Linear Algebra Thm Archive - for Mid-Term Exam

by Gyeonggi Science High School for the Gifted 'Linear Algebra' Participants
Main Author : 14129 황동욱
Revised by : 14121 하석민
LaTeX Technician : 14041 박승원

Last Compilation Time : Tuesday 24th January, 2017 16:51

Preface and License

This work is a theorem archive for 'Linear Algebra' subject in Gyeonggi Science High School for the Gifted. It is initiated and mainly written by 황동욱, and being revised by 하석민, 박승원.

Each 'Theorem' is identically numbered as textbook. (Except Chapter 3.5) On the other hand, 'Extra Theorem' is things that aren't discussed or proved in textbook.

Anyway, good luck on your mid-term exam on Friday!

Theorems in this archive can have some errors. Please come to us if you find some of them, then we will revise them.

Contents

List of Theorems

List of Extra Theorems

Chapter 1

Vectors

1.1 Terminology relating to vectors

A vector can be represented either in geometric way or algebraic way. In geometric definition of vectors, a vector is a **directed line segment**. A vector from point A (**initial point**, or **tail**) to point B (**terminal point**, or **head**) is denoted as $\overrightarrow{AB}$. Vectors with their tails in the origin is called **position vectors**, and they are at **standard position**.
In algebraic view of vectors, a vector is an **ordered pair** of **components**. We denote the set of all vectors containing n components in $\mathbb{R}$ as $\mathbb{R}^n$. Similarly, set of all vectors containing n integer components is $\mathbb{Z}^n$.
A vector is written in form of **column vectors** and **row vectors**. We use square brackets for denoting vectors, such as

$$\mathbf{v} = \begin{bmatrix} 4 \\ 1 \\ 6 \end{bmatrix}, \begin{bmatrix} 4, & 1, & 6 \end{bmatrix}$$

A **zero vector** is a vector which components are all zero. A zero vector is denoted as $\mathbf{0}$.
Two vectors are equal if and only if all the components of two vectors are equal. (Of course, the number of components should be same.)
Standard unit Vectors have components which one of them is 1 and rest of them are all 0. Unit vector which has 1 in ith component is denoted as $\mathbf{e}_i$, and

$$\mathbf{e}_i = \begin{bmatrix} 0, & 0, & \cdots & 1, & \cdots & 0 \end{bmatrix}$$

* *Note* We should always denote vector either with arrows ($\vec{v}$), or with boldface letters ($\mathbf{v}$). Scalar denotations (v) are not allowed.
A set of vectors with n components taken from finite set of integers $\mathbb{Z}_m = \{0, 1, 2, \cdots, m-1\}$ is denoted as $\mathbb{Z}_m^n$. $\mathbb{Z}_m^n$ is closed with respect to operations of vector addition and scalar multiplication (which is defined later). We can perform those operations in same way, but with modulo operations. Vectors in $\mathbb{Z}_2^n$ (all components are 0 or 1) are called **binary vectors**.

1.2 Basic Operations of Vectors

Definition. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$be $\mathbf{u} = \begin{bmatrix} u_1, & u_2, & \cdots & u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1, & v_2, & \cdots & v_n \end{bmatrix}$. Then their **sum** $\mathbf{u}+\mathbf{v}$ is defined as

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1, & u_2 + v_2, & \cdots & u_n + v_n \end{bmatrix}$$

Definition. Let $\mathbf{v} \in \mathbb{R}^n$be $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$, and let c be a real number. Then their **scalar**

multiple $c\mathbf{v}$ is denifed as

$$c\mathbf{v} = \begin{bmatrix} cv_1, & cv_2, & \cdots & cv_n \end{bmatrix}$$

Definition. A **negative** of $\mathbf{v}$ is defined as $-\mathbf{v} = (-1)\mathbf{v}$.

Definition. The **difference** of vectors is defined as $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$.

Definition. Two vectors are **parallel** if and only if one vector is a scalar multiple of another. (Thus, zero vector is parallel with all vectors.)

Theorem 1.1 : Algebraic Properties of Vectors : Basic Vector Operations
Let $\mathbf{u}$,$\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$and let c and d be scalars.

a. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (Commutative Property of Vector Addition)

b. $(\mathbf{u} + \mathbf{v}) + \mathbf{u} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ (Associative Property of Vector Addition)

c. $\mathbf{u} + \mathbf{0} = \mathbf{u}$

d. $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$

e. $c(\mathbf{u}+\mathbf{v}) = c\mathbf{u}+c\mathbf{v}$ (Left-Distributive Property of Scalar Multiplication over Vector Addition)

f. $(c+d)\mathbf{u} = c\mathbf{u}+d\mathbf{u}$ (Right-Distributive Property of Scalar Multiplication over Vector Addition)

g. $c(d\mathbf{u}) = (cd)\mathbf{u}$

h. $1\mathbf{u} = \mathbf{u}$

Proof. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$be $\mathbf{u} = \begin{bmatrix} u_1, & u_2, & \cdots & u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1, & v_2, & \cdots & v_n \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_1, & w_2, & \cdots & w_n \end{bmatrix}$, and let c, d be scalars.

a.

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= \begin{bmatrix} u_1 + v_1, & u_2 + v_2, & \cdots & u_n + v_n \end{bmatrix} \\ &= \begin{bmatrix} v_1 + u_1, & v_2 + u_2, & \cdots & v_n + u_n \end{bmatrix} = \mathbf{v} + \mathbf{u} \end{aligned}$$

b.

$$\begin{aligned} (\mathbf{u} + \mathbf{v}) + \mathbf{w} &= \begin{bmatrix} u_1 + v_1, & u_2 + v_2, & \cdots & u_n + v_n \end{bmatrix} + \begin{bmatrix} w_1, & w_2, & \cdots & w_n \end{bmatrix} \\ &= \begin{bmatrix} (u_1 + v_1) + w_1, & (u_2 + v_2) + w_2, & \cdots & (u_n + v_n) + w_n \end{bmatrix} \\ &= \begin{bmatrix} u_1 + (v_1 + w_1), & u_2 + (v_2 + w_2), & \cdots & u_n + (v_n + w_n) \end{bmatrix} \\ &= \mathbf{u} + (\mathbf{v} + \mathbf{w}) \end{aligned}$$

c.

$$\begin{aligned} \mathbf{u} + \mathbf{0} &= \begin{bmatrix} u_1 + 0, & u_2 + 0, & \cdots & u_n + 0 \end{bmatrix} \\ &= \begin{bmatrix} u_1, & u_2, & \cdots & u_n \end{bmatrix} = \mathbf{u} \end{aligned}$$

d. (Excercise 1.1 24)

$$\begin{aligned}\mathbf{u}+(-\mathbf{u}) &= \begin{bmatrix}u_1+(-u_1), & u_2+(-u_2), & \cdots & u_n+(-u_n)\end{bmatrix} \\ &= \begin{bmatrix}0, & 0, & \cdots & 0\end{bmatrix} = \mathbf{0}\end{aligned}$$

e. (Excercise 1.1 24)

$$\begin{aligned}c(\mathbf{u}+\mathbf{v}) &= c\begin{bmatrix}u_1+v_1, & u_2+v_2, & \cdots & u_n+v_n\end{bmatrix} \\ &= \begin{bmatrix}c(u_1+v_1), & c(u_2+v_2), & \cdots & c(u_n+v_n)\end{bmatrix} \\ &= \begin{bmatrix}cu_1+cv_1, & cu_2+cv_2, & \cdots & cu_n+cv_n\end{bmatrix} \\ &= \begin{bmatrix}cu_1, & cu_2, & \cdots & cu_n\end{bmatrix} + \begin{bmatrix}cv_1, & cv_2, & \cdots & cv_n\end{bmatrix} \\ &= c\mathbf{u}+c\mathbf{v}\end{aligned}$$

f. (Excercise 1.1 24)

$$\begin{aligned}(c+d)\mathbf{u} &= \begin{bmatrix}(c+d)u_1, & (c+d)u_2, & \cdots & (c+d)u_n\end{bmatrix} \\ &= \begin{bmatrix}cu_1+du_1, & cu_2+du_2, & \cdots & cu_n+du_n\end{bmatrix} \\ &= \begin{bmatrix}cu_1, & cu_2, & \cdots & cu_n\end{bmatrix} + \begin{bmatrix}du_1, & du_2, & \cdots & du_n\end{bmatrix} \\ &= c\mathbf{u}+d\mathbf{u}\end{aligned}$$

g. (Excercise 1.1 24)

$$\begin{aligned}c(d\mathbf{u}) &= c\begin{bmatrix}du_1, & du_2, & \cdots & du_n\end{bmatrix} \\ &= \begin{bmatrix}cdu_1, & cdu_2, & \cdots & cdu_n\end{bmatrix} \\ &= cd\begin{bmatrix}u_1, & u_2, & \cdots & u_n\end{bmatrix} \\ &= (cd)\mathbf{u}\end{aligned}$$

