sum_{k=1}^{n} \lambda_i a_{ki} u_k = \sum_{k=1}^{n} \left(\sum_{i=1}^{n} \lambda_i a_{ki} \right) u_k$$

Let $[v]_U = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}$. Then v = $\sum_{k=1}^{n} \gamma_k u_k$ Therefore, $\gamma_k = \sum_{i=1}^{n} a_{ki}\lambda_i$ $\forall$ k = 1,2,...n

This yields that $[v]_U = P[v]_S$.

Remark: Let U, S be bases of vector space V of n–dimension, and P be the change-of-basis matrix from U to S. Then, since S is linearly independent, it follows that rank P = rank S = n. Hence, P is nonsingular. It is straightforward to see that P^{-1} is the change-of-basis matrix from S to U.

Chapter 6: Linear Mappings and Transformations

I. Basic definitions

1.1. Definition: Let V and W be two vector spaces over the same field **K** (say **R** or **C**). A mapping F: V $\to$ W is called a linear mapping (or vector space homomorphism) if it satisfies the following conditions:

$$\left| \begin{array}{l} \text{1) For any } u, v \in V, F(u+v) = F(u)+ F(v) \\ \text{2) For all } k \in \mathbf{K}; \text{ and } v \in V, F(kv) = k F(v) \end{array} \right.$$

In other words; F: V $\to$ W is linear if it preserves the two basic operations of vector spaces, that of vector addition and that of scalar multiplication.

Remark: 1) The two conditions (1) and (2) in above definition are equivalent to:

"$\forall \lambda, \mu \in \mathbf{K}, \forall u,v \in V \Rightarrow F(\lambda u + \mu v) = \lambda F(u) + \mu F(v)$"

2) Taking k = 0 in condition (2) we have that $F(O_v) = O_w$ for a linear mapping

F: V $\to$ W.

1.2. Examples:

1) Given $A \in M_{mxn}$ (**R**); Then, the mapping

F: M_{nx1} (**R**) $\to$ M_{mx1} (**R**); F(X) = AX $\forall$ X $\in$ M_{nx1} (**R**) is a linear mapping because,

a) $F(X_1 + X_2) = A(X_1 + X_2) = AX_1 + AX_2$

$= F(X_1) + F(X_2) \quad \forall X_1, X_2 \in M_{nx1}$ (**R**)

b) $F(\lambda X) = A(\lambda X) = \lambda AX$ $\forall \lambda \in$ **R** and X $\in$ M_{nx1} (**R**).

2) The projection F: $\mathbf{R}^3 \to \mathbf{R}^3$, F(x,y,z) = (x,y,0) is a linear mapping.

3) The translation f: $\mathbf{R}^2 \to \mathbf{R}^2$, f(x,y) = (x+1, y +2) is not a linear mapping since f(0,0) = (1,2) $\neq$ (0,0) $\in \mathbf{R}^2$.

4) The zero mapping F: V $\to$W, F(v) = O_w $\forall$ v $\in$ V is a linear mapping.

5) The identity mapping F: V $\to$ V; F(v)= v $\forall$ v $\in$ V is a linear mapping.

6) The mapping F: $\mathbf{R}^3 \to \mathbf{R}^2$; F(x,y,z) = (2x + y +z, x - 4y – 2z) is a linear mapping because $F(\lambda(x,y,z)+ \mu(x_1,y_1,z_1)) = F(\lambda x+\mu x_1, \lambda y + \mu y_1, \lambda z + \mu z_1) =$

$=(2(\lambda x + \mu x_1)+\lambda y+\mu y_1+\lambda z+\mu z_1,\ 4(\lambda x+\mu x_1) - 2(\lambda y+\mu y_1) - (\lambda z + \mu z_1))$

$= \lambda(2x + y + z,\ 4x-2y-z) + \mu(2x_1+y_1+ z_1,\ 4x_1 - 2y_1 - z_1)$

$= \lambda F(x,y,z) + \mu F(x_1,y_1,z_1) \forall \lambda,\mu \in R$ and $(x,y,z) \in \mathbf{R}^3$, $(x_1,y_1,z_1) \in \mathbf{R}^3$

The following theorem say that, for finite–dimensional vector spaces, a linear mapping is completely determined by the images of the elements of a basis.

1.3. Theorem: Let V and W be vector spaces of finite – dimension, let $\{v_1, v_2,\dots, v_n\}$ be a basis of V and let $u_1,u_2,..., u_n$ be any n vectors in W. Then, there exists a unique linear mapping $F: V \to W$ such that $F(v_1) = u_1$; $F(v_2) = u_2,..., F(v_n) = u_n$.

PROOF. We put $F: V \to W$ by

$F(v) = \sum_{i=1}^{n} \lambda_i u_i$ for $v=\sum_{i=1}^{n} \lambda_i v_i \in V$. Since for each $v \in V$ there exists a unique

$(\lambda_1,\dots \lambda_n) \in \mathbf{K}^n$ such that $v = \sum_{i=1}^{n} \lambda_i v_i$. This follows that F is a mapping satisfying $F(v_1) = u_1$; $F(v_2) = u_2,..., F(v_n) = u_n$. The linearity of F follows easily from the definition. We now prove the uniqueness. Let G be another linear mapping such that $G(v_i) = u_i\ \forall i= 1, 2,...,n$. Then, for each $v = \sum_{i=1}^{n} \lambda_i v_i \in V$ we have that.

$$G(v) = G\left(\sum_{i=1}^{n} \lambda_i v_i\right) = \sum_{i=1}^{n} \lambda_i G(v_i) = \sum_{i=1}^{n} \lambda_i u_i = \sum_{i=1}^{n} \lambda_i F(v_i) =$$

$$F\left(\sum_{i=1}^{n} \lambda_i v_i\right) = F(v).$$

This yields that G = F.

1.4. Definition: Two vector spaces V and W over **K** are said to be isomorphic if there exists a bijective linear mapping $F: V \to W$. The mapping F is then called an isomorphism between V and W; and we denote by $V \cong W$.

Example: Let V be a vector space of n–dimension over **K**, and S be a basis of V.

Then, the mapping F: V $\to$ $\mathbf{K}^n$

v $\mapsto$ ($v_1, v_2, ...v_n$), where $v_1, v_2, ...v_n$ is coordinators of v with respect to S, is an isomorphism between V and $\mathbf{K}^n$. Therefore, V $\cong$ $\mathbf{K}^n$.

II. Kernels and Images

2.1. Definition: Let F: V $\to$ W be a linear mapping. The image of F is the set

F(V) = {F (v) |v $\in$V }. It can also be expressed as

F(V) = {u$\in$W | $\exists$ v $\in$ V such that F(v) = u}; and we denote by Im F = F(V).

2.2. Definition: The kernel of a linear mapping F: V $\to$ W, written by Ker F, is the set of all elements of V which map into O$\in$ W (null vector of W). That is,

$$\text{Ker } F = \{v\in V \mid F(v) = O\} = F^{-1}(\{O\})$$

2.3. Theorem: Let F: V $\to$ W be a linear mapping. Then, the image ImF of W is a subspace of W; and the kernel Ker F of F is a subspace of V.

PROOF: Since F is a linear mapping, we obtain that F(O) = O.

Therefore, O $\in$ Im F. Moreover, for $\lambda, \mu \in \mathbf{K}$ and u, w $\in$Im F we have that there exist v, v' $\in$V such that F(v) = u; F(v') = w. Therefore,

$$\lambda u + \mu w = \lambda f(v) + \mu F(v') = F(\lambda v + \mu v').$$

This yields that $\lambda v + \mu w \in$ Im F. Hence, ImF is a subspace of W.

Similarly, ker F is a subspace of V.

Example: 1) Consider the projection F: $\mathbf{R}^3 \to \mathbf{R}^3$, F(x,y,z) = (x,y,0) $\forall$ (x,y,z) $\in \mathbf{R}^3$. Then, Im F = {F(x,y,z)| (x,y,z) $\in \mathbf{R}^3$} = {(x,y,0)| x,y $\in \mathbf{R}$}. This is the plane xOy.

Ker F = {(x,y,z) $\in \mathbf{R}^3$| F(x,y,z) = (0,0,0) $\in \mathbf{R}^3$} = {(0,0,z)| z $\in \mathbf{R}$}. This is the Oz – axis

2) Let A $\in M_{mxn}(\mathbf{R})$, consider F: $M_{nx1}(\mathbf{R}) \to M_{mx1}(\mathbf{R})$ defined by

$$F(X) = AX \ \forall \ X \in M_{nx1}(\mathbf{R}).$$

ImF = {AX | X $\in M_{nx1}(\mathbf{R})$}. To compute ImF we let $C_1, C_2, ... C_n$ be columns of A, that is A = [$C_1\, C_2 ... C_n$]. Then,

Im F = {$x_1 C_1 + x_2 C_2 + ... + x_n C_n \mid x_1 ... x_n \in \mathbf{R}$}

= Span{$C_1, C_2, ... C_n$} = Column space of A = Colsp (A)

Ker F = $\{X \in M_{nx1}(\mathbf{R}) \mid AX = 0\}$ = null space of A = Null A.

Note that Null A is the solution space of homogeneous linear system AX = 0

2.4. Theorem: Suppose that $v_1, v_2, \ldots v_m$ span a vector space V and the mapping $F: V \to W$ is linear. Then, $F(v_1), F(v_2), \ldots F(v_m)$ span Im F.

PROOF. We have that span $\{v_1, \ldots v_m\}$ = V. Let now $u \in$ Im F.

Then, $\exists$ $v \in V$ such that u = F(v). Since span $\{v_1, \ldots, v_m\}$ = V, there exist $\lambda_1 \ldots \lambda_m$ such that $v = \sum_{i=1}^{m} \lambda_i v_i$. This yields,

$$U = F(v) = F(\sum_{i=i}^{m} \lambda_i v_i) = \sum_{i=1}^{m} \lambda_i F(v_i).$$

Therefore, $u \in$ Span $\{F(v_1), F(v_2) \ldots. F(v_m)\}$

$\Rightarrow$ Im F $\subset$ Span $\{F(v_1), \ldots. F(v_m)\}$.

Clearly Span $\{F(v_1), \ldots. F(v_m)\} \subset$ Im F. Hence, Im F = Span $\{F(v_1) \ldots. F(v_m)\}$. q.e.d.

2.5. Theorem: Let V be a vector space of finite – dimension, and $F: V \to W$ be a linear mapping. Then,

$$\dim V = \dim \operatorname{Ker} F + \dim \operatorname{Im} F. \qquad \text{(Dimension Formula)}$$

PROOF. Let $\{e_1, e_2, \ldots, e_k\}$ be a basis of KerF. Extend $\{e_1, \ldots, e_k\}$ by $e_{k+1}, \ldots, e_n$ such that $\{e_1, \ldots, e_n\}$ is a basis of V. We now prove $\{F(e_{k+1}) \ldots F(e_n)\}$ is a basis of Im F. Indeed, by Theorem 2.4, $\{F(e_1), \ldots, F(e_n)\}$ spans Im F but $F(e_1) = F(e_2) = \ldots = F(e_k) = 0$, it follows that $\{F(e_{k+1}), \ldots, F(e_n)\}$ spans ImF. We now prove $\{F(e_{k+1}) \ldots F(e_n)\}$ is linearly independent. In fact, if $\lambda_{k+1} F(e_{k+1}) + \ldots + \lambda_n F(e_n) = 0 \Rightarrow F(\lambda_{k+1} e_{k+1} + \ldots + \lambda_n e_n) = 0$

$\Rightarrow \lambda_{k+1} e_{k+1} + \ldots + \lambda_n e_n \in$ ker F = span $\{e_1, e_2, \ldots, e_k\}$

Therefore, there exist $\lambda_1, \lambda_2, \ldots \lambda_k$ such that $\sum_{i=1}^{k} \lambda_i e_i = \sum_{J=k+1}^{n} \lambda_J e_J$

$$\Rightarrow \sum_{i=1}^{k} (-\lambda_i) e_i + \sum_{J=k+1}^{n} \lambda_J e_J = 0.$$

Since $e_1, e_2, \ldots, e_n$ are linearly independent, we obtain that

$\lambda_1 = \lambda_2 = ... = \lambda_k = \lambda_{k+1} = ... = \lambda_n = 0$. Hence, F ($e_{k+1}$),..., F($e_n$) are also linearly independent. Thus, {F(e_{k+1})... F(e_n)} is a basis of Im F. Therefore, dim Im F = n – k = n – dim ker F. This follow that dim ImF + dim Ker F = n = dim V.

2.6. Definition: Let F: V $\to$ W be a linear mapping. Then,

1) the rank of F, denoted by rank F, is defined to be the dimension of its image.

2) the nullity of F, denoted by null F, is defined to be the dimension of its kernel.

That is, rank F = dim (ImF), and null F = dim (Ker F).

Example: F: $\mathbf{R}^4 \to \mathbf{R}^3$

F(x,y,s,t) = (s – y + s + t, x +2s – t, x+y + 3s – 3t) $\forall$ (x,y,s,t) $\in \mathbf{R}^4$

Let find a basis and the dimension of the image and kernel of F.

Solution:

Im F = span{F(1,0,0,0), F(0,1,0,0), F(0,0,1,0), F(0,0,0,1)}

= span{(1,1,1); (-1,0,1); (1,2,3); (1. -1, -3)}.

Consider

$$A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 2 & 3 \\ 1 & -1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{(by elementary row operations)}$$

Therefore, dim ImF = rank A = 2; and a basis of Im F is

S = {(1,1,1); (0,1,2)}.

Ker F is the solution space of homogeneous linear system:

$$\begin{cases} x - y + s + t = 0 \\ x + 2s - t = 0 \\ x + y + 3s - 3t = 0 \end{cases} \qquad \text{(I)}$$

The dimension of ker F is easily found by the dimension formula

dim Ker F = dim ($\mathbf{R}^4$) – dim ker F = 4-2 = 2

To find a basis of ker F we have to solve the system (I), say, by Gauss elimination.

$$\begin{pmatrix} 1 & -1 & 1 & 1 \\ 1 & 0 & 2 & -1 \\ 1 & 1 & 3 & -3 \\ & & & \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 1 & 1 \\ 0 & 1 & 1 & -2 \\ 0 & 2 & 2 & -4 \\ & & & \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 1 & 1 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 \\ & & & \end{pmatrix}$$

Hence, (I) $\Leftrightarrow \begin{cases} x - y + s + t = 0 \\ y + s - 2t = 0 \end{cases}$

$\Leftrightarrow$y = -s +2t; x = -2s – 3t for arbitrary s and t.