□

1.3 Dot Product

Definition. Let $\mathbf{u}$,$\mathbf{v} \in \mathbb{R}^n$be $\mathbf{u} = \begin{bmatrix}u_1, & u_2, & \cdots & u_n\end{bmatrix}, \mathbf{v} = \begin{bmatrix}v_1 & v_2 & \cdots & v_n\end{bmatrix}$. Then the **dot product** $\mathbf{u}\cdot\mathbf{v}$ of $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\mathbf{u}\cdot\mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

Theorem 1.2 : Properties of Dot Product
Let $\mathbf{u}$,$\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$and let c be a scalar.

a. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$ (Commutative Property of Dot Product Operation)

b. $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w}) = \mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$ (Distributive Property of Dot Product Operation over Vector Addition)

c. $(c\mathbf{u})\cdot\mathbf{v} = c(\mathbf{u}\cdot\mathbf{v})$

d. $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Proof. Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$be $\mathbf{u} = \begin{bmatrix} u_1, & u_2, & \cdots & u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1, & v_2, & \cdots & v_n \end{bmatrix}, \mathbf{w} = \begin{bmatrix} w_1, & w_2, & \cdots & w_n \end{bmatrix}$, and let c be a scalar.

a.

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= u_1v_1 + u_2v_2 + \cdots + u_nv_n \\ &= v_1u_1 + v_2u_2 + \cdots + v_nu_n = \mathbf{v} \cdot \mathbf{u} \end{aligned}$$

b.

$$\begin{aligned} \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) &= \begin{bmatrix} u_1, & u_2, & \cdots & u_n \end{bmatrix} \cdot \begin{bmatrix} v_1 + w_1, & v_2 + w_2, & \cdots & v_n + w_n \end{bmatrix} \\ &= u_1(v_1 + w_1) + u_2(v_2 + w_2) + \cdots + u_n(v_n + w_n) \\ &= (u_1v_1 + u_2v_2 + \cdots + u_nv_n) + (u_1w_1 + u_2w_2 + \cdots + u_nw_n) \\ &= \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \end{aligned}$$

c.

$$\begin{aligned} (c\mathbf{u}) \cdot \mathbf{v} &= \begin{bmatrix} cu_1, & cu_2, & \cdots & cu_n \end{bmatrix} \cdot \begin{bmatrix} v_1, & v_2, & \cdots & v_n \end{bmatrix} \\ &= cu_1v_1 + cu_2v_2 + \cdots + cu_nv_n \\ &= c(u_1v_1 + u_2v_2 + \cdots + u_nv_n) = c(\mathbf{u} \cdot \mathbf{v}) \end{aligned}$$

d.

$$\mathbf{u} \cdot \mathbf{u} = u_1^2 + u_2^2 + \cdots + u_n^2 \geq 0$$

and since $u_1^2 + u_2^2 + \cdots + u_n^2 = 0$ if and only if $u_1 = u_2 = \cdots = u_n = 0$, $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

□

Definition. The **length** or **norm** of $\mathbf{v} \in \mathbb{R}^n$is defined as

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

Theorem 1.3 : Properties of Norm
For all vectors $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$and scalar c,

a. $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$

b. $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$

Proof. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$be $\mathbf{u} = \begin{bmatrix} u_1, & u_2, & \cdots & u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1, & v_2, & \cdots & v_n \end{bmatrix}$, and let c be a scalar.

a. Since $\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = 0$ if and only if $v_1 = v_2 = \cdots = v_n = 0$, $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

b.

$$\begin{aligned}\|c\mathbf{v}\| &= \sqrt{(cv_1)^2+(cv_2)^2+\cdots+(cv_n)^2}\\ &= |c|\sqrt{v_1^2+v_2^2+\cdots+v_n^2} = |c|\|\mathbf{v}\|\end{aligned}$$

□

Theorem 1.4 : The Cauchy-Schwarz Inequality
For all vectors $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$, $|\mathbf{u}\cdot\mathbf{v}| \le \|\mathbf{u}\|\|\mathbf{v}\|$.

Proof. (Exercise 1.2 71) For any $t\in\mathbb{R}$, $(u_1t-v_1)^2+(u_2t-v_2)^2+\cdots+(u_nt-v_n)^2 \ge 0$. Therefore, the determinant of the quatratic equation $(u_1^2+u_2^2+\cdots+u_n^2)t^2-2(u_1v_1+u_2v_2+\cdots+u_nv_n)t+(v_1^2+v_2^2+\cdots+v_n^2)=0$ is negative or 0. Then

$$\begin{aligned}D/4 = (u_1v_1+u_2v_2+\cdots+u_nv_n)^2-(u_1^2+u_2^2+\cdots+u_n^2)(v_1^2+v_2^2+\cdots+v_n^2) &\le 0\\ |u_1v_1+u_2v_2+\cdots+u_nv_n| \le \sqrt{u_1^2+u_2^2+\cdots+u_n^2}&\sqrt{v_1^2+v_2^2+\cdots+v_n^2}\\ \therefore |\mathbf{u}\cdot\mathbf{v}| \le \|\mathbf{u}\|\|\mathbf{v}\|&\end{aligned}$$

□

Theorem 1.5 : The Triangle Inequality
For all vectors $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^n$, $\|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\|+\|\mathbf{v}\|$.

Proof.

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})\\ &= \mathbf{u}\cdot\mathbf{u}+2(\mathbf{u}\cdot\mathbf{v})+\mathbf{v}\cdot\mathbf{v}\\ &\le \|\mathbf{u}\|^2+2|\mathbf{u}\cdot\mathbf{v}|+\|\mathbf{v}\|^2\\ &\le \|\mathbf{u}\|^2+2\|\mathbf{u}\|\|\mathbf{v}\|+\|\mathbf{v}\|^2\\ &= (\|\mathbf{u}\|+\|\mathbf{v}\|)^2\end{aligned}$$

□

Definition. The **distance** between $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$is defined as

$$\mathrm{d}(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|$$

Definition. The **angle** between nonzero vectors $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$is defined as

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

* *Note.* Since $|\mathbf{u}\cdot\mathbf{v}| \le \|\mathbf{u}\|\|\mathbf{v}\|$, $-1\le\cos\theta\le 1$, which corresponds with properties of cosine function.
Definition. $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** to each other if and only if $\mathbf{u}\cdot\mathbf{v}=0$. Since $\mathbf{0}\cdot\mathbf{v}=0$ for every vector $\mathbf{v}$ in $\mathbb{R}^n$, $\mathbf{0}$ is orthonogal with every vector.

Theorem 1.6 : Pythagoras' Theorem
For vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2+\|\mathbf{v}\|^2$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

Proof. Since $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2$, $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ if and only if $\mathbf{u}\cdot\mathbf{v} = 0$. $\square$

Definition. For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$and $\mathbf{u} \neq \mathbf{0}$, the **projection of v onto u** is denoted as $\text{proj}_{\mathbf{u}}(\mathbf{v})$ and defined as

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}$$

Derivation. The projection of $\mathbf{v}$ onto $\mathbf{u}$ is scalar multiple of $\mathbf{u}$, and its scale is $\|\mathbf{v}\|\cos\theta$. Therefore,

$$\begin{aligned}
\text{proj}_{\mathbf{u}}(\mathbf{v}) &= \|\mathbf{v}\|\cos\theta\left(\frac{1}{\|\mathbf{u}\|}\right)\mathbf{u} \\
&= \|\mathbf{v}\|\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}\right)\left(\frac{1}{\|\mathbf{u}\|}\right)\mathbf{u} \\
&= \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|^2}\right)\mathbf{u} \\
&= \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}
\end{aligned}$$

1.4 Geometry in Vectors

	Normal Form	**General Form**	**Vector Form**	**Parametric Form**
Lines	$\mathbf{n}\cdot\mathbf{x} = \mathbf{n}\cdot\mathbf{p}$	$ax + by = c$	$\mathbf{x} = \mathbf{p} + t\mathbf{d}$	$\begin{cases} x = p_x + td_x \\ y = p_y + td_y \end{cases}$

$(t \in \mathbb{R})$

Table 1.1: Equations of Lines in $\mathbb{R}^2$.

	Normal Form	**General Form**	**Vector Form**	**Parametric Form**
Lines	$\begin{cases} \mathbf{n}_1\cdot\mathbf{x} = \mathbf{n}_1\cdot\mathbf{p}_1 \\ \mathbf{n}_2\cdot\mathbf{x} = \mathbf{n}_2\cdot\mathbf{p}_2 \end{cases}$	$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \end{cases}$	$\mathbf{x} = \mathbf{p} + t\mathbf{d}$	$\begin{cases} x = p_x + td_x \\ y = p_y + td_y \\ z = p_z + td_z \end{cases}$
Planes	$\mathbf{n}\cdot\mathbf{x} = \mathbf{n}\cdot\mathbf{p}$	$ax + by + cz = d$	$\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$	$\begin{cases} x = p_x + su_x + tv_x \\ y = p_y + su_y + tv_y \\ z = p_z + su_z + tv_z \end{cases}$

$(t, s \in \mathbb{R})$

Table 1.2: Equations of Lines and Planes in $\mathbb{R}^3$.

The distance from the point A to a line l: $\mathbf{x} = \mathbf{p} + t\mathbf{d}$ is represented as

$$\begin{aligned}
\text{d}(A, l) &= \|(\mathbf{a}-\mathbf{p}) - \text{proj}_{\mathbf{d}}(\mathbf{a}-\mathbf{p})\| \\
&= \left\|(\mathbf{a}-\mathbf{p}) - \left(\frac{\mathbf{d}\cdot(\mathbf{a}-\mathbf{p})}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d}\right\|
\end{aligned}$$

In $\mathbb{R}^2$, if $A(x_0, y_0)$ and $l : ax + by + c = 0$,

$$\mathrm{d}(A, l) = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$

Derivation. (Excercise 1.3 39) Let $\mathbf{a} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$, and $\mathbf{d} = \begin{bmatrix} b \\ -a \end{bmatrix}$.

$$\begin{aligned}
\mathrm{d}(A, l) &= \|(\mathbf{a} - \mathbf{p}) - \mathrm{proj}_{\mathbf{d}}(\mathbf{a} - \mathbf{p})\| \\
&= \left\| \begin{bmatrix} (x_0 - x_1) - \frac{b(x_0 - x_1) - a(y_0 - y_1)}{a^2 + b^2} b \\ (y_0 - y_1) + \frac{b(x_0 - x_1) - a(y_0 - y_1)}{a^2 + b^2} a \end{bmatrix} \right\| \\
&= \left\| \begin{bmatrix} \frac{a}{a^2 + b^2} \{(ax_0 + by_0) - (ax_1 + by_1)\} \\ \frac{b}{a^2 + b^2} \{(ax_0 + by_0) - (ax_1 + by_1)\} \end{bmatrix} \right\| \\
&= \left\| \begin{bmatrix} \frac{a}{a^2 + b^2} (ax_0 + by_0 + c) \\ \frac{b}{a^2 + b^2} (ax_0 + by_0 + c) \end{bmatrix} \right\| \\
&= \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}
\end{aligned}$$

The distance from the point A to a plane $\mathcal{P}$ (or a hyperplane) with a normal vector $\mathbf{n}$, and point P is on the plane is represented as

$$\begin{aligned}
\mathrm{d}(A, \mathcal{P}) &= \|\mathrm{proj}_{\mathbf{n}}(\mathbf{a} - \mathbf{p})\| \\
&= \left\| \frac{\mathbf{n} \cdot (\mathbf{a} - \mathbf{p})}{\mathbf{n} \cdot \mathbf{n}} \right\| \|\mathbf{n}\|
\end{aligned}$$

In $\mathbb{R}^3$, if $A(x_0, y_0, z_0)$ and $\mathcal{P}$: $ax + by + cz + d = 0$,

$$\mathrm{d}(A, \mathcal{P}) = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}$$

Derivation. (Excercise 1.3 40) Let $\mathbf{a} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$, and $\mathbf{n} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

$$\begin{aligned}
\mathrm{d}(A, \mathcal{P}) &= |\frac{\mathbf{n} \cdot (\mathbf{a} - \mathbf{p})}{\mathbf{n} \cdot \mathbf{n}}| \|\mathbf{n}\| \\
&= |\frac{a(x_0 - x_1) + b(y_0 - y_1) + c(z_0 - z_1)}{a^2 + b^2 + c^2}| \sqrt{a^2 + b^2 + c^2} \\
&= \frac{|(ax_0 + by_0 + cz_0) - (ax_1 + by_1 + cz_1)|}{\sqrt{a^2 + b^2 + c^2}} \\
&= \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}
\end{aligned}$$

1.5 Solutions for Excercises Worthy to Solve

a. **Excercise 1.2 55**

Proof.

$$\begin{aligned}\mathrm{d}(\mathbf{u},\mathbf{v}) &= \|\mathbf{u}-\mathbf{v}\| \\ &= \|(-1)(\mathbf{v}-\mathbf{u})\| = |-1|\|\mathbf{v}-\mathbf{u}\| \\ &= \|\mathbf{v}-\mathbf{u}\| = \mathrm{d}(\mathbf{v},\mathbf{u})\end{aligned}$$

□

b. **Excercise 1.2 56**

Proof.

$$\begin{aligned}\mathrm{d}(\mathbf{u},\mathbf{w}) = \|\mathbf{u}-\mathbf{w}\| &= \|(\mathbf{u}-\mathbf{v})+(\mathbf{v}-\mathbf{w})\| \\ &\leq \|\mathbf{u}-\mathbf{v}\| + \|\mathbf{v}-\mathbf{w}\| = \mathrm{d}(\mathbf{u},\mathbf{v}) + \mathrm{d}(\mathbf{v},\mathbf{w})\end{aligned}$$

□

c. **Excercise 1.2 57**

Proof. By Theorem 1.3 (c), $\mathrm{d}(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| = 0$ if and only if $\mathbf{u}-\mathbf{v} = \mathbf{0}$. Thus, $\mathrm{d}(\mathbf{u},\mathbf{v}) = 0$ if and only if $\mathbf{u} = \mathbf{v}$. □

d. **Excercise 1.2 59**

Proof.

$$\begin{aligned}\|\mathbf{u}\| &= \|(\mathbf{u}-\mathbf{v})+\mathbf{v}\| \\ &\leq \|\mathbf{u}-\mathbf{v}\| + \|\mathbf{v}\| \\ \therefore \|\mathbf{u}-\mathbf{v}\| &\geq \|\mathbf{u}\| - \|\mathbf{v}\|\end{aligned}$$

□

e. **Excercise 1.2 60**

Solution. A well-known counterexample for $\mathbf{u}\cdot\mathbf{v} = \mathbf{u}\cdot\mathbf{w} \to \mathbf{v} = \mathbf{w}$ is the case when $\mathbf{u}\cdot\mathbf{v} = \mathbf{u}\cdot\mathbf{w} = 0$. Vectors orthogonal with $\mathbf{u}$ satisfy the condition, though they are not equal.

f. **Excercise 1.2 69**

Proof.

$$\begin{aligned}\mathbf{u}\cdot\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) &= \mathbf{u}\cdot\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}\right) \\ &= \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)(\mathbf{u}\cdot\mathbf{u}) = \mathbf{u}\cdot\mathbf{v} \\ \therefore \mathbf{u}\cdot(\mathbf{v}-\mathrm{proj}_{\mathbf{u}}(\mathbf{v})) &= \mathbf{u}\cdot\mathbf{v} - \mathbf{u}\cdot\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \mathbf{u}\cdot\mathbf{v} - \mathbf{u}\cdot\mathbf{v} = 0\end{aligned}$$

Therefore, $\mathbf{u}$ is orthogonal to $\mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$. □

g. **Excercise 1.2 70**

Proof. (a)

$$\begin{aligned}
\text{proj}_{\mathbf{u}}(\text{proj}_{\mathbf{u}}(\mathbf{v})) &= \left(\frac{\mathbf{u}\cdot\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}\right)}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} \\
&= \left(\frac{\left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)(\mathbf{u}\cdot\mathbf{u})}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} \\
&= \left(\frac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} = \text{proj}_{\mathbf{u}}(\mathbf{v})
\end{aligned}$$

(b) By Excercise 1.2 69, $\mathbf{u}$ is orthogonal to $\mathbf{v}-\text{proj}_{\mathbf{u}}(\mathbf{v})$, thus $\mathbf{u}\cdot(\mathbf{v}-\text{proj}_{\mathbf{u}}(\mathbf{v})) = 0$. Therefore, $\text{proj}_{\mathbf{u}}(\mathbf{v}-\text{proj}_{\mathbf{u}}(\mathbf{v})) = \mathbf{0}$.