This yields (x,y,s,t) = (-s +2t, -2s – 3t, s,t)

= s(-1,-2, 1, 0)+ t(2,-3,0,1) $\forall$ s,t $\in$ 1

$\Rightarrow$ Ker F = span {(-1,-2, 1, 0), (2,-3,0,1)}. Since this two vectors are linearly independent, we obtain a basis of KerF as {(-1,-2, 1, 0), (2,-3,0,1)}.

2.7. Applications to systems of linear equations:

Consider the homogeneous linear system AX = 0 where A $\in$ $M_{mxn}(\mathbf{R})$ having rank A = r. Let F: M_{nx1} (**R**) $\to$ M_{mx1} (**R**) be the linear mapping defined by FX = AX $\forall$ X $\in$ M_{nx1} (**R**); and let N be the solution space of the system AX = 0. Then, we have that N = ker F. Therefore,

dim N = dim KerF = dim(M_{nx1}(**R**) – dim ImF = n – dim(colsp (A))

= n – rankA = n – r.

We thus have found again the property that the dimension of solution space of homogeneous linear system AX = 0 is equal to n – rankA, where n is the number of the unknowns.

2.8. Theorem: Let F: V $\to$ W be a linear mapping. Then the following assertions hold.

1) F is in injective if and only if Ker F = {0}

2) In case V and W are of finite – dimension and dim V = dim W, we have that F is in injective if and only if F is surjective.

PROOF:

1) "$\Rightarrow$" let F be injective. Take any x $\in$ Ker F

Then, F(x) = O = F(O) $\Rightarrow$ x = O. This yields, Ker F = {O}

"$\Leftarrow$": Let Ker F = {O}; and x_1, x_2 $\in$ V such that $f(x_1) = f(x_2)$. Then

$O = F(x_1) – F(x_2) = F(x_1 – x_2) \Rightarrow x_1 – x_2 \in \ker F = \{O\}$.

$\Rightarrow x_1 – x_2 = O \Rightarrow x_1 = x_2$. This means that F is injective.

2) Let dim V = dim W = n. Using the dimension formula we derive that: F is injective $\Leftrightarrow$ Ker F = {0} $\Leftrightarrow$ dim Ker F = 0

$\Leftrightarrow$ dim Im F = n (since dim Ker F + dim Im F = n)

$\Leftrightarrow$ Im F =W (since dim W = n) $\Leftrightarrow$ F is surjective. (q.e.d.)

The following theorem gives a special example of vector space.

2.9. Theorem: Let V and W be vector spaces over **K**. Denote by

Hom (V,W)= {f: V $\to$ W | f is linear}. Define the vector addition and scalar multiplication as follows:

$(f+g)(x) = f(x) + g(x)\ \forall f,g \in \text{Hom}(V,W);\ \forall x \in V$

$(\lambda f)(x) = \lambda f(x)\ \forall \lambda \in \mathbf{K};\ \forall f \in \text{Hom}(V,W);\ \forall x \in V$

Then, Hom (V,W) is a vector space over **K**.

III. Matrices and linear mappings

3.1. Definition. Let V and W be two vector spaces of finite dimension over **K**, with dim V = n; dim W = m. Let $S = \{v_1, v_2, ...v_n\}$ and $U= \{u_1, u_2 ,..., u_m\}$ be bases of V and W, respectively. Suppose the mapping F: V $\to$ W is linear. Then, we can express

$F(v_1) = a_{11} u_1 +a_{21} u_2 +... + a_{m1}u_m$

$F(v_2) = a_{12} u_1 +a_{22} u_2 +... + a_{m2}u_m$

$\vdots$

$F(v_n) = a_{1n} u_1 +a_{2n} u_2 +... + a_{mn}u_m$

The transposition of the above matrix of coefficients, that is the matrix $\begin{pmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} \dots & a_{mn} \end{pmatrix}$,

denoted by $[F]_U^S$, is called the matrix representation of F with respect to bases S and U (we write [F] when the bases are clearly understood)

Example: Consider the linear mapping F: $\mathbf{R}^3 \to \mathbf{R}^2$; F(x,y,z) = (2x -4y +5z, x+3y-6z); and the two usual bases S and U of $\mathbf{R}^3$ and $\mathbf{R}^2$, respectively. Then,

F(1,0,0) = (2,1) = 2(1,0) + 1(0,1)

F(0,1,0) = (-4,3) = -4(1,0) + 3(0,1)

F(0,0,1) = (5,-6) = 5(1,0) + (-6)(0,1)

Therefore, $[F]_U^S = \begin{bmatrix} 2 & -4 & 5 \\ 1 & 3 & -6 \end{bmatrix}$

3.2. Theorem: Let V and W be two vector spaces of finite dimension, F: V → W be linear, and S, U be bases of V and W respectively. Then, for any X ∈ V,

$$[F]_U^S\ [X]_S = [F(X)]_U. \qquad (3.1)$$

PROOF:

$$[F]_U^S\ [X]_S = [a_{jk}]\ [X]_S = \begin{pmatrix} \sum_{k=1}^{n} a_{1k}x_k \\ \vdots \\ \sum_{k=1}^{n} a_{mk}x_k \end{pmatrix} \text{ for } [X]_S = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Compute now, $F(X) = F\left(\sum_{J=1}^{n} x_J v_J\right) = \sum_{J=1}^{n} x_J F(v_J) = \sum_{J=1}^{n} x_J \sum_{k=1}^{m} a_{kJ} u_k$

$$= \sum_{J=1}^{m}\sum_{k=1}^{n} x_J a_{Jk} u_k = \sum_{k=1}^{n}\left(\sum_{J=1}^{m} a_{kJ} x_J\right) u_k .$$

For S = $\{v_1, v_2, ... v_n\}$; U = $\{u_1, u_2, ... u_m\}$

$$\text{Therefore } [F(X)]_U = \begin{pmatrix} \sum_{J=1}^{m} a_{1J}x_J \\ \sum_{J=1}^{m} a_{2J}x_J \\ \vdots \\ \sum_{J=1}^{m} a_{mJ}x_J \end{pmatrix} = [F]_U^S\ [X]_S. \qquad \text{(q.e.d)}$$

Example: Return to the above example, F: $\mathbf{R}^3 \to \mathbf{R}^2$

F(x,y,z) = (2x-4y+5z, x+3y-6z). Then,

$$[F(x,y,z)]_U = \begin{pmatrix} 2x - 4y + 5z \\ x + 3y - 6z \end{pmatrix} = \begin{bmatrix} 2 & -4 & 5 \\ 1 & 3 & -6 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = [F]_U^S\ [(x,y)]_S$$

3.3. Corollary: If A is a matrix such that A $[X]_S = [F(X)]_U$ for all X $\in$ V, then

$$A = [F]_U^S .$$

Note: For a linear mapping F: V $\to$ W with finite–dimensional spaces V and W. If we fix two bases S and U of V and W, respectively, we can replace the action of F by the multiplication of its matrix representation $[F]_U^S$ through equality (3.1). Moreover, the mapping

$$\Im: \text{Hom}(V,W) \to M_{mxn}(\mathbf{R})$$

$$F \mapsto [F]_U^S$$

is an isomorphism. In other words, Hom(V,W) $\cong$ M_{mxn} (**K**). The following theorem gives the relation between the matrices and the composition of mappings.

3.4. Theorem. Let V, E, W be vector spaces of finite dimension, and R, S, U be bases of V, E, W, respectively. Suppose that f: V $\to$E and g: E $\to$W are linear mappings then $[g\circ f]_U^R = [g]_U^S [f]_S^R$

3.5. Change of bases: Let V, W be vector spaces of finite–dimension, F: V $\to$ W be a linear mapping; S and U be bases of V and W; S', U' be other bases of V and W, respectively. We want to find the relation bet ween $[F]_U^S$ and $[F]_{U'}^{S'}$. To do that, let P be the change–of–basis matrix from S to S', and Q be change–of–basis matrix from U to U', respectively.

By Theorem 3.2 and Corollary 3.3, we have that.

$[F]_U^S$ $[X]_S = [F(X)]_U$

$[X]_S = P[X]_{S'}$ and $[F(X)]_U = Q[F(X)]_{U'}$

Therefore, $[F]_U^S$ $[X]_S = [F]_U^S$ $P[X]_{S'} = [F(X)]_U = Q[F(X)]_{U'}$

Hence, Q^{-1} $[F]_\varepsilon^S$ $P[X]_{S'} = [F(X)]_{U'}$ for all X $\in$ V

Therefore, by Corollary 3.3,

$$[F]_{U'}^{S'} = Q^{-1}\ [F]_U^S\ P. \tag{3.2}$$

Example: Consider F: $\mathbf{R}^3 \to \mathbf{R}^2$

F(x,y,z) = (2x-4y+z, x + 4y -5z).

Let S, U be usual bases of $\mathbf{R}^3$ and $\mathbf{R}^2$, respectively. Then

$$[F]_U^S = \begin{bmatrix} 2 & -4 & 1 \\ 1 & 3 & -5 \end{bmatrix}$$

Now, consider the basis S' = {(1,1,0); (1,-1,1); (0,1,1)} of $\mathbf{R}^3$ and the basis U' = {(1,1); (-1,1)} of $\mathbf{R}^2$. Then, the change – of – basis matrix from S to S' is P = $\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$

and the change – of – basis matrix from U to U' is Q = $\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

Therefore, $[F]_{U'}^{S'} = Q^{-1} [F]_U^S P = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{bmatrix} 2 & -4 & 1 \\ 1 & 3 & -5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -5/2 \\ 3 & -7 & 1/2 \end{bmatrix}$

IV. Eigenvalues and Eigenvectors

Consider a field **K**; (**K** = **R** or **C**). Recall that $M_n(\mathbf{K})$ is the set of all n-square matrices with entries belonging to **K**.

4.1. Definition: Let A be an n-square matrix, $A \in M_n(\mathbf{K})$. A number $\lambda \in \mathbf{K}$ is called an eigenvalue of A if there exists a non–zero column vector $X \in M_{nx1}(\mathbf{K})$ for which AX = λX; and every vector satisfying this relation is then called an eigenvector of A corresponding to the eigenvalue λ. The set of all eigenvalues of A, which belong to **K**, is called the spectrum on **K** of A, denoted by $Sp_{\mathbf{K}}A$. That is to say,

$$Sp_{\mathbf{K}}A = \{ \lambda \in \mathbf{K} \mid \exists X \in M_{nx1}(\mathbf{K}), X \neq 0 \text{ such that } AX = \lambda X\}.$$

Example: A = $\begin{pmatrix} -5 & 2 \\ 2 & -2 \end{pmatrix} \in M_2(\mathbf{R})$. We find the spectrum of A on **R** , that is $Sp_{\mathbf{R}}$ A.

By definition we have that

$$\lambda \in Sp_{\mathbf{R}}A \Leftrightarrow \exists X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ such that } AX = \lambda X$$

$\Leftrightarrow$ The system (A - λI)X = 0 has nontrivial solution X ≠ 0

$\Leftrightarrow$ det (A - λI) = 0

$\Leftrightarrow \begin{vmatrix} -5-\lambda & 2 \\ 2 & -2-\lambda \end{vmatrix} = 0 \Leftrightarrow \lambda^2 + 7\lambda + 6 = 0$

$\Leftrightarrow \left[\begin{array}{l} \lambda = -1 \\ \lambda = -6 \end{array}\right.$, therefore, $Sp_{\mathbf{R}}A = \{-1, -6\}$

To find eigenvectors of A, we have to solve the systems

$$(A - \lambda I)X = 0 \text{ for } \lambda \in Sp_{\mathbf{R}}A. \qquad (4.1)$$

For $\lambda = \lambda_1 = -1$: we have that

$$(4.1) \Leftrightarrow \begin{bmatrix} -5+1 & 2 \\ 2 & -2+1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{cases} -4x_1 + 2x_2 = 0 \\ 2x_1 - x_2 = 0 \end{cases} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Therefore, the eigenvectors corresponding to $\lambda_1 = -1$ are of the form $k\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ for all $k \neq 0$

For $\lambda = \lambda_2 = -6$: $\quad (4.1) \Leftrightarrow \begin{cases} x_1 + 2x_2 = 0 \\ 2x_1 + 4x_2 = 0 \end{cases} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_2\begin{pmatrix} -2 \\ 1 \end{pmatrix}$

Therefore, the eigenvectors corresponding to $\lambda_2 = -6$ are of the form $k\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ for all $k \neq 0$.

4.2. Theorem: Let $A \in M_n(\mathbf{K})$. Then, the following assertion are equivalent.

(i) The scalar $\lambda \in \mathbf{K}$ is an eigenvalue of A

(ii) $\lambda \in \mathbf{K}$ satisfies det (A - λI) = 0

PROOF:

(i) $\Leftrightarrow \lambda \in Sp_{\mathbf{K}}(A) \Leftrightarrow (\exists X \neq 0, X \in M_{nx1}(\mathbf{K})$ such that $AX = \lambda X)$

$\Leftrightarrow$ The system (A - λI)X = 0 has a nontrivial solution

$\Leftrightarrow$ det (A - λI) = 0 (ii)

4.3. Definition. Let $A \in M_n(\mathbf{K})$. Then, the polynomial $\chi(\lambda)$ = det (A - λI) is called the **characteristic polynomial** of A and the equation $\chi(\lambda) = 0$ is called **characteristic**

equation of A. Then, the set $E_\lambda = \{X \in M_{nx1}(\mathbf{K}) \mid AX = \lambda X\}$ is called the **eigenspace** of A corresponding to λ.

Remark: If X is an eigenvector of A corresponding to an eigenvalue $\lambda \in \mathbf{K}$, so is kX for all $k \neq 0$. Indeed, since $AX = \lambda X$ we have that $A(kX) = kAX = k(\lambda X) = \lambda(kX)$. Therefore, kX is also an eigenvector of A for all $k \neq 0$.

It is easy to see that, for $\lambda \in Sp_{\mathbf{K}}(A)$, E_λ is a subspace of $M_{nx1}(\mathbf{K})$.

Note: For $\lambda \in Sp_{\mathbf{K}}(A)$; the eigenspace E_λ coincides with Null(A - λI); and therefore $\dim E_\lambda = n - \text{rank}(A - \lambda I) > 0$.