$\square$

h. **Excercise 1.3 17**

Proof. Consider two lines with slopes m_1 and m_2. Then the direction vectors of two lines are $\begin{bmatrix}1\\m_1\end{bmatrix}$, and $\begin{bmatrix}1\\m_2\end{bmatrix}$, respectively. Since two vectors are orthogonal (or, perpendicular) if and only if $\begin{bmatrix}1\\m_1\end{bmatrix}\cdot\begin{bmatrix}1\\m_2\end{bmatrix} = 1+m_1m_2 = 0$, two lines are perpendicular if and only if $m_1m_2=-1$. $\square$

i. **Excercise 1.3 41**

Proof. Suppose $\mathbf{n}_1\cdot\mathbf{x}_1 = c_1$ and $\mathbf{n}_2\cdot\mathbf{x}_2 = c_2$. Then the distance between two lines is

$$\begin{aligned}
\|\text{proj}_{\mathbf{n}}(\mathbf{x}_2-\mathbf{x}_1)\| &= \left|\frac{\mathbf{n}\cdot(\mathbf{x}_2-\mathbf{x}_1)}{\mathbf{n}\cdot\mathbf{n}}\right|\|\mathbf{n}\| \\
&= \frac{|\mathbf{n}\cdot\mathbf{x}_2-\mathbf{n}\cdot\mathbf{x}_1|}{\|\mathbf{n}\|} \\
&= \frac{|c_2-c_1|}{\|\mathbf{n}\|}
\end{aligned}$$

$\square$

j. **Excercise 1.3 42**

Proof. Suppose $\mathbf{n}_1\cdot\mathbf{x}_1 = d_1$ and $\mathbf{n}_2\cdot\mathbf{x}_2 = d_2$. Then the distance between two planes is

$$\begin{aligned}
\|\text{proj}_{\mathbf{n}}(\mathbf{x}_2-\mathbf{x}_1)\| &= \left|\frac{\mathbf{n}\cdot(\mathbf{x}_2-\mathbf{x}_1)}{\mathbf{n}\cdot\mathbf{n}}\right|\|\mathbf{n}\| \\
&= \frac{|\mathbf{n}\cdot\mathbf{x}_2-\mathbf{n}\cdot\mathbf{x}_1|}{\|\mathbf{n}\|} \\
&= \frac{|d_2-d_1|}{\|\mathbf{n}\|}
\end{aligned}$$

$\square$

k. **Excercise 1.3 47**
Solution.

$$\begin{aligned}
\mathbf{p} \cdot \mathbf{n} &= (\mathbf{v} - c\mathbf{n}) \cdot \mathbf{n} \\
&= \mathbf{v} \cdot \mathbf{n} - c\|\mathbf{n}\|^2 = 0 \\
\therefore c &= \frac{\mathbf{v} \cdot \mathbf{n}}{\|\mathbf{n}\|^2} \\
\mathbf{p} &= \mathbf{v} - \frac{\mathbf{v} \cdot \mathbf{n}}{\|\mathbf{n}\|^2}\mathbf{n}
\end{aligned}$$

Chapter 2

Systems of Linear Equations

2.1 Terminology

Definition. A vector $\mathbf{v}$ is a **linear combination** of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ if there exist scalars $c_1, c_2, \cdots, c_n$ such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

The scalars $c_1, c_2, \cdots, c_n$ are called **coefficients** of linear combination.

Definition. A **linear equation** in the n variables $x_1, x_2, \cdots, x_n$ is an equation that can be written in the form of

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where the **coefficients** $a_1, a_2, \cdots, a_n$ and the **constant term** b are constants.

A **set of linear equations** is a finite set of linear equations with same variables. A **solution** of a linear equation $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ is a vector $\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ which satisfies the equation. A solution of a set of linear equations is a vector which is simultaneously a solution of all linear equations in the system. A **solution set** of a system of linear equations is the set of all solutions of the system.

A system of linear equations is **consistent** if there exists a solution. **Inconsistent** set of linear equations has an empty solution set. Two linear systems are **equivalent** if they have same solution set.

The **coefficient matrix** contains the coefficients of variables in the set of linear equations. The **augmented matrix** is the coefficient matrix augmented by a vector containing constant terms. For the system

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

the coefficient matrix is

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

and the augmented matrix is

$$\begin{bmatrix} A | \mathbf{b} \end{bmatrix} = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

Definition. A matrix is in **row echelon form** (REF) if:

a. All rows consisting entirely of zeros are at the bottom.

b. The **leading entry** (the first nonzero entry) of each rows is located to the left of any leading entries below it.

Definition. A matrix is in **reduced row echelon form** (RREF) if:

a. It is in REF.

b. All leading entries are 1. (**leading 1**)

c. Each columns containing a leading 1 has 0 everywhere else.

REF of a matrix is not unique, but all matrices have unique RREF.

Definition. A system of linear equations is **homogeneous** if the constant term in each equation is zero.

2.2 Solving Linear Systems

A system of linear equations with *real coefficients* has either a unique solution, or infinitely many solutions, or no solutions.

Proof. (Another proof is shown at Theorem 3.22) Consider a consistent set of m linear equations

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

Suppose this set has more than one solutions. Then there exists two solutions $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}$, where $\mathbf{x} \neq \mathbf{y}$. For any real number k,

$$\begin{aligned} & a_{i1}(kx_1 + (1-k)y_1) + a_{i2}(kx_2 + (1-k)y_2) + \cdots + a_{in}(kx_n + (1-k)y_n) \\ & = k(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n) + (1-k)(a_{i1}y_1 + a_{i2}y_2 + \cdots + a_{in}y_n) \\ & = kb_i + (1-k)b_i = b_i \end{aligned}$$

where $1 \leq i \leq m$. Thus, a vector

$$\mathbf{v}_k = \begin{bmatrix} kx_1 + (1-k)y_1 & kx_2 + (1-k)y_2 & \cdots & kx_n + (1-k)y_n \end{bmatrix}$$

is a also solution of the set of linear equations. Therefore, if the set of linear equations has more than one solutions, then there are infinitely many solutions. □

Definition. The **elementary row operations**

a. Interchanging two rows ($R_i \leftrightarrow R_j$)

b. Multiplying nonzero constant to a row (kR_i)

c. Adding a multiple of a row to another row ($R_i + kR_j$)

can be performed to a matrix, and the matrices before and after applying the elementary row operations are equivalent. Applying certain sequence of elementary row operations to bring a matrix into REF is called **row reduction**.

Definition. Matrices A and B are **row equivalent** if there exists a sequence of elementary row operation that converts A to B.

Theorem 2.1
Matrices A and B are row equivalent if and only if they can be reduced to same REF.

Proof. Suppose that matrices A and B can be reduced to REF R. Then there exists a sequence of elementary row operations which converts B to R. Reversing the sequence will convert R to B. Then combining the sequence of operations $A \to R$ and $R \to B$ will convert A to B. Therefore, A and B are row equivalent. □

Definition. The **rank** of a matrix is the number of nonzero rows in REF of a matrix.

The variables in the linear system corresponding to leading entires in REF of coefficient matrix is called **leading variables**. The rest are called **free variables**.

Note. A system of linear equations with at least one free variable has infinitely many solutions.

Theorem 2.2 : The Rank Theorem
Let A be the coefficient matrix of a system of linear equations with n variables. If the system is consistent, then

$$\text{number of free variables} = n - \text{rank}(A)$$

Proof. The number of leading variables in the system is equal to the number of nonzero rows of the coefficient matrix A, which is $\text{rank}(A)$. Therefore, the number of free variables is $n - \text{rank}(A)$. □

Theorem 2.3
A homogeneous system of linear equations $[A|\mathbf{0}]$ with n variables and m equations has infinitely many solutions if $m < n$.

Proof. Since there exists a trivial solution $\mathbf{0}$, the system is consistent. By the Rank Theorem (Thm 2.2),

$$\text{number of free variables} = n - \text{rank}(A) \geq n - m > 0$$

Therefore there exists at least one free variable, hence the system has infinitely many solutions. □

2.3 Solving Linear Systems : Example

The **Gaussian Elimination** is an algorithm which solves a linear system. The procedure is

a. Reduce the augmented matrix of the system into REF.

b. Using back substitution, solve the REF of the matrix.

The **Gauss-Jordan Elimination** also solves a linear system, but in easier way. The procedure is

a. Reduce the augmented matrix of the system into RREF.

b. Solve the leading variables in terms of free variables.

Solving a given system of linear equations

$$\begin{cases} x_1 - x_2 - x_3 + 2x_4 = 1 \\ 2x_1 - 2x_2 - x_3 + 3x_4 = 3 \\ -x_1 + x_2 - x_3 = -3 \end{cases}$$

using Gaussian Elimination starts from representing the system into augmented matrix

$$\left[\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 2 & -2 & -1 & 3 & 3 \\ -1 & 1 & -1 & 0 & -3 \end{array}\right]$$

Then reduce the augmented matrix using elementary row operations.

$$\left[\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 2 & -2 & -1 & 3 & 3 \\ -1 & 1 & -1 & 0 & -3 \end{array}\right] \xrightarrow{R_2-2R_1, R_3+R_1} \left[\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & -2 & 2 & -2 \end{array}\right]$$

$$\xrightarrow{R_3+2R_2} \left[\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The system is consistent because constants of zero-rows are all 0. (If not, the system will be inconsistent) The leading entries are in first and third column, so leading variables are x_1 and x_3, and free variables are x_2 and x_4. Using back substitution, solve x_1 and x_3 in terms of x_2 and x_4.

$$x_3 - x_4 = 1 \rightarrow x_3 = x_4 + 1$$
$$x_1 - x_2 - x_3 + 2x_4 = 1 \rightarrow x_1 = x_2 - x_4 + 2$$

You can obtain same result with reducing REF into RREF.

$$\left[\begin{array}{cccc|c} 1 & -1 & -1 & 2 & 1 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{R_1+R_2} \left[\begin{array}{cccc|c} 1 & -1 & 0 & 1 & 2 \\ 0 & 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Representing the variables into form of vector will give you the solution set. (x_2 corresponds to s and x_4 to t in the solution.)

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad (s, t \in \mathbb{R})$$

2.4 Spanning Sets and Linear Independence

Theorem 2.4
A system of linear equations with augmented matrix $[A|\mathbf{b}]$ is consistent if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Proof. Suppose a system of linear equations with augmented matrix

$$[A|\mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

is consistent. Then there exist real numbers $c_1, c_2, \cdots, c_n$ which satisfy

$$a_{i1}c_1 + a_{i2}c_2 + \cdots + a_{in}c_n = b_i, \text{for all } 1 \leq i \leq m$$

Therefore, b can be represented with linear combination of columns of A.

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = c_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + c_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + c_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

Similarly, if b is a linear combination of the columns of A, the coefficients of linear combination satisfy the linear equations of system. Therefore, the system is consistent. □

Definition. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$. Then the set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ is called the **span** of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ and denoted by $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)$ or $\text{span}(S)$. If $\text{span}(S) = \mathbb{R}^n$, then S is a **spanning set** for $\mathbb{R}^n$.

Note. To prove that the spanning set $\text{span}(S)$ is equal to another set T, proving **both** $S \subset T$ and $T \subset S$ is essencial.

Definition. A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_n$ is **linearly dependent** if there exists scalars $c_1, c_2, \cdots c_n$, at least one of which is nonzero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

A set of vectors which are not linearly dependent is **linearly independent**.

Theorem 2.5
Vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent if and only if at least one of the vectors can be represented with a linear combination of other vectors.

Proof. If vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent, there exist scalars $c_1, c_2, \cdots, c_n$ such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$ and at least one of scalars is nonzero. Suppose $c_1 \neq 0$. Then

$$\mathbf{v}_1 = -\frac{c_2}{c_1}\mathbf{v}_2 - \cdots - \frac{c_n}{c_1}\mathbf{v}_n$$

Thus $\mathbf{v}_1$ is represented with a linear combination of other vectors.
Suppose that $\mathbf{v}_1$ can be expressed as linear combination of $\mathbf{v}_2, \mathbf{v}_3, \cdots, \mathbf{v}_n$. Then there exist scalars $c_2, c_3, \cdots, c_n$ which satisfy $\mathbf{v}_1 = c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_n\mathbf{v}_n$. Then

$$\mathbf{v}_1 - c_2\mathbf{v}_2 - \cdots - c_n\mathbf{v}_n = \mathbf{0}$$

thus vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent. □

Theorem 2.6
Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be vectors in $\mathbb{R}^n$and let A be the $n \times m$ matrix $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix}$. Then $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent if and only if the homogeneous linear system with augmented matrix $[A|\mathbf{0}]$ has a nontrivial solution (or, infinitely many solutions.)

Proof. Vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent if and only if there exist scalars $c_1, c_2, \cdots, c_n$ such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$ where at least one of scalars is nonzero. Since the nonzero vector $\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ is a solution of the linear system with augmented matrix $[A|\mathbf{0}]$, the system has nontrivial solution. Similarly, if there exists a nontrivial solution $\begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{bmatrix}$ of the homogeneous linear system, the components satisfy

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

Since at least one of $c_1, c_2, \cdots, c_n$ is nonzero, vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent. □

Extra Theorem : Elementary Row Operations and Linear Combination
(Example 2.25) If matrix B is generated by applying elementary row operations at matrix A, then rows of B can be represented as nontrivial linear combination of rows of A.

Proof. Let $A = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \vdots \\ \mathbf{R}_n \end{bmatrix}$. By Theorem 3.10, sequence of elementary row operations that would convert A into B will also convert I_n to E, where $B = EA$. Since there exists the sequence of elementary row operations converting I_n to E, I_n and E are row equivalent, so $\text{rank}(I_n) = \text{rank}(E) = n$. Then the rows of E are all nonzero rows. Therefore, rows of $B = EA$ can be expressed as nontrivial linear combination of rows of A. □

Theorem 2.7
Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be row vectors in $\mathbb{R}^n$and let A be the $m \times n$ matrix $\begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_n \end{bmatrix}$ Then $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent if and only if $\text{rank}(A) < n$.

Proof. Assume that $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent. By Theorem 2.5, at least one of the vectors can be written as a linear combination of other vectors. With relabeling such vector to $\mathbf{v}_n$, there exist scalars $c_1, c_2, \cdots, c_{n-1}$ such that

$$\mathbf{v}_n = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{n-1}\mathbf{v}_{n-1}$$

Then performing the elementary row operations $(R_n - c_1R_1), (R_n - c_2R_2), \cdots, (R_n - c_{n-1}R_{n-1})$ will create zero row in nth row. Therefore, $\text{rank}(A) < n$.
Conversely, if $\text{rank}(A) < n$, A would have zero row in its REF, thus there exists certain sequence of elementary operation that would create zero row from A. There exist scalars $c_1, c_2, \cdots, c_n$ such that at least one of scalars is nonzero, and

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

Therefore, $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent. □

Theorem 2.8
Any set of m vectors in $\mathbb{R}^n$is linearly dependent if $m > n$.

Proof. (also solution for Exercise 2.3 45) Let A be the $m \times n$ matrix with the vectors as its rows. Since $\text{rank}(A) < n < m$, by Theorem 2.7, the vectors are linearly dependent. □

2.5 Solutions of Exercises Worthy to Solve

a. **Exercise 2.2 39**

Proof. The augmented matrix of the system is $\left[\begin{array}{cc|c} a & b & r \\ c & d & s \end{array}\right]$. Applying elementary row operations $(cR_1), (aR_2), (R_2 - R_1)$, the REF of the augmented matrix is $\left[\begin{array}{cc|c} ac & bc & rc \\ 0 & ad - bc & sa - rc \end{array}\right]$. If $ad - bc \neq 0$, the rank of augmented matrix is 2, therefore the system has unique solution. □

b. **Exercise 2.2 51**

Proof. Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$. Since $\mathbf{u} \cdot \mathbf{x} = \mathbf{v} \cdot \mathbf{x} = 0$, $\mathbf{x}$ is a solution of the linear system with augmented matrix $\left[\begin{array}{ccc|c} u_1 & u_2 & u_3 & 0 \\ v_1 & v_2 & v_3 & 0 \end{array}\right]$. The REF of augmented matrix is $\left[\begin{array}{ccc|c} u_1 & u_2 & u_3 & 0 \\ 0 & u_1v_2 - u_2v_1 & u_1v_3 - u_3v_1 & 0 \end{array}\right]$, and the solution is

$$\mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}$$

where $t \in \mathbb{R}$. □

c. **Exercise 2.2 59**

Proof. By Theorem 2.2, there are $n - \text{rank}(A)$ free variables and $\text{rank}(A)$ leading variables in the linear system. Since the values of leading variables are fixed for given values of free variables, the number of solution is $p^{n-\text{rank}(A)}$. $\square$

d. **Exercise 2.2 60**

Solution. Complications arise when we try to solve this directly in $\mathbb{Z}_6$. Therefore, we split the system to $\mathbb{Z}_2$ and $\mathbb{Z}_3$, then find the solutions simultaneously satisfying them.