Examples: For $A = \begin{pmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{pmatrix}$ we have that

$|A - \lambda I| = 0 \Leftrightarrow -\lambda^3 - \lambda^2 + 21\lambda + 45 = 0$

$\Leftrightarrow \lambda_1 = 5; \lambda_2 = \lambda_3 = -3$. Therefore, $Sp_{\mathbf{K}}(A) = \{5; -3\}$

For $\lambda_1 = 5$, consider $(A - 5I)X = 0 \Leftrightarrow X = x_1 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \forall x_1$

Therefore, $E_5 = \text{Span}\left\{\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}\right\}$.

For $\lambda_2 = \lambda_3 = -3$, solve $(A+3I)X=0 \Leftrightarrow X = x_2 \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \forall x_2, x_3.$

Therefore, $E_{(-3)} = \text{Span}\{\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}\}$

4.5. An application: **Stretching of an elastic membrane**

An elastic membrane in the x_1x_2 plane with boundary $x_1^2 + x_2^2 = 1$ is stretched so that a point $P(x_1,x_2)$ goes over into the point $Q(y_1,y_2)$ given by

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = AX = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Find the "principal directions", that is, directions of the position vector $X \neq 0$ of P for which the direction of the position vector Y of Q is the same or exactly opposite (or, in other words, X and Y are linearly dependent). What shape does the boundary circle take under this deformation?

Solution: We are looking for X such that $Y = \lambda X$; $X \neq 0$

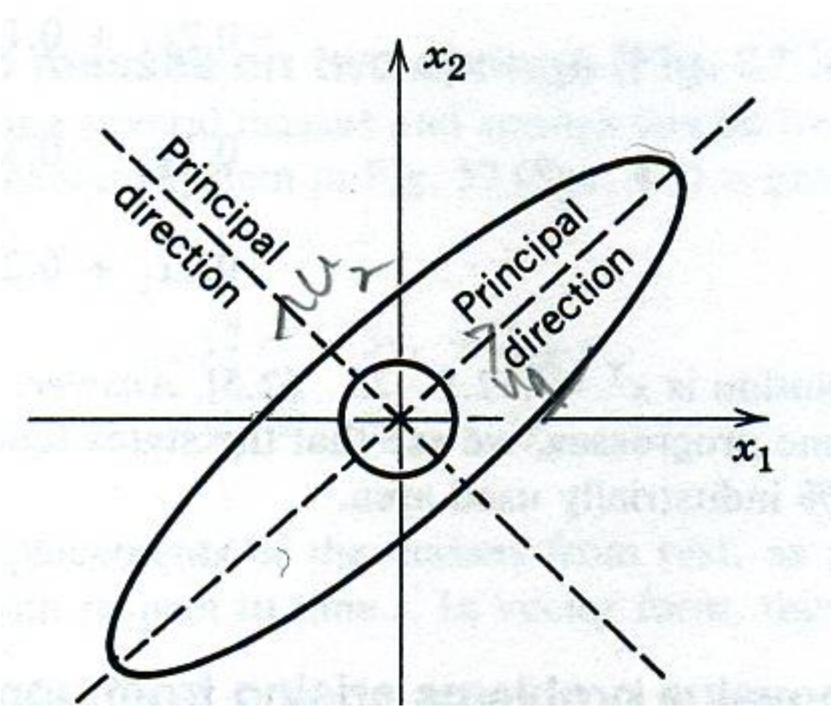
$\Leftrightarrow AX = \lambda X$. This means that we have to find eigenvalues and eigenvectors of A.

Solve $\det(A - \lambda I) = 0 \Leftrightarrow \begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} = 0 \Leftrightarrow \lambda_1 = 8, \lambda_2 = 2$

For $\lambda_1 = 8$; $E_8 = \text{Span}\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$

For $\lambda_2 = 2$; $E_2 = \text{Span}\left\{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}$

We choose $u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$; $u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Then, u_1, u_2 give the principal directions. The eigenvalues show that in the principal directions the membrane is stretched by factors 8 and 2, respectively.



Accordingly, if we choose the principal directions as directions of a new Cartesian v_1,v_2 – coordinate system, say,

$$v_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right);\ v_2 = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$$

Then $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$; $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

$$\Rightarrow \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{8}{\sqrt{2}}x_1 + & \frac{8}{\sqrt{2}}x_2 \\ \frac{-2}{\sqrt{2}}x_1 & \frac{2}{\sqrt{2}}x_2 \end{pmatrix}$$

$\Rightarrow \frac{z_1^2}{32} + \frac{z_2^2}{2} = 2(x_1^2 + x_2^2) = 2 \Leftrightarrow \frac{z_1^2}{64} + \frac{z_2^2}{4} = 1$ and we obtain a new shape as an ellipse.

V. Diagonalizations

5.1. Similarities: An nxn matrix B is called similar to an nxn matrix A if there is an nxn – nonsingular matrix T such that.

$$B = T^{-1}AT \qquad (5.1)$$

5.2. *Theorem:* If B is similar to A, then B and A have the same characteristic polynomial, therefore have the same set of eigenvalues.

PROOF:

$B = T^{-1}AT \Rightarrow |B - \lambda I| = |T^{-1}AT - T^{-1}\lambda IT|$

$= |T^{-1}(A - \lambda I)T| = |T^{-1}|.|A - \lambda I|.|T| = |A - \lambda I|$

5.3. Lemma: Let $\lambda_1; \lambda_2 \dots \lambda_k$ be distinct eigenvalues of an nxn matrix A. Then, the corresponding eigenvectors $x_1, x_2, \dots x_k$ form a linearly independent set.

PROOF: Suppose that the conclusion is false. Let r be the largest integer such that $\{x_1; x_2;\dots x_r\}$ is linearly independent. Then, $r < k$ and the set $\{x_1,..,x_{r+1}\}$ is linearly dependent. This means that, there are scalars $C_1, C_2,\dots C_{r+1}$ not all zero, such that.

$$C_1x_1 + C_2x_2 + \dots + C_{r+1}x_{r+1} = 0 \qquad (3.3)$$

$$\Rightarrow \lambda_{r+1}\sum_{i=1}^{r+1} C_i x_i = 0 \text{ and } A\left(\sum_{i=1}^{r+1} C_i x_i\right) = 0$$

Then $\begin{cases} \sum_{i=1}^{r+1} \lambda_{r+1} C_i x_i = 0 \\ \sum_{i=1}^{r+1} \lambda_i C_i x_i = 0 \end{cases} \Rightarrow \sum_{i=1}^{r} (\lambda_{r+1} - \lambda_i) C_i x_i = 0$

Since $x_1, x_2, ...x_r$ are linearly independent, we have that $(\lambda_{r+1} - \lambda_i)C_i = 0\ \forall\ i = 1,2...,r$ it follows that $C_i = 0\ \forall i = 1,2,...r$. By (3.3), $C_{r+1}x_{r+1} = 0$

Therefore $C_{r+1} = 0$ because $x_{r+1} \neq 0$. This contradicts with the fact that not all $C_1,... C_{r+1}$ are zero.

5.4. Theorem: If an nxn matrix A has n distinct eigenvalues, then it has n linearly independent vectors.

Note: There are nxn matrices which do not have n distinct eigenvalues but they still have n linearly independent eigenvectors, e.g.,

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}; \ |A - \lambda I| = (\lambda+1)^2(\lambda-2) = 0 \Leftrightarrow \left[\begin{array}{l} \lambda = -1 \\ \lambda = 2 \end{array}\right.,$$

and A has 3 linearly independent vectors $u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$; $u_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$; $u_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

5.5. Definition: A square nxn matrix A is said to be diagonalizable if there is a nonsingular nxn matrix T so that $T^{-1}AT$ is diagonal, that is, A is similar to a diagonal matrix.

5.6. Theorem: Let A be an nxn matrix. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.

PROOF: If A is diagonalizable, then $\exists T$, $T^{-1}AT = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$ - a diagonal matrix.

Let $C_1, C_2,... C_n$ be columns of T, that is, $T = [C_1 C_2 ... C_n]$.

Then AT = T$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \Rightarrow \begin{matrix} AC_1 = \lambda_1 C_1 \\ AC_2 = \lambda_2 C_2 \\ \vdots \\ AC_n = \lambda_n C_n \end{matrix}$

$\Rightarrow$ A has n eigenvectors $C_1, C_2, ..., C_n$. Obviously, $C_1, C_2, ... C_n$ are linearly independent since T is nonsingular.

Conversely, A has n linearly independent eigenvectors $C_1, C_2, ... C_n$ corresponding to eigenvalues $\lambda_1, \lambda_2, ...\lambda_n$, respectively.

Set T = [C_1 C_2 ... C_n]. Clearly, AT = T$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$

Therefore $T^{-1}AT = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$ is a diagonal matrix. (q.e.d)

5.7. Definition: Let A be a diagonalizable matrix. Then, the process of finding of T such that $T^{-1}AT$ is a diagonal matrix, is called the diagonalization of A.

We have the following **algorithm of diagonalization** of nxn matrix A.

Step 1: Solve the characteristic equation to find all eigenvalues $\lambda_1, \lambda_2, ...\lambda_n$

Step 2: For each λ_i solve $(A - \lambda_i I)X = 0$ to find all the linearly independent eigenvectors of A. If A has less than n linearly independent eigenvectors, then conclude that A can not be diagonalized. If A has n linearly independent eigenvectors, then come to next step.

Step 3: Let $u_1, u_2, ..., u_n$ be n linearly independent eigenvectors of A found in Step 2. Then, set T = [u_1 u_2...u_n] (that is columns of T are $u_1, u_2, ..., u_n$, respectively). By theorem 5.6 we have that $T^{-1}AT = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$ where u_i is the eigenvectors corresponding to the eigenvalue λ_i, i = 1,2,... n, respectively.

Example: A = $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$;

Characteristic equation: $|A - \lambda I| = (\lambda+1)^2(\lambda-2) = 0 \Leftrightarrow \left[\begin{array}{l} \lambda_1 = \lambda_2 = -1 \\ \lambda_3 = 2 \end{array}\right.$

For $\lambda_1 = \lambda_2 = -1$: Solve (A+I)X = 0 for X = $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ we have $x_1+x_2+x_3 = 0$;

There are two linearly independent eigenvectors $u_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$; and $u_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ corresponding to the eigenvector -1.

For $\lambda_3 = 2$, solve (A-2I)X = 0 for X = $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ we obtain that

$\begin{cases} -2x_1 + x_2 + x_3 = 0 \\ x_1 - 2x_2 + x_3 = 0 \\ x_1 + x_2 - 2x_3 = 0 \end{cases} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. There is one linearly independent eigenvector $u_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ corresponding to the eigenvector $\lambda_3 = 2$. Therefore, A has 3 linearly independent eigenvectors u_1, u_2, u_3. Setting now T = $\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}$ we obtain that

$T^{-1}AT = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ finishing the diagonalization of A.

VI. Linear operators (transformations)

6.1. Definition: A linear mapping F: V→ V(from V to itself) is called a linear operator (or transformation).

6.2. Definition: A linear operator F: V $\to$V is called nonsingular if Ker F = {0}.

By Theorem 2.8 we obtain the following results on linear operators.

6.3. Theorem: Let V be a vector space of finite dimension and F: V $\to$V be a linear operator. Then, the following assertions are equivalent.

i) F is nonsingular.

ii) F is injective.

iii) F is surjective.

iv) F is bijective.

6.4. Linear operators and matrices: Let V a vector space of finite dimension, and F: V $\to$V be a linear operator. Let S be a basis of V. Then, by definition 3.1, we can construct the matrix $[F]_S^S$. This matrix is called the matrix representation of F corresponding to S, denoted by $[F]_S = [F]_S^S$.

By Theorem 3.2. and Corollary 3.3 we have that $[F]_S [x]_S = [Fx]_S \ \forall\ x \in V$.

Conversely, if A is an nxn square matrix satisfying $A[x]_S = [Fx]_S$ for all $x \in V$, then $A = [F]_S$.

Example: Let F: $\mathbf{R}^3 \to \mathbf{R}^3$; F(x,y,z) = (2x – 3y+z,x +5y -3z, 2x –y – 5z). Let S be the usual basis in $\mathbf{R}^3$, S = {(1,0,0); (0,1,0); (0,0,1)}. Then, $[F]_S = \begin{pmatrix} 2 & -3 & 1 \\ 1 & 5 & -3 \\ 2 & -1 & -5 \end{pmatrix}$

6.5. Change of bases: Let F: V $\to$V be an operator where V is of n–dimension, and let S and U be two different bases of V. Putting $A = [F]_S$; $B = [F]_U$ and supposing that P is the change–of–basis matrix from S to U, we have that

$$B = P^{-1}AP$$

(this follows from the formula (3.2) in Section 3.4). Therefore, we obtain that two matrices A and B represent the same linear operator F if and only if they are similar.

6.6. Eigenvectors and eigenvalues of operators

Similarly to square matrices, we have the following definition of eigenvectors and eigenvalues of an operator T: V $\to$V where V is a vector space over **K**.

Definition: A Scalar $\lambda \in \mathbf{K}$ is called an eigenvalue of T if there exists a nonzero vector $v \in V$ for which $T(v) = \lambda v$. Then, every vector satisfying this relation is called an eigenvector of T corresponding to λ.

The set of all eigenvalues of T in **K** is denoted by $Sp_{\mathbf{K}}T$ and is called spectral set of T on **K**.

Example: Let $f : \mathbf{R}^2 \to \mathbf{R}^2$; $f(x,y) = (2x+y,3y)$

$\lambda \in Sp_{\mathbf{K}}f \Leftrightarrow \exists(x,y) \neq (0,0)$ such that $f(x,y) = \lambda(x,y)$

$\Leftrightarrow \exists\ (x,y) \neq (0,0)$ such that $\begin{cases} (2-\lambda)x + y = 0 \\ (3-\lambda)y = 0 \end{cases}$

$\Leftrightarrow$ the system $\begin{cases} (2-\lambda)x + y = 0 \\ (3-\lambda)y = 0 \end{cases}$ has a nontrivial solution $(x,y) \neq (0,0)$

$\Leftrightarrow \begin{vmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{vmatrix} = 0 \Leftrightarrow (2-\lambda)(3-\lambda)=0$

$\Leftrightarrow \begin{cases} \lambda = 2 \\ \lambda = 3 \end{cases} \Leftrightarrow \lambda \in Sp_{\mathbf{R}}[f]_S$ where $[f]_S = \begin{bmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix}$ is the matrix representation of f with respect to the usual basis S of $\mathbf{R}^2$.