In $\mathbb{Z}_2$,

$$\left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 1 & 0 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{cc|c} 0 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

then the solution is

$$\mathbf{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \end{bmatrix} (t = 0, 1)$$

In $\mathbb{Z}_3$,

$$\left[\begin{array}{cc|c} 2 & 0 & 1 \\ 1 & 0 & 2 \end{array}\right] \xrightarrow{R_2 + R_1} \left[\begin{array}{cc|c} 2 & 0 & 1 \\ 0 & 0 & 0 \end{array}\right]$$

then the solution is

$$\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \end{bmatrix} (t = 0, 1, 2)$$

Combining these solutions, we have

$$\mathbf{x} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 3 \\ 2 \end{bmatrix} (t \in \mathbb{Z}_6)$$

e. **Exercise 2.3 20**

Proof. **(a)** Suppose $\mathbf{v} \in \text{span}(S)$. Then $\mathbf{v}$ can be expressed as linear combination of vectors in S, so there exist scalars $c_1, c_2, \cdots, c_k$ such that $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$. Then $\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_m$, so $\mathbf{v}$ is also a linear combination of vectors in T. Therefore, $\text{span}(S) \subset \text{span}(T)$.

(b) By (a), $\text{span}(S) = \mathbb{R}^n \subset \text{span}(T)$. Also, $\text{span}(T) \subset \mathbb{R}^n$ since $\text{span}(T)$ is a set of vectors in $\mathbb{R}^n$. Therefore, $\mathbb{R}^n = \text{span}(T)$. $\square$

f. **Exercise 2.3 21**

Proof. **(a)** Since each $\mathbf{u}_i$ are linear combinations of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$, there exist scalars $c_{i1}, c_{i2}, \cdots, c_{im}$ for each i such that $\mathbf{u}_i = c_{i1}\mathbf{v}_1 + c_{i2}\mathbf{v}_2 + \cdots + c_{im}\mathbf{v}_m$. Suppose vector $\mathbf{w}$ is in $\text{span}(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k)$. Since $\mathbf{w}$ is a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k$, there exist scalars $d_1, d_2, \cdots, d_k$ such that $\mathbf{w} = d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_k\mathbf{u}_k$. Then

$$\begin{aligned}\mathbf{w} &= d_1\mathbf{u}_1 + d_2\mathbf{u}_2 + \cdots + d_k\mathbf{u}_k \\ &= d_1(c_{11}\mathbf{v}_1 + \cdots + c_{1m}\mathbf{v}_m) + \cdots + d_k(c_{k1}\mathbf{v}_1 + \cdots + c_{km}\mathbf{v}_m) \\ &= (c_{11}d_1 + c_{21}d_2 + \cdots + c_{k1}d_k)\mathbf{v}_1 + \cdots + (c_{1m}d_1 + c_{2m}d_2 + \cdots + c_{km}d_k)\mathbf{v}_m\end{aligned}$$

Therefore, $\mathbf{w}$ can be expressed as a linear combination of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$, so $\mathbf{w} \in \text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)$.

(b) The same logic can be applied to prove that $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m) \subset \text{span}(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k)$. Therefore, $\text{span}(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k) = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)$.

(c) Let $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}1\\1\\1\end{bmatrix}$, then $\mathbf{v}_1 = \mathbf{e}_1$, $\mathbf{v}_2 = \mathbf{e}_1 + \mathbf{e}_2$, and $\mathbf{v}_3 = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$. Also, $\mathbf{e}_1 = \mathbf{v}_1$, $\mathbf{e}_2 = \mathbf{v}_2 - \mathbf{v}_1$, and $\mathbf{e}_3 = \mathbf{v}_3 - \mathbf{v}_2$. Since $\text{span}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \mathbb{R}^3$, by (b), $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \text{span}(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) = \mathbb{R}^3$. □

g. **Exercise 2.3 44**

Proof. Suppose that vectors $\mathbf{v}_1, \mathbf{v}_2$ are linearly dependent. Then there exist scalars c_1, c_2 such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$.

(i) $c_1c_2 \neq 0$

$$\mathbf{v}_1 = -\frac{c_2}{c_1}\mathbf{v}_2$$

(ii) If not, without loss of generality, $c_1 \neq 0$ and $c_2 = 0$.

$$\begin{aligned}c_1\mathbf{v}_1 + c_2\mathbf{v}_2 &= \mathbf{0} \\ c_1\mathbf{v}_1 &= \mathbf{0} \\ \therefore \mathbf{v}_1 &= \mathbf{0} = 0\mathbf{v}_2\end{aligned}$$

□

h. **Exercise 2.3 46**

Proof. Given a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$, suppose that vectors $\mathbf{v}_{a_1}, \mathbf{v}_{a_2}, \cdots, \mathbf{v}_{a_m}$ are linearly dependent, where $\{a_1, a_2, \cdots, a_m\} \subset \{1, 2, \cdots n\}$. Then there exists scalars $d_1, d_2, \cdots, d_m$ such that at least one of scalars is nonzero and $d_1\mathbf{v}_{a_1} + d_2\mathbf{v}_{a_2} + \cdots + d_m\mathbf{v}_{a_m} = \mathbf{0}$.

Let

$$c_i = \begin{cases} d_j & \text{if } i = a_j \\ 0 & \text{if } i \notin \{a_1, a_2, \cdots, a_m\}\end{cases}$$

then $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = d_1\mathbf{v}_{a_1} + d_2\mathbf{v}_{a_2} + \cdots + d_m\mathbf{v}_{a_m} = \mathbf{0}$. Therefore, the set $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ is linearly dependent. The contrapositive proposition that the subset of linearly independent set of vectors is always linearly independent is also true. □

i. **Exercise 2.3 47**

Proof. We prove this proposition by Exercise 2.3 21.

(i) For every $\mathbf{v}_i \in S'$, $\mathbf{v}_i = 0\mathbf{v}_1 + \cdots + 1\mathbf{v}_i + \cdots + 0\mathbf{v}_k + 0\mathbf{v}$. Thus, every vectors in S' can be expressed as a linear combination of vectors in S.

(ii) For every $\mathbf{v}_i \in S$, $\mathbf{v}_i = 0\mathbf{v}_1 + \cdots + 1\mathbf{v}_i + \cdots + 0\mathbf{v}_k$, and $\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$. Thus, every vectors in S can be expressed as a linear combination of vectors in S'.

By Exercise 2.3 21(b), $\text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k, \mathbf{v}) = \text{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$. □

j. **Exercise 2.3 48**

Proof. Suppose that $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k (c_1 \neq 0)$ and vectors $\mathbf{v}, \mathbf{v}_2, \cdots, \mathbf{v}_k$ are linearly dependent. Then □

Chapter 3

Matrices

3.1 Terminology

Definition. A **matrix** is a rectangular array of numbers, which are called as **entries** or **elements**. If the matrix has n rows and m columns, the **size** of the matrix is $n \times m$.

A $1 \times n$ matrix is called a **row matrix**, or **row vector**. A $n \times 1$ matrix is called a **column matrix**, or **column vector**. (*A vector is considered as a matrix*) We can denote matrices using row vectors or column vectors, such as

$$A = \begin{bmatrix} \mathbf{A}_1^C & \mathbf{A}_2^C & \cdots & \mathbf{A}_m^C \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1^R \\ \mathbf{A}_2^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix}$$

where $\mathbf{A}_i^C$ is the ith column of A and $\mathbf{A}_i^R$ is the ith row of A.

The element at ith row and jth column is denoted by A_{ij}. We can also denote matrices using elements, such as $A = [A_{ij}]$.

Definition. The **diagonal entries**of A are A_{ii}.

Definition. The **square matrix** is a matrix which has same number of rows and columns (so the size is $n \times n$). **Diagonal matrix** is a square matrix which has its nondiagonal entries as 0. A diagonal matrix with all of its diagonal entries are the same are **scalar matrix**. If the value of diagonal entries are all 1, it is **identity matrix**.
A $n \times n$ identity matrix is denoted as I_n, and

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

A **zero matrix** O is a matrix which all of its entires are zero.

Definition. Two matrices are **equal** if and only if
(i) The size of two matrices are the same.
(ii) The corresponding entries of the matrices are the same.

3.2 Matrix Operations

Definition. If A and B are both $n \times m$ matrices, the **sum** of A and B is defined as

$$A + B = [A_{ij} + B_{ij}]$$

where $A + B$ is also an $n \times m$ matrix.

Definition. If A is an $n \times m$ matrix and c is a scalar, the **scalar multiplication** cA is defined as

$$cA = [cA_{ij}]$$

where cA is also an $n \times m$ matrix.

Definition. If A is an $m \times n$ matrix and B is an $n \times r$ matrix, then the **product** AB is defined as

$$AB = [A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj}]$$

where AB is an $m \times r$ matrix.

* *Note.* $(AB)_{ij} = \mathbf{A}_i^R \cdot \mathbf{B}_j^C$

Defintion. A **transpose** of an $n \times m$ matrix A is denoted as A^T, which is an $m \times n$ matrix and

$$(A^T)_{ij} = A_{ji}$$

Definition. A square matrix A is **symmetric** if $A^T = A$.

Matrices can be divided into **submatrices** by partitioning the matrix into ceratin blocks. We introduce partitioned matrix in order to perform matrix multiplication in easier way.

Extra Theorem : Multiplication of Partitioned Matrices

If matrices A and B are partitioned as $A = \begin{bmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nm} \end{bmatrix}$ and $B = \begin{bmatrix} B_{11} & \cdots & B_{1r} \\ \vdots & & \vdots \\ B_{m1} & \cdots & B_{mr} \end{bmatrix}$, then AB can be partitioned as

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} + \cdots + A_{1m}B_{m1} & \cdots & A_{11}B_{1r} + A_{12}B_{2r} + \cdots + A_{1m}B_{mr} \\ \vdots & & \vdots \\ A_{n1}B_{11} + A_{n2}B_{21} + \cdots + A_{nm}B_{m1} & \cdots & A_{n1}B_{1r} + A_{n2}B_{2r} + \cdots + A_{nm}B_{mr} \end{bmatrix}$$

assuming that all the products are defined.

Proof. Let A be an $n \times m$ matrix and let B be an $m \times r$ matrix so that AB is defined. Then

$$AB = A\begin{bmatrix}\mathbf{B}_1^C & \mathbf{B}_2^C & \cdots & \mathbf{B}_r^C\end{bmatrix} = \begin{bmatrix}A\mathbf{B}_1^C & A\mathbf{B}_2^C & \cdots & A\mathbf{B}_r^C\end{bmatrix} \text{ (matrix-column representation)}$$

$$= \begin{bmatrix}\mathbf{A}_1^R \\ \mathbf{A}_2^R \\ \vdots \\ \mathbf{A}_n^R\end{bmatrix} B = \begin{bmatrix}\mathbf{A}_1^R B \\ \mathbf{A}_2^R B \\ \vdots \\ \mathbf{A}_n^R B\end{bmatrix} \text{ (row-matrix representation)}$$

$$= \begin{bmatrix}\mathbf{A}_1^R \\ \mathbf{A}_2^R \\ \vdots \\ \mathbf{A}_n^R\end{bmatrix}\begin{bmatrix}\mathbf{B}_1^C & \mathbf{B}_2^C & \cdots & \mathbf{B}_r^C\end{bmatrix} = \begin{bmatrix}\mathbf{A}_1^R\mathbf{B}_1^C & \cdots & \mathbf{A}_1^R\mathbf{B}_r^C \\ \vdots & & \vdots \\ \mathbf{A}_n^R\mathbf{B}_1^C & \cdots & \mathbf{A}_n^R\mathbf{B}_r^C\end{bmatrix} \text{ (row-column representation)}$$

$$= \begin{bmatrix}\mathbf{A}_1^C & \mathbf{A}_2^C & \cdots & \mathbf{A}_m^C\end{bmatrix}\begin{bmatrix}\mathbf{B}_1^R \\ \mathbf{B}_2^R \\ \vdots \\ \mathbf{B}_m^R\end{bmatrix} \text{ (column-row representation)}$$

$$= \mathbf{A}_1^C\mathbf{B}_1^R + \cdots + \mathbf{A}_m^C\mathbf{B}_m^R \text{ (outer product expansion)}$$

□

Theorem 3.1
Let A be an $n \times m$ matrix. Then

a. $\mathbf{e}_i A = \mathbf{A}_i^R$

b. $A\mathbf{e}_j = \mathbf{A}_j^C$

Proof. **(a)** (Exercise 3.1 41) Since the size of $\mathbf{e}_i$ is $1 \times n$ and the size of A is $n \times m$, the size of $\mathbf{e}_i A$ is $1 \times m$, which is same with the size of $\mathbf{A}_i^R$. Also,

$$(\mathbf{e}_i A)_{1k} = 0A_{1k} + \cdots + 1A_{ik} + \cdots + 0A_{nk} = A_{ik}$$

Therefore $\mathbf{e}_i A = \begin{bmatrix}A_{i1} & A_{i2} & \cdots & A_{im}\end{bmatrix} = \mathbf{A}_i^R$.

(b) Since the size of A is $n \times m$ and the size of $\mathbf{e}_j$ is $m \times 1$, the size of $A\mathbf{e}_j$ is $n \times 1$, which is same with the size of $\mathbf{A}_i^C$. Also,

$$(A\mathbf{e}_j)_{k1} = 0A_{k1} + \cdots + 1A_{ki} + \cdots + 0A_{km} = A_{ki}$$

Therefore $A\mathbf{e}_j = \begin{bmatrix}A_{1i} \\ A_{2i} \\ \vdots \\ A_{ni}\end{bmatrix} = \mathbf{A}_i^C$. □

Theorem 3.2 : Algebraic Properties of Matrix Addition and Scalar Multiplication
Let A, B, and C be matrices of the same size and let c and d be scalars.

a. $A + B = B + A$ (Commutativity of Matrix Addition)

b. $(A+B)+C = A+(B+C)$ (Associativity of Matrix Addition)

c. $A+O = A$

d. $A+(-A) = O$

e. $c(A+B) = cA+cB$ (Left Distributivity of Scalar Multiplication over Matrix Addition)

f. $(c+d)A = cA+dA$ (Right Distributivity of Scalar Multiplication over Matrix Addition)

g. $c(dA) = (cd)A$

h. $1A = A$

Proof. The size of all matrices in both sides of equation will be equal to the size of A, B, and C. The proof in componentwise perspective is on below.

a. (Exercise 3.2 17)

$$(A+B)_{ij} = A_{ij}+B_{ij} = B_{ij}+A_{ij} = (B+A)_{ij}$$

b. (Exercise 3.2 17)

$$\begin{aligned}((A+B)+C)_{ij} = (A+B)_{ij}+C_{ij} &= (A_{ij}+B_{ij})+C_{ij} \\ &= A_{ij}+(B_{ij}+C_{ij}) = A_{ij}+(B+C)_{ij} = (A+(B+C))_{ij}\end{aligned}$$

c. (Exercise 3.2 17)

$$(A+O)_{ij} = A_{ij}+O_{ij} = A_{ij}+0 = A_{ij}$$

d. (Exercise 3.2 17)

$$(A+(-A))_{ij} = A_{ij}+(-A)_{ij} = A_{ij}+(-A_{ij}) = 0$$

e. (Exercise 3.2 18)

$$(c(A+B))_{ij} = c(A+B)_{ij} = c(A_{ij}+B_{ij}) = cA_{ij}+cB_{ij} = (cA)_{ij}+(cB)_{ij} = (cA+cB)_{ij}$$

f. (Exercise 3.2 18)

$$((c+d)A)_{ij} = (c+d)A_{ij} = cA_{ij}+dA_{ij} = (cA)_{ij}+(dA)_{ij} = (cA+dA)_{ij}$$

g. (Exercise 3.2 18)

$$(c(dA))_{ij} = c(dA)_{ij} = c(dA_{ij}) = (cd)A_{ij} = ((cd)A)_{ij}$$

h. (Exercise 3.2 18)

$$(1A)_{ij} = 1A_{ij} = A_{ij}$$

$\square$

Theorem 3.3 : Properties of Matrix Multiplication
Let A, B, and C be matrices and let k be the scalar. If the operations can be defined,

a. $A(BC) = (AB)C$

b. $A(B + C) = AB + AC$

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

d. $k(AB) = (kA)B = A(kB)$

e. $I_nA = A = AI_m$, where size of A is $n \times m$

Proof. The proof of (a) is also done in page 229 of the textbook. (Not Included in Exam)

a. Let A be a $n \times m$ matrix, B be a $m \times p$ matrix, and C be a $p \times q$ matrix. Then the size of BC is $m \times q$, so the size of $A(BC)$ is $n \times q$. Also, the size of AB is $n \times p$, so the size of $(AB)C$ is $n \times q$. Therefore, the size of $A(BC)$ and $(AB)C$ are the same.

$$\begin{aligned}
((AB)C)_{ij} &= \sum_{k=1}^{p}((AB)_{ik}C_{kj}) \\
&= \sum_{k=1}^{p}(\sum_{l=1}^{m} A_{il}B_{lk})C_{kj} \\
&= \sum_{k=1}^{p}\sum_{l=1}^{m}(A_{il}B_{lk}C_{kj}) \\
&= \sum_{l=1}^{m} A_{il}(\sum_{k=1}^{p} B_{lk}C_{kj}) \\
&= \sum_{l=1}^{m} A_{il}(BC)_{lj} \\
&= (A(BC))_{ij}
\end{aligned}$$

b. Let A be a $n \times m$ matrix, and let B and C be $m \times r$ matrices. Then the size of $B + C$ is $m \times r$, so the size of $A(B + C)$ is $n \times r$. The size of AB and AC are also $n \times r$, so the size of matrices at both sides of equations are the same.