Since $f(v) = \lambda v \Leftrightarrow [f(v)]_U = \lambda[v]_U \Leftrightarrow [f]_U[v]_U = \lambda[v]_U$ for any basis U of $\mathbf{R}^2$, we can easily obtain that λ is an eigenvalue of f if and only if λ is an eigenvalue of $[f]_U$ for any basis U of $\mathbf{R}^2$.

Moreover, by the same arguments we can deduce the following theorem.

6.7. Theorem. Let V be a vector space of finite dimension and $F: V \to V$ be a linear operator. Then, the following assertions are equivalent.

i) $\lambda \in \mathbf{K}$ is an eigenvalue of F

ii) $(F - \lambda I)$ is a singular operator, that is, $Ker\,(F - \lambda I) \neq \{0\}$

iii) λ is an eigenvalue of the matrix $[F]_S$ for any basis S of V.

Example: Let $F: \mathbf{R}^3 \to \mathbf{R}^3$ be defined by

$F(x,y,z) = (y+z,x+z,x+y)$; and S be the usual basis of $\mathbf{R}^3$.

Then $[F]_S = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ and $Sp_{\mathbf{R}}F = Sp_{\mathbf{R}}[F]_S = \{-1,2\}$

Remark: Let T: V→ V be an operator on finite dimensional space V and S be a basis of V. Then,

v∈ V is an eigenvector of T ⇔ $[v]_S$ is an eigenvector of $[T]_S$

This equivalence follows from the fact that $[T]_S[v]_S = [Tv]_S$.

Therefore, we can deduce the finding of eigenvectors and eigenvalues of an operator T to that of its matrix representation $[T]_S$ for any basis S of V.

6.8. Diagonalization of a linear operator:

Definition: The operator T: V→ V (where V is a finite-dimensional vector space) is said to be diagonalizable if there is a basis S of V such that $[T]_S$ is a diagonal matrix. The process of finding S is called the diagonalization of T. Since, in finite-dimensional spaces, we can replace the action of an operator T by that of its matrix representation, we thus have the following theorem whose proof is left as an excercise.

Theorem: let V be a vector space of finite dimension with dim V =n, and T: V→ V be linear operator. Then, the following assertions are equivalent.

i) T is diagonalizable.

ii) T has n linearly independent eigenvectors.

iii) There is a basis of V, which are consisted of n eigenvectors of T.

Example: Let consider above example T: $\mathbf{R}^3 \to \mathbf{R}^3$ defined by

$$T(x,y,z) = (y+z,x+z,x+y)$$

To diagonalize T we choose any basis of $\mathbf{R}^3$, say, the usual basis S. Then, we write

$$[T]_S = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

The eigenvalues of T coincide with the eigenvalues of $[T]_S$; and they are easily computed by solving the characteristic equation

$$\text{Det}([T]_S - \lambda I) = \begin{vmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{vmatrix} = 0 \Leftrightarrow (\lambda+1)^2(\lambda-2) = 0$$

Next, using the fact that v ∈ V is a eigenvector of T if and only if $[v]_S$ is an eigenvector of $[T]_S$, we have that for $\lambda = -1$, we solve $([T]_S - \lambda I)X = 0 \Leftrightarrow ([T]_S + I)X = 0$

$$\Leftrightarrow \begin{cases} x_1 + x_2 + x_3 = 0 \\ x_1 + x_2 + x_3 = 0 \\ x_1 + x_2 + x_3 = 0 \end{cases} \text{(for x = } \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ corresponding to v = } (x_1, x_2, x_3))$$

Therefore, there are two linearly independent eigenvectors corresponding to $\lambda = -1$; these are $v_1 = (1,-1,0)$ and $v_2 = (1,0,-1)$.

For $\lambda = 2$, we solve $([T]_S - 2I)X = 0$ (for x = $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ corresponding to v = (x_1,x_2,x_3)).

We then have

$$\Leftrightarrow \begin{cases} -2x_1 + x_2 + x_3 = 0 \\ x_1 - 2x_2 + x_3 = 0 \\ x_1 + x_2 - 2x_3 = 0 \end{cases} \Leftrightarrow (x_1,x_2,x_3) = x_1(1,1,1) \ \forall x_1$$

Thus, there is one linearly independent eigenvector corresponding to $\lambda = 2$, that may be chosen as $v_3 = (1,1,1)$.

Clearly, v_1, v_2, v_3 are linearly independent and the set $U = \{v_1,v_2,v_3\}$ is a basis of $\mathbf{R}^3$ for which the matrix representation $[T]_U$ is diagonal.

Chapter 7: Euclidean Spaces

I. Inner product spaces

1.1. Definition: let V be a vector space over **R**. Suppose to each pair of vectors u, v $\in$ V there is assigned a real number, denoted by $<u,v>$. Then, we obtain a function:

$$V \times V \to \mathbf{R}$$

$$(u,v) \mapsto <u,v>$$

This function is called a real inner product on V if it satisfies the following axioms:

(I_1): $< au + bw,v > = a< u,v > + b < w,v > \ \forall\, u,v,w \in V$ and $a,b \in \mathbf{R}$ (Linearity)

(I_2): $< u,v > = <v, u> \ \forall\, u,v \in V$ (Symmetry)

(I_3): $<u,u> \geq 0 \quad \forall\, u \in V$; and $<u,u> = 0$ if and only if u = O-the null vector of V

(Positive definite).

The vector space V with a inner product is called a (real) Inner Product Space (we write IPS to stand for the phrase: "Inner Product Space").

Note:

a) The axiom (I_1) is equivalent to

1) $< u_1 +u_2,v> = <u_1,v> + <u_2,v> \ \forall\, u_1, u_2, v \in V$.

2) $<ku,v> = k<u,v> \forall\, u, v \in V$

b) Using (I_1) and (I_2) we obtain that

$$< u, \alpha v_1+\beta v_2> = <\alpha v_1 + \beta v_2, u> = \alpha <v_1,u> +\beta<v_2,u>$$

$$= \alpha <u,v_1> + \beta<u,v_2> \ \forall\, \alpha,\beta \in \mathbf{R} \text{ and } u,v_1,v_2 \in V$$

Examples: Take V= $\mathbf{R}^n$; let an inner product be defined by

$$< (u_1,u_2...u_n), (v_1,v_2,...v_n)> = \sum_{i=1}^{n} u_i v_i \ .$$

Then, we can check

(I_1): For $u = (u_1,u_2...u_n)$;$w = (w_1,w_2,...w_n)$ and $v= (v_1,v_2,...v_n)$, and $a,b \in \mathbf{R}$, we have

$$<au +bw,v> = \sum_{i=1}^{u}(au_i + bw_i)v_i = \sum_{i=1}^{u} au_i v_i + b\sum_{i=1}^{u} w_i v_i = a<u,v> + b<w,v>$$

I_2): <u,v> = $\sum_{i=1}^{u} w_i v_i = \sum_{i=1}^{u} v_i u_i = < v, u >$

I_3): <u,u> = $\sum_{i=1}^{u} u_i^2 \geq 0$ and <u.u> = 0 $\Leftrightarrow \sum_{i=1}^{u} u_i^2 = 0$

$\Leftrightarrow u_1 = u_2 = \cdots = u_n = 0 \Leftrightarrow u = (0,0,\ldots,0)$-null vector of $\mathbf{R}^n$.

Therefore, we obtain that < · , · > is an inner product making $\mathbf{R}^n$ the IPS. This inner product is called usual scalar product (or dot product) on $\mathbf{R}^2$, and is sometimes denoted by u · v = <u,v>.

2) Let C[a,b] = {f: [a,b] $\to\mathbf{R}$ | f is continuous} be the vector space of all real, continuous function defined on [a,b] $\subset \mathbf{R}$. Then, one can show that the following assignment

$$<f,g>= \int_a^b f(x).g(x)dx \quad \forall\ f, g \in C[a,b]$$

is an inner product on C[a,b] making C[a,b] an IPS.

3) Consider M_{mxn} (**R**) – the vector space of all real matrices of the size mxn. Then, the following assignment

<A,B> = Tr (A^TB), where Tr ($[C_{ij}]$) = $\sum_{i=1}^{n} C_{ii}$ for an n-square matrix $[C_{ij}]$, is an inner product on M_{mxn} (**R**). It is called the unual inner product on M_{mxn}(**R**).

Specially, when n= 1 we have that the (M_{mx1}(**R**), <.,.>) is an IPS with

$$<X,Y> = X^TY = \sum_{i=1}^{n} x_i y_i \text{ for } X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix};\ Y= \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

1.2. Remark: 1) <O,u> = 0 because <O,u> = < 0.O,u> = 0< O,u> = 0.

2) If u ≠ O then <u,u> > 0.

1.3. Definition: An Inner Product Space of finite dimension is called an **Euclidean space**.

Example of Euclidean spaces:

1) ($\mathbf{R}^n$, < · , · >), where < · , · > is usual scalar product .

2) (M_{mxn}(**R**), < · , · >), where < · , · > is usual inner product .

II. Length (or Norm) of vectors

2.1. Definition: Let (V, $<\cdot,\cdot>$) be an IPS. For each u $\in$V we define $\|u\| = \sqrt{<u,u>}$ and call it the length (or norm) of u.

We now justify this definition by proving the following properties of the length of a vector.

2.2. Proposition. The length of vectors in V has the following properties.

1) $\|u\| \geq 0$ $\forall$ u $\in$ V and $\|u\| = 0 \Leftrightarrow$ u = O.

2) $\|\lambda u\| = |\lambda|\,\|u\|$ $\forall$ $\lambda \in \mathbf{R}$; u$\in$V.

3) $||u+v|| \leq ||u||+||v||$ $\forall$ u, v $\in$ V.

PROOF. The proofs of (1) and (2) are straightforward. We prove (3). Indeed,

(3) $\Leftrightarrow$ $\|u+v\|^2 \leq \|u\|^2 + \|v\|^2 + 2\|u\|\,\|v\|$

$\Leftrightarrow$ $<u+v, u+v> \leq <u,u> + <v,v> + 2\sqrt{<u,u><v,v>}$

By the linearity of the inner product the above inequality is equivalent to

$<u,u> + 2<u,v> + <v,v> \leq <u,u> + 2<v,v> + 2\sqrt{<u,u><v,v>}$

$\Leftrightarrow$ $<u,v> \leq \sqrt{<u,u><v,v>}$

This last inequality is a consequence of the following Cauchy – Schwarz inequality

$$<u,v>^2 \leq <u,u>\,<v,v>. \qquad \text{(C-S)}$$

We now prove (C-S). In fact, for all t $\in$ **R** we have that $<tu+v, tu+v> \geq 0$

$\Leftrightarrow$ $t^2 <u,u> + 2<u,v>t + <v,v> \geq 0$ $\forall$ t $\in$ **R**

If u = 0, the inequality (C-S) is obvious.

If u $\neq$ 0, then we have that $\Delta' \leq 0 \Leftrightarrow <u,v>^2 \leq <u,u>\,<v,v>$ (q.e.d)

2.3. Remarks:

1) If $\|u\| = 1$, then u is called a unit vector.

2) The non–negative real number d(u,v) = $\|u - v\|$ is called the distance between u and v.

3) For nonzero vectors u, v $\in$ V, the angle between u and v is defined to be the angle θ, $0 \le \theta \le \pi$, such that $\cos\theta = \dfrac{<u,v>}{\|u\|.\|v\|}$

4) In the space $\mathbf{R}^n$ with usual scalar product <.,.> we have that the Cauchy –Schwarz inequality (C – S) becomes

$$(x_1y_1 + x_2y_2 + ...+ x_ny_n)^2 \le (x_1^2 + x_2^2 + ... + x_n^2)(y_1^2 + y_2^2 + ... + y_n^2)$$

for x = (x_1, x_2,...x_n) and y = (y_1, y_2,... y_n) $\in \mathbf{R}^n$, which is known as Bunyakovskii inequality.

III. Orthogonality

Throughout this section, (V, < · , · >) is an IPS.

3.1. Definition: the vectors u,v $\in$V are said to be orthogonal if <u,v> = 0, denoted by u $\perp$ v (we also say that u is orthogonal to v).

Note: 1) If u is orthogonal to every v $\in$V, then <u,u> = 0 $\Rightarrow$ u = 0.

2) For u,v $\neq$ 0; if u $\perp$ v, the angle between u and v is $\dfrac{\pi}{2}$

Example: let V = $\mathbf{R}^n$, and < · , · >- the usual scalar product.

(1, 1, 1) $\perp$ (1,-1,0) because <(1,1,-1), (1,-1,0)> = 1.1-1.1+0.0 = 0

3.2. Definition: Let S be a subset of V. The orthogonal complement of S, denoted by $S^\perp$ (read S "perp") consists of those vectors in V which are orthogonal to every vectors of S, that is to say,

$$S^\perp = \{v \in V \mid <v,u> = 0 \ \forall\ u \in S\}.$$

In particular, for a given u $\in$ V; $u^\perp = \{u\}^\perp = \{v \in V \mid v \perp u\}$.

The following properties of $S^\perp$ are easy to prove.

3.3. Proposition: Let V be an ISP. For S $\subset$ V, $S^\perp$ is a subspace of V and $S^\perp \cap S \subset \{0\}$. Moreover, if S is a subspace of V then $S^\perp \cap S = \{0\}$.

Examples:

1) Consider $\mathbf{R}^3$ with usual scalar product. Then, we can compute

$(1,3,-4)^\perp = \{(x,y,z) \in \mathbf{R}^3 \mid x+3y-4z = 0\}$

= Span {(3,-1,0); (0,4,3)} $\subset \mathbf{R}^3$

2) Similarly {(1,-2,1); (-2,1,1)}$^\perp$ = $\left\{(x,y,z)\in R^3 \mid \begin{cases} x-2y+z=0 \\ -2x+y+z=0\end{cases}\right\}$

= Span{(1,1,1)}.

3.4. Definition: The set S $\subset$ V is called orthogonal if each pair of vectors in S are orthogonal; and S is called orthonormal if S is orthogonal and each vector in S has unit length.

To be more concretely, let S = $\{u_1,u_2,...,u_k\}$. Then,

+) S is orthogonal $\Leftrightarrow \langle u_i,u_j\rangle = 0\ \forall\ i \neq j$; where $1 \le i \le k$; $1 \le j \le k$

+) S is orthonormal $\Leftrightarrow \langle u_i,u_j\rangle = \begin{bmatrix} 0 & if\ i\neq j \\ 1 & if\ i=j\end{bmatrix}$ where $1\le i \le k$; $1\le j \le k$.

The concept of orthogonality leads to the following definition of orthogonal and orthonormal bases.

3.5. Definition: A basis S of V is called an orthogonal (orthonormal) basis if S is an orthogonal (orthonormal, respectively) set of vectors. We write ON–basis to stand for "orthonormal basis".

3.6. Theorem: Suppose that S is an orthogonal set of nonzero vectors in V. Then S is linealy independent.

PROOF: Let S = $\{u_1, u_2,\dots, u_k\}$ with $u_i \neq 0\ \forall\ i$ and $\langle u_i, u_j\rangle = 0\ \forall\ i \neq j$

Then, suppose $\lambda_1, \lambda_2 .... \lambda_k \in \mathbf{R}$ such that: $\sum_{i=1}^{k}\lambda_i u_i = 0$.

It follows that

0 = $\left\langle \sum_{i=1}^{k}\lambda_i u_i, u_J\right\rangle = \sum_{i=1}^{k}\langle \lambda_i u_i, u_J\rangle = \sum_{i=1}^{k}\lambda_i\langle u_i, u_J\rangle = \lambda_J\langle u_J, u_J\rangle$ for fixed j $\in\{1,2,...k\}$.

Since $\langle u_j,u_j\rangle \neq 0$ we must have that $\lambda_j = 0$; and this is true for any $j \in \{1,2,...,k\}$. That means, $\lambda_1 = \lambda_2 = ... = \lambda_k = 0$ yielding that S is linearly independent.

Corollary: Let S be an orthogonal set of *n* nonzero vectors in a euclidean space V with dim V = *n*. Then, S in an orthogonal basis of V.

Remark: Let S = { e_1, e_2, ...e_n} be an ON – basis of V and u,v ∈V. Then, the coordinate representation of the inner product in V can be written as

$$<u,v> = \left\langle \sum_{i=1}^{n} u_i e_i , \sum_{J=1}^{n} v_J e_J \right\rangle = \sum_{i=1}^{n}\sum_{J=1}^{n} \langle e_i , e_J \rangle u_i v_J = \sum_{i=1}^{n} u_i v_i = [u]_S^T [v]_S$$

(since $<e_i , e_J> = \begin{cases} 0 & if\ i \neq j \\ 1 & if\ i = j \end{cases}$)

3.7. Pithagorean theorem: Let u ⊥v. Then, we have that $\|u\|^2 + \|v\|^2 = \|u + v\|^2$

PROOF: $\|u + v\|^2$ = <u +v,u+v> = <u,u> + <v,v> +2<u,v> = $\|u\|^2 + \|v\|^2$

Example: Let V = C[-π, π], and <f,g> = $\int_{-\pi}^{\pi} f(t)g(t)dt$ for all f, g ∈ V

Consider S = {1, cost, con2t,..., cosnt, sint, ... sinnt,...}. Due to the fact that $\int_{-\pi}^{\pi} \cos nt.\cos mt$ dt = 0 ∀ m ≠ n and $\int_{-\pi}^{\pi} \sin nt$ sinmt dt = 0 ∀ m ≠ n, and $\int_{-\pi}^{\pi} \cos nt \sin mt$ dt = 0 ∀ m,n, we obtain that S is an orthogonal set.

3.8. Theorem: Suppose S = {u_1, u_2,...u_n} be an orthogonal basis of V and v ∈V. Then

$$v = \sum_{i=1}^{n} \frac{< v,u_i >}{< u_i ,u_i >}.u_i$$

In orther word, the coordinate of v with respect to S is

$$\left(\frac{< v,u_1 >}{< u_1.u_1 >}, \frac{< v,u_2 >}{< u_2 ,u_2 >},..., \frac{< v,u_n >}{< u_n ,u_n >} \right)$$

PROOF. Let V = $\sum_{k=1}^{n}$ $\lambda_k u_k$. We now determine λ_k∀ k = 1,...n. Taking the inner product <v , u_j>, we have that

$<v,u_j> = \left\langle \sum_{k=1}^{n} \lambda_k u_k , u_j \right\rangle = \sum_{k=1}^{n} \lambda_k < u_k , u_j > = \lambda_j <u_j,u_j>$ for fixed j ∈ {1,2,…,n},

because, $<u_k,u_j>=0$ if k ≠ j.

Therefore $\lambda_j = \dfrac{< v,u_j >}{< u_j ,u_j >}$ for all j ∈ {1,2,...,n}. (q.e.d)

Example: V = $\mathbf{R}^3$ with usual scalar product < .,. >

S = {(1,2,1); (2,1,-4); (3,-2,1)} is an orthogonal basis of $\mathbf{R}^3$. Let v = (3, 5, -4). Then to compute the coordinate of v with respect to S we just have to compute

$$\lambda_1 = \frac{<(1,2,1),(3,5,-4)>}{\|(1,2,-1)\|^2} = \frac{9}{6} = \frac{3}{2}$$

$$\lambda_2 = \frac{<(2,1,-4),(3,5,-4)>}{\|(1,2,-4)\|^2} = \frac{27}{21} = \frac{9}{7}$$

$$\lambda_3 = \frac{<(3,-2,1),(3,5,-4)>}{\|(3,-2,1)\|^2} = \frac{-5}{14}$$

Therefore: $(v)_S = \left(\frac{3}{2}, \frac{9}{7}, \frac{-5}{14}\right)$

Note: If S = $\{u_1,u_2,...u_n\}$ is an ON – basis of V; then for $v \in V$ we have that

$$v = \sum_{i=1}^{n} <v,u_i> u_i$$

In other words, $(v)_S = (<v, u_1>; <v,u_2>,..., <v,u_n>)$.

3.9. Gram-Schmidt orthonormalization process:

As shown a bove, the ON – Basis has many important properties. Therefore, it is natural to pose the question: For any Euclidean space, does an ON–Basis exist? Precisely, we have the following problem.

Problem: Let $\{v_1, v_2,...v_n\}$ be a basis of Euclidean space V. Find an ON – basis $\{e_1,e_2,...e_n\}$ of V such that

$$\text{Span}\{e_1,e_2,...,e_k\} = \text{Span}\{v_1,v_2,...,v_k\} \;\forall\, k = 1,2,...,n. \qquad (*)$$

Solution: Firstly, we find an orthogonal basis $\{u_1,u_2,...u_n\}$ of V satisfying:

$$\text{Span}\{u_1,u_2,...,u_n\} = \text{Span}\{v_1,v_2,...,v_k\} \;\forall\, k = 1,2...n.$$

To do that, let us start by putting:

$u_1 = v_1$, then,

$u_2 = v_2 + \alpha_{21}u_1$. Let find the real number α_{21} such that $<u_2, u_1> = 0$.

This is equivalent to $\alpha_{21} = -\frac{<v_2,u_1>}{<u_1,u_1>}$, and hence $u_2 = v_2 - \frac{<v_2,u_1>}{<u_1,u_1>}u_1$.

It is straightforward to see that

$$\text{span}\,\{u_1,u_2\} = \text{span}\,\{v_1,v_2\}.$$

Proceeding in this way, we find u_k in the form

$$u_k = v_k - \frac{< v_k, u_1 >}{< u_1, u_1 >}.u_1 - \frac{< v_k, u_2 >}{< u_2, u_2 >}.u_2 - \ldots - \frac{< v_k, u_{k-1} >}{< u_{k-1}, u_{k-1} >}.u_{k-1}$$

for all $k = 2,\ldots, n$, yielding the set of vectors $\{u_1,u_2,\ldots,u_n\}$ satisfying that $<u_i, u_j> = 0\ \forall\ i \neq j$ and

$$\text{Span}\,\{u_1,u_2,\ldots,u_k\} = \text{Span}\,\{v_1,v_2,\ldots v_k\}\ \text{ for all }\quad k = 1,2,\ldots, n.$$

Secondly, putting: $e_1 = \frac{u_1}{\|u_1\|}$; $e_2 = \frac{u_2}{\|u_2\|}$;... ; $e_n = \frac{u_n}{\|u_n\|}$ we obtain an ON – basis of V satisfying that Span $\{e_1,e_2,\ldots e_k\}$ = span $\{v_1,v_2,\ldots,v_k\}\forall\ k = 1,\ldots,n$.

The above process is called the Gram – Schmidt orthonormalization process.

Example: Let V = $\mathbf{R}^3$ with usual scalar product < .,. >, and

$S = \{v_1 = (1,1,1); v_2 = (0,1,1); v_3 = (0,0,1)\}$ be a basis of V.

We implement the Gram – Schmidt orthonormalization process as follow:

$u_1 = v_1= (1,1,1)$

$$u_2 = v_2 - \frac{< v_2, u_1 >}{< u_1, u_1 >}.u_1 = (0,1,1) - \frac{2}{3}(1,1,1) = \left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)$$

$$u_3 = v_3 - \frac{1}{3}(1,1,1) - \frac{1}{2}\left(-\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right) = \left(0, -\frac{1}{2}, \frac{1}{2}\right)$$

Now, putting: $e_1 = \frac{u_1}{\|u_1\|} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$; $e_2 = \frac{u_2}{\|u_2\|} = \left(-\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}\right)$;

$e_3 = \frac{u_3}{\|u_3\|} = \left(0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, we obtain an ON – basis $\{e_1,e_2,e_3\}$ satisfying that

Span $\{e_1\}$ = span $\{v_1\}$; Span $\{e_1,e_2\}$ = Span$\{v_1,v_2\}$ and span $\{e_1,e_2,e_3\}$ = Span $\{v_1,v_2,v_3\}$ = $\mathbf{R}^3$.

3.10. Corollary: 1) Every Euclidean space has at least one orthonormal basis.

2) Let S = $\{u_1,u_2,...u_k\}$ be an orthonormal subset of vectors is an Euclidean space V with dim V = n > k. Then, we can extend S to an ON – basis U = $\{u_1,u_2, ... ,u_k,u_{k+1},..., u_n\}$ of V.

IV. Projection and least square approximations

4.1. Theorem: Let V be an Euclidean space, W be a subspace of V with W ≠ {0}.

Then for each $v \in V$, there exists a unique couple of vectors $w_1 \in W$ and $w_2 \in W^{\perp}$ such that $v = w_1 + w_2$.

PROOF. By Gram – Schmidt ON process, W has an ON–basis, say S = $\{u_1,u_2,...,u_k\}$. Then, we extend S to an ON – basis $\{u_1,u_2, ... ,u_k, u_{k+1},..., u_n\}$ of V. Now, for $v \in V$ we have the expression

$$v = \sum_{J=1}^{n}\langle v,u_J\rangle u_J = \sum_{J=1}^{k}\langle v,u_J\rangle u_J + \sum_{l=k+1}^{n}\langle v,u_l\rangle u_l .$$

Putting $w_1 = \sum_{J=1}^{k}\langle v,u_J\rangle u_J$; $w_2 = \sum_{l=k+1}^{n}\langle v,u_l\rangle u_l$ we have that $w_1 \in W$ because $\{u_1,..., u_k\}$ is a basis of W. We now prove that $w_2 \in W^{\perp}$. Indeed, taking any $u \in W$, we have that

$$u = \sum_{i=1}^{k}\langle u,u_i\rangle u_i .$$

Therefore, $\langle w_2,u\rangle = \left\langle \sum_{l=k+1}^{n}\langle v,u_l\rangle u_l, \sum_{i=1}^{k}\langle u,u_i\rangle u_i \right\rangle = \sum_{l=k+1}^{n}\sum_{i=1}^{k}\langle u,u_i\rangle\langle v,u_l\rangle\langle u_l,u_i\rangle = 0$

(because $l \neq i$ for $l = k+1,...,n$ and $i = 1,...,k$).

Hence, $w_2 \perp u$ for all $u \in W$. this means that $w_2 \in W^{\perp}$. We next prove that the expression $v = w_1+w_2$ for $w_1 \in W$ and $w_2 \in W^{\perp}$ is unique. To do that, let $v = w_1' + w_2'$ be another expression such that $w_1' \in W$ and $w_2' \in W^{\perp}$. It follows that $v = w_1+w_2 = w_1' + w_2' \Rightarrow w_1 - w_1' = w_2' - w_2$. Hence, $w_1 - w_1' \in W$ and also $w_1 - w_1' = w_2' - w_2 \in W^{\perp}$. This yields $w_1 - w_1' \in W \cap W^{\perp} = \{0\} \Rightarrow w_1 - w_1' = 0 \Rightarrow w_1 = w_1'$. Similarly, $w_2' = w_2$. Therefore, the expression $V = w_1 + w_2$ for $w_1 \in W$ and $w_2 \in W^{\perp}$ is unique.

4.2. Definition:

1) Lef W be a nontrivial subspace of V. Since for all $v \in V$, v can be uniquely expressed as $v = w_1 + w_2$ with $w_1 \in W$ and $w_2 \in W^{\perp}$, we write this relation as $V = W \oplus W^{\perp}$, and call V the direct sum of W and $W^{\perp}$.

2) We denote by P the mapping $P : V \to W$

$$P(v) = w_1 \text{ if } v = w_1 + w_2$$

where $w_1 \in W$ and $w_2 \in W^{\perp}$. Then, P is called the **orthogonal projection** on to W.

Remarks:

1) Looking at the proof of Theorem 4.1, we obtain that:

For an ON – basis $S = \{u_1,..., u_k\}$ of W, the orthogonal projection P on W can be determined by

$$P: V \to W$$

$$P(v) = \sum_{i=1}^{k} \langle v, u_i \rangle u_i \quad \text{for all } v \in V.$$

2) For an othogonal projection $P: V \to W$ we have that, P is a linear operator satisfying properties that $P^2 = P$; Ker $P = W^{\perp}$ and Im$P = W$.

Example: Let V = $\mathbf{R}^3$ with usual scalar product and W = Span{(1,1,0); (1,-1,1)} and $P: V \to W$ be the orthogonal projection onto W, and v = (2,1,3). We now find P(v). To do so, we first find an ON – basis of W. This can be easily done by using Gram-Schmidt process to obtain

$$S = \left\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)\right\} = \{u_1, u_2\}$$

Then $P(v) = \langle v, u_1 \rangle u_1 + \langle v, u_2 \rangle u_2 = (1, -1, -2)$.

4.3. Lemma: Let V be an Eudidean space; W be a subspace of V, and $P: V \to W$ be the orthogonal projection on W. Then $\|v - u\| \geq \|v - P(v)\|$ for all $v \in V$ and $u \in W$.

PROOF: We start by computing

$$\|v - u\|^2 = \|v - P(v) + P(v) - u\|^2$$

Since v – P(v) $\in W^{\perp}$ and P(v) –u $\in$W (because v = P(v)+v – P(v) where P(v) $\in$W and v – P(v) $\in W^{\perp}$), we obtain using Pithagorean Theorem that

$$\|v - P(v) + P(v) - u\|^2 = \|v - P(v)\|^2 + \|P(v) - u\|^2 \geq \|v - P(v)\|^2 .$$

Therefore, $\|v - u\| \geq \|v - P(v)\|$; and the equality happens if and only if u = P(v).

***4.4. Application:* Least square approximation**

For A $\in M_{mxn}(\mathbf{R})$; B $\in M_{mx1}(\mathbf{R})$ consider the following problem.

Problem: Let AX = B have no solution, that is, B $\notin$ Colsp(A). Then, we pose the following question:

Which vector X will minimize the norm $\| AX-B\|^2$?

Such a vector, if it exists, is called a least square solution of the system AX = B.

Solition: We will use Lemma 4.3 to find the least square solution. To do that, we consider V = $M_{mx1}(\mathbf{R})$ with the usual inner product $<\mathbf{u},\mathbf{v}>= \mathbf{u}^T\mathbf{v}$ for all **u** and **v** $\in$ V .

Putting colsp (A) = W and taking into account that AX $\in$W for all X $\in M_{nx1}(\mathbf{R})$, we obtain, by Lemma 4.3, that $\| AX-B\|^2$ is smallest for such an X = $\tilde{X}$ that A$\tilde{X}$ = P(B), where P: V $\to$W is the orthogonal projection on to W, (since $\| AX-B\|^2 = \| B - AX\|^2 \geq \| B- P(B)\|^2$ and the equality happens if and only if AX = P(B)).

We now find $\tilde{X}$ such that A$\tilde{X}$ = P(B). To do so, we write

$$A\tilde{X} - B = P(B) - B \in W^{\perp}.$$

This is equivalent to

$A\tilde{X} - B \perp U$ for all U $\in$ W = Colsp (A)

$\Leftrightarrow A\tilde{X} - B \perp C_i\ \forall$ i = 1,2... n (where C_i is the i^{th} column of A)

$\Leftrightarrow < A\tilde{X} - B, C_i > = 0\ \forall$ i = 1,2... n.

$\Leftrightarrow C_i^T (A\tilde{X} - B) = 0\ \forall$ i = 1,2... n.

$\Leftrightarrow A^T (A\tilde{X} - B) = 0 \Leftrightarrow A^TA\tilde{X} - A^TB = 0$

$$\Leftrightarrow \qquad A^TA\tilde{X} = A^TB \qquad (4.1)$$

Therefore, $\tilde{X}$ is a solution of the linear system (4.1).

Example: Consider the system $\begin{pmatrix} 1 & 1 & 2 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$

Since r(A) = 2 < r ($\tilde{A}$) = 3, this system has no solution. We now find the vector $\tilde{X}$ such $\tilde{X}$ minimizes the norm $||AX - B||^2$ where A = $\begin{pmatrix} 1 & 1 & 2 \\ 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$; B = $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

To do so, we have to solve the system

$A^TA\tilde{X} = A^TB$

$$\Leftrightarrow \begin{pmatrix} 2 & 0 & 2 \\ 0 & 3 & 3 \\ 2 & 3 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}$$

$$\Leftrightarrow \begin{cases} x_1 + x_3 = 1 \\ x_2 + x_3 = \dfrac{2}{3} \end{cases} \Leftrightarrow \begin{cases} x_1 = 1 - x_3 \\ x_2 = \dfrac{2}{3} - x_3 \\ x_3 \text{ is arbitrary} \end{cases}$$

We thus obtain $\tilde{X} = \begin{pmatrix} 1-t \\ \frac{2}{3} - t \\ t \end{pmatrix}$ $\forall$ t $\in$ **R**.

V. Orthogonal matrices and orthogonal transformation

5.1. Definition. An nxn matrix is said to be orthogonal if $A^TA = I = AA^T$ (i.e., A is nonsingular and $A^{-1} = A^T$)

Examples: 1) A = $\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$

2) A = $\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{pmatrix}$

5.2. Proposition. Let V be an Euclidean Space. Then, the change –of–basis matrix from an ON – basis to another ON – basis of V is an orthogonal matrix.

PROOF. Let S = $\{e_1, e_2, \dots, e_n\}$ and U=$\{u_1, u_2,\dots,u_n\}$ be two ON – bases of V and A be the change–of–basis matrix from S to U, say A = (a_{ij}). Putting $A^T = (b_{ij})$; $A^TA = (c_{ij})$ where $b_{ij} = a_{ji}$; compute

$$c_{ij} = \sum_{k=1}^{n} b_{ik}a_{kJ} = \sum_{k=1}^{n} a_{ki}a_{kJ} = (a_{1i}\ a_{2i}\ \dots\ a_{ni})\begin{pmatrix} a_{1J} \\ a_{2J} \\ \vdots \\ a_{nJ} \end{pmatrix}$$

$$= \left\langle \sum_{k=1}^{n} a_{ki}e_k, \sum_{l=1}^{n} a_{lJ}e_l \right\rangle = \langle u_i, u_J \rangle = \begin{cases} 1 \text{ if } i=j \\ 0 \text{ if } i \neq j \end{cases}$$

Therefore, $A^TA = I$. This means that A is orthogonal.

Example: Let V = $\mathbf{R}^3$ with the usual scalar product < .,. >

S = {(1,0,0), (0,1,0),(0,0,1)} be the usual basis of $\mathbf{R}^3$;

U = $\left\{\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right); \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right); \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)\right\}$ be another ON – basis of $\mathbf{R}^3$.

Then, the change – of – basis matrix for S to U is

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$ Clearly, A is an orthogonal matrix.

5.3. Definition. Let (V, < .,. >) be an IPS. Then, the linear transformation f: V $\to$ V is said to be orthogonal if < f(x),f(y) > = <x,y> for all x,y $\in$ V.

The following theorem provides another criterion for the orthogonality of a linear transformation in Euclidean spaces.

5.4. Theorem: Let V be an Euclidean space, and f: V $\to$ V be a linear transformation. Then, the following assertions are equivalent.

i) f is orthogonal.

ii) For any ON – basis S of V, the matrix representation $[f]_S$ of f corresponding to S is orthogonal.

iii) $||f(x)|| = ||x||$ for all $x \in V$.

PROOF: Let S be an ON – basis of V. Then, taking the coordinates by this basis, we obtain

$<f(x), f(y)> = [f(x)]_S^T [f(y)]_S = ([f]_S[x]_S)^T [f(x)]_S[f(y)]_S$

$= [x]_S^T [f]_S^T [f]_S[y]_S$

Therefore, for an ON – basis S = $\{u_1,u_2,...,u_n\}$ of V, we have that:

$<f(x), f(y)> = <x,y> \ \forall\ x,y \in S$

$\Leftrightarrow\ [x]_S^T[f]_S^T[f]_S[y]_S = [x]_S^T[y]_S \ \forall\ x,y \in S.$

$\Leftrightarrow [u_i]_S^T[f]_s^T[f]_s[u_j] = [u_i]_S^T[u_j]_S = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases} \ \forall\ i,j \in \{1,2...n\}$

$\Leftrightarrow [f]_S^T[f]_S = I \Leftrightarrow [f]_S$ is an orthogonal matrix.

We thus obtain the equivalence (i) $\Leftrightarrow$ (ii). We now prove the equivalence (i) $\Leftrightarrow$ (iii).

(i) $\Rightarrow$ (iii): Since $< f(x), f(y) >= <x,y>$ holds true for all $x,y \in V$, simply taking x = y, we obtain $||f(x)||^2 = ||x||^2 \Rightarrow ||f(x)|| = ||x|| \ \forall\ x \in V$.

(iii) $\Rightarrow$ (i): We have that $||f(x + y)||^2 = ||x + y||^2$ for all $x,y\in V$. Therefore,

$$\langle f(x+y), f(x+y) \rangle = \langle x+y, x+y \rangle.$$

$\Rightarrow \langle f(x), f(x)\rangle + 2\langle f(x), f(y)\rangle + \langle f(y), f(y)\rangle = \langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle$

$\Rightarrow \langle f(x), f(y)\rangle = \langle x,y\rangle \ \ \forall x,y \in V.$

5.5. Definition: Let A be real nxn matrix. Then A is said to be *orthogonally diagonalizable* if there is an orthogonal matrix P such that P^TAP is a diagonal matrix.

The process of finding such an orthogonal matrix P is called the *orthogonal diagonalization* of A.

Remark: If A is orthogonally diagonalizable, then, by definition, $\exists$ P – orthogonal such that

$P^TAP = D$ – diagonal $\Rightarrow A = PDP^T \Rightarrow A^T = PDP^T = A$.

Therefore, A is symmetric.

The converse is also true as we accept the following theorem.

5.6 Theorem: Let A be a real, nxn matrix. Then, A is orthogonally diagonalizable if and only if A is symmetric.

The following lemma shows an important property of symmetric real matrices.

5.7. Lemma: Let A be a real symmetric matrix; and λ_1; λ_2 be two distinct eigenvalues of A; and X_1, X_2 be eigenvectors corresponding to λ_1, λ_2, respectively. Then, $X_1 \perp X_2$ with respect to the usual inner product in M_{nx1} (**R**) (that is, $\langle X_1, X_2\rangle = X_1^T X_2 = 0$)

PROOF: In fact, $\langle \lambda_1 X_1, X_2\rangle = \lambda_1 X_1^t X_2 = (AX_1)^T X_2$

$= X_1^T A^T X_2 = X_1^T A X_2 = X_1^T \lambda_2 X_2 = \lambda_2 \langle X_1, X_2\rangle$

Therefore, $(\lambda_1 - \lambda_2)\langle X_1, X_2\rangle = 0$. Since $\lambda_1 \neq \lambda_2$, this implies that $\langle X_1, X_2\rangle = 0$.

Next, we have the algorithm of orthogonal diagonalization of a symmetric real matrix A.

5.8. Algorithm of orthogonal diagonalization of symmetric nxn matrix A:

Step 1. Find the eigenvalues of A by solving the characteristic equation

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

Step 2. Find all the linearly independent eigenvectors of A, say , $X_1, X_2, \ldots X_n$.

Step 3. Using Gram – Schmidt process to obtain the ON – basis $\{Y_1, Y_2, \ldots Y_n\}$ from $\{X_1, X_2, \ldots X_n\}$; (**Note:** This Gram-Schmidt process can be conducted in each eigenspace since the two eigenvectors in distinct eigenspaces are already orthogonal due to Lemma 5.7).

Step 4. Set $P = [Y_1\ Y_2 \ldots Y_n]$. We obtain immediately that

$$P^T AP = \begin{bmatrix} \lambda_1 & 0\ldots & 0 \\ 0 & \lambda_2\ldots & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0\ldots & \lambda_n \end{bmatrix}$$

where λ_i is the eigenvalue corresponding to the eigenvector Y_i , i = 1,2..., n, respectively.

Example: Let A = $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

To diagonalize orthogonally A, we follow the algorithm 5.8.

Firstly, we compute the eigenvalues of A by solving: $|A - \lambda I| = 0$

$$\Leftrightarrow \begin{vmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{vmatrix} = 0 \Leftrightarrow (\lambda+1)^2 (\lambda - 2) = 0 \Leftrightarrow \left[\begin{array}{l} \lambda_1 = \lambda_2 = -1 \\ \lambda_3 = 2 \end{array}\right.$$

For $\lambda_1 = \lambda_2 = -1$, we compute eigenvectors X by solving.

$(A + E)X = 0 \Leftrightarrow x_1 + x_2 + x_3 = 0$ for $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

Then, there are two linearly independent eigenvectors corresponding to $\lambda = -1$, that are

$X_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}; X_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$. In other worlds, the eigenspace $E_{(-1)}$ = Span $\left\{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\right\}$.

By Gram – Schmidt process, we have $u_1 = v_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$;

$u_2 = v_2 - \dfrac{(v_2, u_1)}{< u_1, u_1 >}.u_1 = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix}$. Then, we put

$Y_1 = \dfrac{U_1}{\|U_1\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$; $Y_2 = \dfrac{U_2}{\|U_2\|} = \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}$ to obtain two orthonormal eigenvectors Y_1, Y_2 corresponding to $\lambda = -1$.

For $\lambda_3 = 2$, We solve $(A - 2E)X = 0$

$$\Leftrightarrow \begin{cases} -2x_1 + x_2 + x_3 = 0 \\ x_1 - 2x_2 + x_3 = 0 \\ x_1 + x_2 - 2x_3 = 0 \end{cases} \Leftrightarrow \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Therefore, there is only linearly independent eigenvector corresponding to $\lambda = 2$. The eigenspace is $E_2 = \text{Span}\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\}$. To implement the Gram-Schmidt process, we just have to put $Y_3 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}$

We now obtain the ON – basis $\{Y_1, Y_2, Y_3\}$ of $M_{3x1}(\mathbf{R})$ which contains the linearly independent eigenvectors of A. Then, we put

$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}$ to obtain $P^TAP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ finishing the orthogonal diagonalization of A.

IV. Quadratic forms

6.1 Definition: Consider the space $\mathbf{R}^n$. A quadratic form q on $\mathbf{R}^n$ is a mapping

q: $\mathbf{R}^n \to \mathbf{R}$ defined by

$$q(x_1, x_2,... x_n) = \sum_{1\le i\le j\le n} c_{ij}x_i x_j \quad \text{for all } (x_1, x_2,... x_n) \in \mathbf{R}^n \qquad (6.1)$$

where the constants $c_{ij} \in \mathbf{R}$ are given, $1 \le i \le j \le n$.

Example: Let $q : \mathbf{R}^3 \to \mathbf{R}$ be defined by

$$q(x_1, x_2, x_3) = x_1^2 - 2x_1x_2 - x_1x_3 + x_2^2 - 2x_2x_3 - x_3^2 \ \forall \ (x_1, x_2, x_3) \in \mathbf{R}^3.$$

Then q is a quadratic form on $\mathbf{R}^3$.

6.2. Definition: The quadratic form q on $\mathbf{R}^n$ is said to be in canonical form (or in diagonal form) if

$$q(x_1, x_2, ..., x_n) = c_{11}x_1^2 + c_{22}x_2^2 + ... + c_{nn}x_n^2 \text{ for all } (x_1,x_2,...x_n) \in \mathbf{R}^n.$$

(That is, q has no cross product terms x_ix_j with $i\neq j$).

We will show that, every quadratic form q can be transformed to the canonical form by choosing a relevant coordinate system.

In general, the quadratic form (6.1) can be expressed uniquely in the matrix form as

$$q(x)=q(x_1, x_2, ..., x_n)=[x]^TA[x] \text{ for } x=(x_1, x_2, ..., x_n) \in \mathbf{R}^n$$

where $[x] = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is the coordinate vector of X with respect to the usual basis of $\mathbf{R}^n$ and

$A = [a_{ij}]$ is a symmetric matrix with $a_{ij} = a_{ji} = c_{ij}/2$ for $i \neq j$, and $a_{ii} = c_{ii}$ for $i = 1,2...n$.

This fact can be easily seen by directly computing the product of matrices of the form:

$$q(x_1,x_2..., x_n) = (x_1 \ x_2 \ ... \ x_n)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = [x]^T A[x]$$

The above symmetric matrix A is called the matrix representation of the quadratic form q with respect to usual basis of $\mathbf{R}^n$.

6.3. Change of coordinates:

Let E be the usual (canonical) basis of $\mathbf{R}^n$. For $x = (x_1,...,x_n) \in \mathbf{R}^n$, as above, we denote by [x] the coordinate vector of x with respect to E. Let $q(x) = [x]^T A[x]$ be a quadratic form on $\mathbf{R}^n$ with the matrix representation A. Now, consider a new basis S of $\mathbf{R}^n$, and let P be the change-of–basis matrix from E to S. Then, we have $[x] = [x]_E = P[x]_S$.

Therefore, $q = [x]^T A[x] = [x]_S^T P^T AP[x]_S$.

Putting $B = P^T AP$; $Y = [x]_S$, we obtain that, in the new coordinate system with basis S, q has the form:

$$q = Y^T BY, \text{ where } B = P^T AP \text{ and } Y = [x]_S.$$

Example: Consider the quadratic form

q: $\mathbf{R}^3 \to \mathbf{R}$ defined by $q(x_1,x_2,x_3) = 2x_1x_2 + 2x_2x_3 + 2x_3x_1$, or in matrix representation by

$$q = (x_1\ x_2\ x_3)\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = X^T AX \text{ for } X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$ -the coordinate vector of (x_1,x_2,x_3) with respect to the usual basis of $\mathbf{R}^3$. Let S = {(1,0,0); (1,1,0); (1,1,1)} be another basis of $\mathbf{R}^3$. Then, we can compute the change-of –basis matrix P from the usual basis to the basis S as

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus, the relation between the old and new coordinate vector $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ and $Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = [(x_1, x_2, x_3)]_S$ is $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$. Therefore, changing to the new coordinates, we obtain

$$q = X^T AX = Y^T P^T APY$$

$$= (y_1\ y_2\ y_3)\begin{pmatrix}1&0&0\\1&1&0\\1&1&1\end{pmatrix}\begin{pmatrix}0&1&1\\1&0&1\\1&1&0\end{pmatrix}\begin{pmatrix}1&1&1\\0&1&1\\0&0&1\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}$$

$$= (y_1\ y_2\ y_3)\begin{pmatrix}1&1&2\\1&2&4\\2&4&6\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}$$

$$= y_1^2 + 2y_2^2 + 6y_3^2 + 2y_1y_2 + 4y_2y_3 + 8y_3y_1.$$

6.4. Transformation of quadratic forms to principal axes (or canonical forms):

Consider a quadratic form q in matrix representation: $q = X^TAX$, where X = [x] is the coordinate vector of x with respect to the usual basis of $\mathbf{R}^n$, and A is the matrix representation of q.

Since A is symmetric, A is orthogonally diagonalizable. This means that there exists an orthogonal matrix P such that P^TAP is diagonal, i.e.,

$$P^TAP = \begin{pmatrix}\lambda_1&0&\cdots&0\\0&\lambda_2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&\lambda_n\end{pmatrix} \text{- a diagonal matrix.}$$

We then change the coordinate system by putting X = PY. This means that we choose a new basis S such that the change–of–basis matrix from the usual basis to the basis S is the matrix P. Hence, in the new coordinate system, q has the form:

$$q = X^TAX = Y^TP^TAPY = Y^T\begin{pmatrix}\lambda_1&0&\cdots&0\\0&\lambda_2&\cdots&0\\\vdots&&&\\0&0&\cdots&\lambda_n\end{pmatrix}Y$$

$$= \lambda_1y_1^2 + \lambda_2y_2^2 + \ldots + \lambda_ny_n^2 \text{ for } Y = \begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}.$$

Therefore, q has a canonical form; and the vectors in the basis S are called the *principal axes* for q.

The above process is called the **transformation of q to the principal axes** (or the diagonalization of q). More concretely, we have the following algorithm to diagonalize a quadratic form q.

6.5. Algorithm of diagonalization of a quadratic form $q = X^TAX$:

Step 1: Orthogonally diagonalize A, that is, find an orthogonal matrix P so that P^TAP is diagonal. This can always be done because A is a symmetric matrix.

Step 2. Change the coordinates by putting X = PY (here P acts as the change–of–basis matrix). Then, in the new coordinate system, q has the diagonal form:

$$q = Y^TP^TAPY = (y_1\ y_2 \ldots y_n)\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$$= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \ldots + \lambda_n y_n^2$$ finishing the process.

Note that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all eigenvalues of A

Example: Consider $q = 2x_1x_2 + 2x_2x_3 + 2x_3x_1 = (x_1\ x_2\ x_3)\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$

To diagonalize q, we first orthogonally diagonalize A. This is already done in Example after algorithm 5.8, by this example, we obtain

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \text{ for which } P^T AP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

We next change the coordinates by simply putting

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = P \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Then, in the new coordinate system, q has the form:

$$q = (y_1\ y_2\ y_3)\, P^T A P \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = (y_1\ y_2\ y_3) \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = -y_1^2 - y_2^2 + 2y_3^2$$

6.6. Law of inertia: Let q be a quadratic form on $\mathbf{R}^n$. Then, there is a basis of $\mathbf{R}^n$ (a coordinate system in $\mathbf{R}^n$) in which q is represented by a diagonal matrix, every other diagonal representation of q has the same number *p* of positive entries and the same number *m* of negative entries. The difference s= *p* - *m* is called the signature of q.

Example: In the above example $q = 2x_1x_2 + 2x_2x_3 + 2x_3x_1$ we have that *p* = 1; *m*=2. Therefore, the signature of q is s = 1-2 = -1. To illustrate the Law of inertia, let us use another way to transform q to canonical form as follow.

Firstly, putting $\begin{cases} x_1 = y_1' - y_2' \\ x_2 = y_2' + y_2' \\ x_3 = y_3' \end{cases}$ we have that

$$q = 2{y_1'}^2 - 2{y_2'}^2 + 4y_1'y_3'$$

$$= 2\left({y_1'}^2 + 2y_1'y_3' + {y_3'}^2\right) - 2{y_2'}^2 - 2y_2' - 2{y_3'}^2$$

$$= 2\,(y_1' + y_3')^2 - 2{y_2'}^2 - 2{y_3'}^2.$$

Putting $\begin{cases} y_1 = y_1' + y_3' \\ y_2 = y_2' \\ y_3 = y_3' \end{cases}$ we obtain $q = 2y_1^2 - 2y_2^2 - 2y_3^2$.

Then, we have the same *p* = 1; *m* = 2; s = 1-2 = -1 as above.

6.7 Definition: A quadratic form q on $\mathbf{R}^n$ is said to be positive definite if q(v) > 0 for every nonzero vector $v \in \mathbf{R}^n$.

By the diagonalization of a quadratic form, we obtain immediately the following theorem.

6.8. Theorem: A quadratic form is positive definite if and only if all the eigenvalues of its matrix representation are positive.

VII. Quadric lines and surfaces

7.1. Quadric lines: Consider the coordinate plane *xOy*.

A quadric line is a line on the plane *xOy* which is described by the equation

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + b_1x + b_2y + c = 0,$$

or in matrix form:

$$(x\ y)\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 & b_2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + c = 0$$

where the 2x2 symmetric matrix $A = (a_{ij}) \neq 0$.

We can see that the equation of a quadric line is the sum of a quadratic form and a linear form. Also, as known in the above section, the quadratic form is always transformed to the principal axes. Therefore, we will see that we can also transform the quadric lines to the principal axes.

7.2. Transformation of quadric lines to the principal axes:

Consider the quadric line described by the equation

$$(x\ \ y)\ A\begin{pmatrix} x \\ y \end{pmatrix} + B\begin{pmatrix} x \\ y \end{pmatrix} + c = 0 \tag{7.1}$$

where A is a 2x2 symmetric nonzero matrix, $B=(b_1\ b_2)$ is a row matrix, and c is constant. Basing on the algorithm of diagonalization of a quadratic form, we then have the following algorithm of transformation a quadric line to the principal axes (or the canonical form).

Algorithm of transformation the quadric line (7.1) to principal axes.

Step 1. Orthogonally diagonalize A to find an orthogonal matrix P such that

$$P^T A P = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Step 2. Change the coordinates by putting $\begin{pmatrix} x \\ y \end{pmatrix}$ =P$\begin{pmatrix} x' \\ y' \end{pmatrix}$.

Then, in the new coordinate system, the quadric line has the form

$$\lambda_1 x'^2 + \lambda_2 y'^2 + b_1' x' + b_2' y' + c = 0$$

where $(b_1' \ b_2') = (b_1 \ \ b_2)P$.

Step 3. Eliminate the first order terms if possible.

Example: Consider the quadric line described by the equation

$$x^2+2xy+y^2+8x+y=0 \quad (7.2)$$

To perform the above algorithm, we write (7.2) in a matrix form

$$(x \quad y)\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + (8 \quad 1)\begin{pmatrix} x \\ y \end{pmatrix} = 0.$$

Firstly, we diagonalize A = $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ orthogonally starting by computing the eigenvalue of A from the characteristic equation

$$\begin{vmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = 0 \Leftrightarrow \left[\begin{array}{l} \lambda_1 = 0 \\ \lambda_2 = 2 \end{array}\right.$$

For $\lambda_1 = 0$, there is one linearly independent eigenvector $u_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

For $\lambda_2 = 2$, there is also only one linearly independent eigenvector $u_2=\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

The Gram-Schmidt process is very simple in this case. We just have to put

$$e_1 = \frac{u_1}{\| u_1 \|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad e_2 = \frac{u_2}{\| u_2 \|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Then, setting P = $\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$ we have that $P^T AP = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$

Next, we change the coordinates by putting $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix}$

Therefore, the equation in new coordinate system is

$$(x'\ y')\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} + (8\quad 1)\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} = 0$$

$$\Leftrightarrow 2y'^2 + \frac{7}{\sqrt{2}}x' + \frac{9}{\sqrt{2}}y' = 0$$

$$\Leftrightarrow 2\left(y' + \frac{9}{4\sqrt{2}}\right)^2 + \frac{7}{\sqrt{2}}\left(x' - \frac{\sqrt{2}.81}{112}\right) = 0$$

(We write this way to eliminate the first order term $\frac{9}{\sqrt{2}}y'$)

Now, we continue to change coordinates by putting

$$\begin{cases} X = x' - \frac{81\sqrt{2}}{112} \\ Y = y' + \frac{9}{4\sqrt{2}} \end{cases}$$

In fact, this is the translation of the coordinate system to the new origin

$$I\left(\frac{81\sqrt{2}}{112}, \frac{-9}{4\sqrt{2}}\right)$$

Then, we obtain the equation in principal axes:

$$2Y^2 + \frac{7}{\sqrt{2}}X = 0.$$

Therefore, this is a parabola.

7.3. Quadric surfaces:

Consider now the coordinate space *Oxyz.*

A quadric surface is a surface in space *Oxyz* which is described by the equation

$$a_{11}x^2+ a_{22}y^2+a_{33}z^2+2a_{12}xy+2a_{23}yz + 2a_{13}zx+b_1x+b_2y+b_3z + c = 0$$

or, in matrix form

$$(x \quad y \quad z)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + (b_1 \quad b_2 \quad b_3)\begin{pmatrix} x \\ y \\ z \end{pmatrix} + c = 0,$$

where A =(a_{ij}) is a 3x3 symmetric matrix; A ≠ 0.

Similarly to the case of quadric lines, we can transform a quadric surface to the principal axes by the following algorithm.

7.4. Algorithm of transformation of a quadric surface to the principal axes (or to the canonical form):

Step 1. Write the equation in the matrix form as above

$$(x \quad y \quad z)A\begin{pmatrix} x \\ y \\ z \end{pmatrix} + B\begin{pmatrix} x \\ y \\ z \end{pmatrix} + c = 0.$$

Then, orthogonally diagonalize A to obtain an orthogonal matrix P such that

$$P^TAP = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

Step 2. Change the coordinates by putting $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = P\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$

Then, the equation in the new coordinate system is

$$\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 + b_1' x' + b_2' y' + b_3' z' + c = 0$$

where $\begin{pmatrix} b_1' & b_2' & b_3' \end{pmatrix} = \begin{pmatrix} b_1 & b_2 & b_3 \end{pmatrix} P$

Step 3. Eliminate the first order terms if possible.

Example: Consider the quadric surface described by the equation:

$$2xy + 2xz + 2yz - 6x - 6y - 4z = 0$$

$$\Leftrightarrow \begin{pmatrix} x & y & z \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} -6 - 6 - 4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = 0.$$

The orthogonal diagonalization of A = $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ was already done in the example after algorithm 5.8 by that we obtain

$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \text{ for which } P^TAP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

We then put $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = P\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$ to obtain equation in new coordinate system as

$$-x'^2 - y'^2 + 2z'^2 + \begin{pmatrix} -6 & -6 & -4 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = 0$$

$$\Leftrightarrow x'^2 - y'^2 + 2z'^2 + \frac{4}{\sqrt{6}}y' - \frac{16}{\sqrt{3}}z' = 0$$

$$\Leftrightarrow -x^2 - \left(y' - \frac{2}{\sqrt{6}}\right)^2 + 2\left(z' - \frac{4}{\sqrt{3}}\right)^2 - 10 = 0$$

Now, we put (in fact, this is a translation)

$$\begin{cases} X = x' \\ Y = y' - \frac{2}{\sqrt{6}} \\ Z = z' - \frac{4}{\sqrt{3}} \end{cases}$$ (the new origin is $I\left(0, \frac{2}{\sqrt{6}}; \frac{4}{\sqrt{3}}\right)$)

to obtain the equation of the surface in principal axes

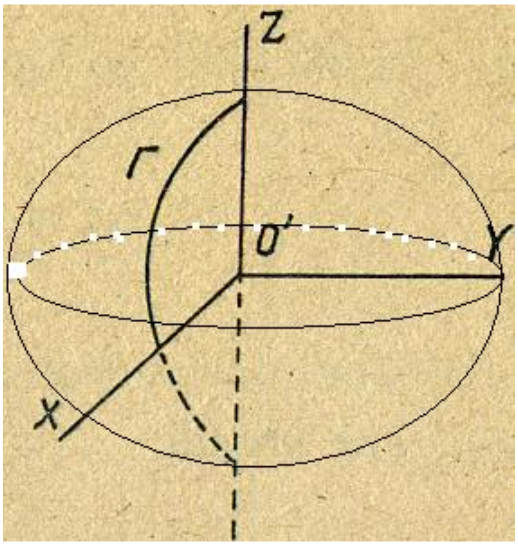
$$X^2 + Y^2 - 2Z^2 = -10.$$

We can conclude that, this surface is a two-fold hyperboloid (see the subsection 7.6).

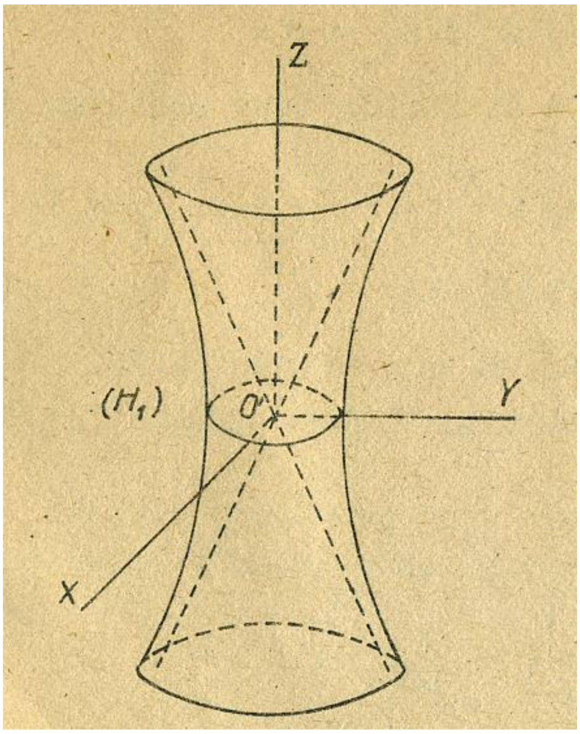
7.6. Basic quadric surfaces in principal axes:

1) Ellipsoid: $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$

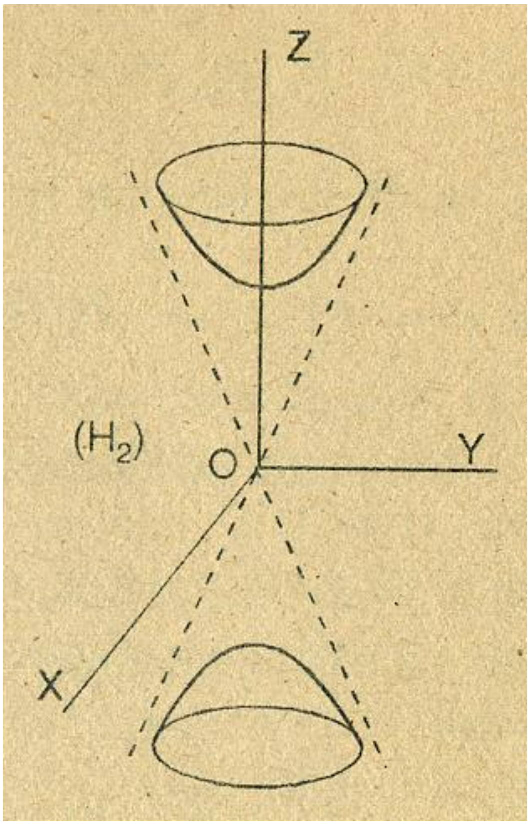
(If a = b = c, this is a sphere)



2) 1 – fold hyperboloid $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$

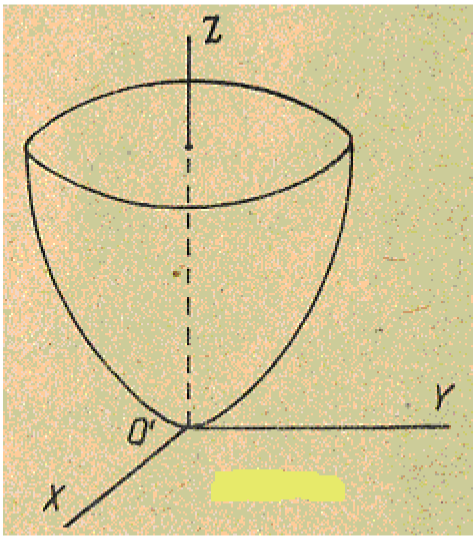


3) 2 – fold hyperboloid $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=-1$

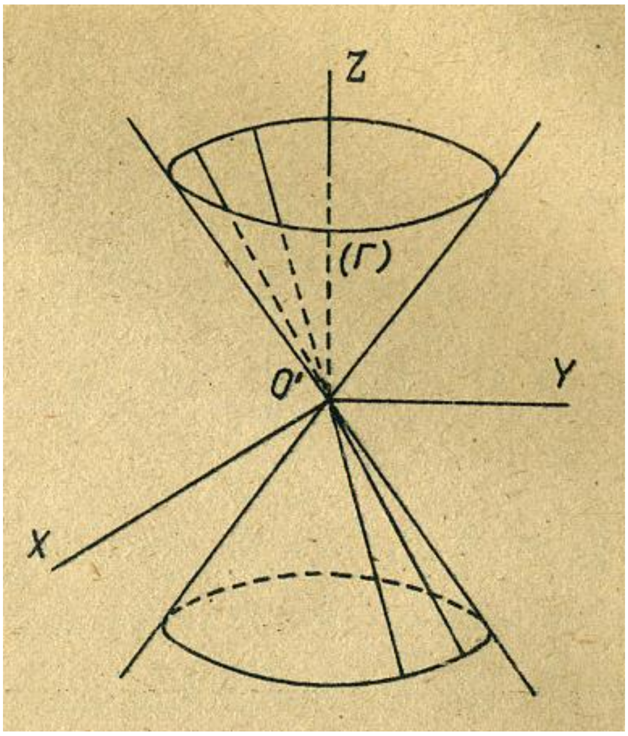


4) Elliptic paraboloid $\frac{x^2}{a^2} + \frac{y^2}{b^2} - z = 0$

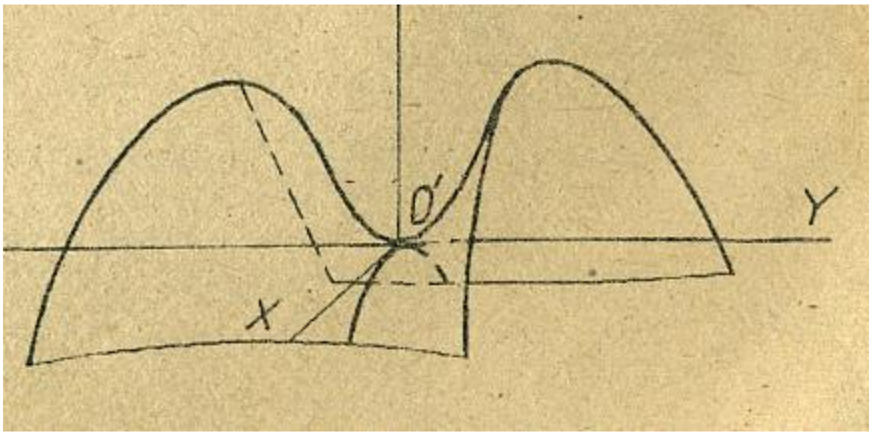
(If a = b this a paraboloid of revolution)



5) Cone: $\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 0$



6) Hyperbolic – paraboloid (or Saddle surface) $\frac{x^2}{a^2} - \frac{y^2}{b^2} - z = 0$



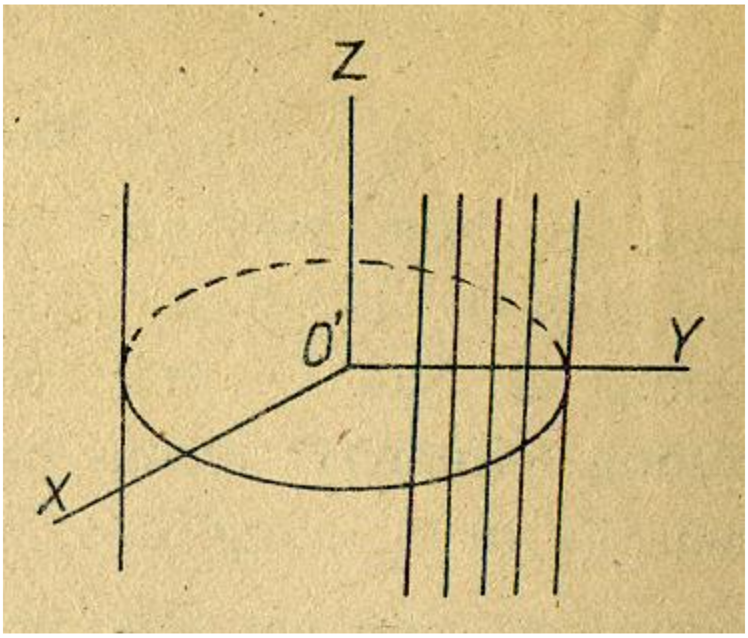
7) Cylinder: A cylinder has one of the following form of equation:

f(x, y) = 0; or f(x, z) = 0 or f(y, z) = 0.

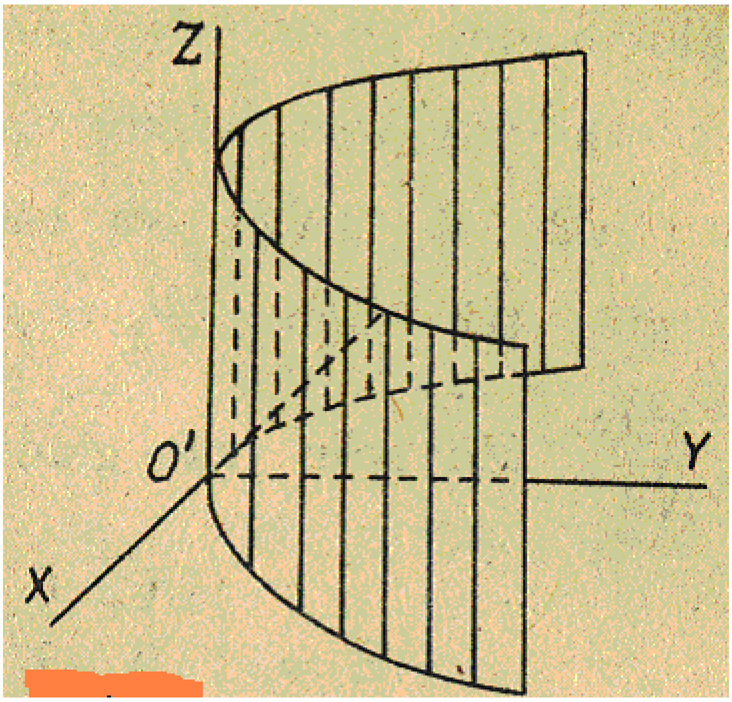
Since the roles of x, y, z are equal, we consider only the case of equation f(x,y) = 0. This cylinder is consisted of generating lines paralleling to z–axis and leaning on a directrix which lies on *xOy* – plane and has the equation f(x,y) = 0 (on this plane).

Example of quadric cylinders:

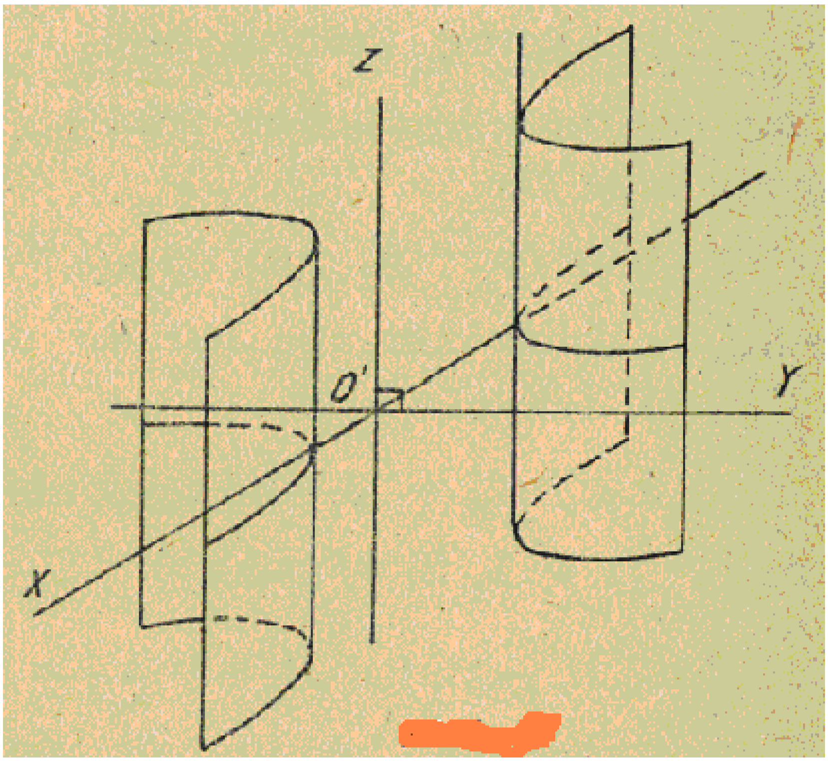
1) $x^2 + y^2 = 1$



2) $x^2 - y = 0$



3) $x^2 - y^2 = 1$



References

1. E.H. Connell, *Elements of abstract and linear algebra*, 2001, [http://www.math.miami.edu/_ec/book/]
2. David C. Lay, *Linear Algebra and its Applications*, 3rd Edition, Addison-Wesley, 2002.
3. S. Lipschutz, *Schaum's Outline of Theory and Problems of Linear Algebra*, (Schaum,1991) McGraw-hill, New York, 1991.
4. Gilbert Strang, *Introduction to Linear Algebra*, Wellesley-Cambridge Press, 1998.