$$\begin{aligned}
(A(B + C))_{ij} &= \mathbf{A}_i^R \cdot ((B + C)_j^C) \\
&= \mathbf{A}_i^R \cdot (\mathbf{B}_j^C + \mathbf{C}_j^C) \\
&= \mathbf{A}_i^R \cdot \mathbf{B}_j^C + \mathbf{A}_i^R \cdot \mathbf{C}_j^C \\
&= (AB)_{ij} + (AC)_{ij} = (AB + AC)_{ij}
\end{aligned}$$

c. (Exercise 3.2 19) Let A and B be $n \times m$ matrices, and let C be a $m \times r$ matrix. Then the size of $A + B$ is $n \times m$, so the size of $(A + B)C$ is $n \times r$. The size of AC and BC are also $n \times r$, so the size of matrices at both sides of equations are the same.

$$\begin{aligned}
((A+B)C)_{ij} &= ((A+B)_i^R) \cdot \mathbf{C}_j^C \\
&= (\mathbf{A}_i^R + \mathbf{B}_i^R) \cdot \mathbf{C}^C)_j \\
&= \mathbf{A}_i^R \cdot \mathbf{C}_j^C + \mathbf{B}_i^R \cdot \mathbf{C}_j^C \\
&= (AC)_{ij} + (BC)_{ij} = (AC + BC)_{ij}
\end{aligned}$$

d. (Exercise 3.2 20) Let A be a $n \times m$ matrix and B be a $m \times r$ matrix. Then the size of $k(AB)$, $(kA)B$, and $A(kB)$ are all $n \times r$, so all the size of the matrices are the same.

$$\begin{aligned}
(k(AB))_{ij} &= k(\mathbf{A}_i^R \cdot \mathbf{B}_j^C) \\
&= (k\mathbf{A}_i^R) \cdot \mathbf{B}_j^C = ((kA)B)_{ij} \\
&= \mathbf{A}_i^R \cdot (k\mathbf{B}_j^C) = (A(kB))_{ij}
\end{aligned}$$

e. (Exercise 3.2 21) Let A be a $n \times m$ matrix. Since $I_n = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}$ and $I_m = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_m \end{bmatrix}$,

$$\begin{aligned}
I_n A &= \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix} A = \begin{bmatrix} \mathbf{e}_1 A \\ \vdots \\ \mathbf{e}_n A \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix} = A \\
AI_m = A \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_m \end{bmatrix} &= \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{e}_m \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1^C & \cdots & \mathbf{A}_m^C \end{bmatrix} = A
\end{aligned}$$

□

Definition. For a square matrix A and nonnegative integer n, define A^n as

$$A^n = \begin{cases} I & \text{if } n = 0 \\ A^{n-1}A = AA^{n-1} & \text{otherwise} \end{cases}$$

Extra Theorem : Matrix Powers
If A is a square matrix and r, s are nonnegative integers,

a. $A^r A^s = A^{r+s}$

b. $(A^r)^s = A^{rs}$

Proof. Let A be an $n \times n$ matrix.

a. Claim 1 : For a nonnegative integer r, $A^r A^s = A^{r+s}$ for all nonnegative integer s.
   (i) $A^r A^0 = A^r I_n = A^r$
   (ii) Suppose that $A^r A^k = A^{r+k}$. Then $A^r A^{k+1} = A^r A^k A = A^{r+k} A = A^{r+k+1}$.

By (i), (ii), for all nonnegative integer s, $A^r A^s = A^{r+s}$.

Claim 2 : For a nonnegative integer s, $A^r A^s = A^{r+s}$ for all nonnegative integer r.
(i) $A^0 A^s = I_n A^s = A^s$
(ii) Suppose that $A^k A^s = A^{k+s}$. Then $A^{k+1} A^s = AA^k A^s = AA^{k+s} = A^{k+s+1}$.
By (i), (ii), for all nonnegative integer r, $A^r A^s = A^{r+s}$.

By Claim 1 and 2, for all nonnegative integers r, s, $A^r A^s = A^{r+s}$.

b. For any nonnegative integer r,
(i) $(A^r)^0 = I_n = A^0 = A^{r\times 0}$
(ii) Suppose that $(A^r)^k = A^{rk}$. Then $(A^r)^{k+1} = (A^r)^k A^r = A^{rk} A^r = A^{rk+r} = A^{r(k+1)}$.
By (i), (ii), for all nonnegative integers r, s, $(A^r)^s = A^{rs}$.

□

Theorem 3.4 : Properties of Transpose
Let A and B be matrices and let k be a scalar. If the operations are defined,

a. $(A^T)^T = A$
b. $(A + B)^T = A^T + B^T$
c. $(kA)^T = k(A^T)$
d. $(AB)^T = B^T A^T$
e. $(A^r)^T = (A^T)^r$ for all r $\in \mathbb{Z}^+$

Proof. Let A and B be matrices that all the operations are defined.

a. (Exercise 3.2 30) If A is an $n \times m$ matrix, A^T is an $m \times n$ matrix, thus $(A^T)^T$ is an $n \times m$ matrix. The size of $(A^T)^T$ and A is the same.

$$((A^T)^T)_{ij} = (A^T)_{ji} = A_{ij}$$

b. (Exercise 3.2 30) If A and B are both $n \times m$ matrices, the size of $(A + B)$ is also $n \times m$, so $(A + B)^T$ and A^T and B^T are all $m \times n$ matrices.

$$((A + B)^T)_{ij} = (A + B)_{ji} = A_{ji} + B_{ji} = (A^T)_{ij} + (B^T)_{ij} = (A^T + B^T)_{ij}$$

c. (Exercise 3.2 30) If A is an $n \times m$ matrix, both $(kA)^T$ and $k(A^T)$ are $m \times n$ matrices.

$$((kA)^T)_{ij} = (kA)_{ji} = k(A_{ji}) = k(A^T)_{ij} = (k(A^T))_{ij}$$

d. If A is an $n \times m$ matrix and B is an $m \times r$ matrices, then $(AB)^T$ is a $r \times n$ matrix. The size of B^T and A^T is $r \times m$ and $m \times n$, so the size of $B^T A^T$ is $r \times n$.

$$((AB)^T)_{ij} = (AB)_{ji} = \mathbf{A}_j^R \cdot \mathbf{B}_i^C = (\mathbf{A}^T)_j^C \cdot (\mathbf{B}^T)_i^R = (B^T A^T)_{ij}$$

e. (Exercise 3.2 31) Let A be a $n \times n$ matrix.
   (i) $(A^0)^T = (I_n)^T = I_n = (A^T)^0$
   (ii) Suppose that for an arbitrary nonnegative integer r, $(A^r)^T = (A^T)^r$. Then

$$(A^{r+1})^T = (A^r A)^T = A^T (A^r)^T = A^T (A^T)^r = (A^T)^{r+1}$$

   By (i) and (ii), $(A^r)^T = (A^T)^r$ for all nonnegative integer r.

□

Theorem 3.5
For any matrix A,

a. If A is a square matrix, then $A + A^T$ is symmetrical.

b. AA^T and $A^T A$ are symmetrical.

Proof.

a. Let A be a square matrix. Then

$$(A + A^T)^T = A^T + (A^T)^T = A^T + A = A + A^T$$

   Therefore $A + A^T$ is symmetrical.

b. (Exercise 3.2 34) Let A be a $n \times m$ matrix. Then the size of A^T is $m \times n$ matrix, so the products AA^T and $A^T A$ are defined.

$$(AA^T)^T = (A^T)^T A^T = AA^T$$
$$(A^T A)^T = A^T (A^T)^T = A^T A$$

   Therefore AA^T and $A^T A$ are symmetrical.

□

3.3 Other Types of Matrices : Exercise 3.2

Definition. An **upper triangular** matrix is a square matrix with all of its entries below the main diagonal are zero. In other words, A is an upper triangular matrix if

$$A_{ij} = 0 \text{ if } i > j$$

Extra Theorem : Properties of Upper Triangular Matrices
If A and B are both upper triangular matrices in the same size, the product AB is also an upper triangular matrix.

Proof. (Exercise 3.2 29)
(i) If A and B are both 1×1 upper triangular matrices, let $A = [A_{11}]$ and $B = [B_{11}]$. Then $AB = [A_{11}B_{11}]$, which is also an upper triangular matrix.
(ii) Suppose that the product of $n \times n$ upper triangular matrices is also upper triangular matrix. Let A and B be both $n+1 \times n+1$ upper triangular matrices. Then we can partition A and B as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_{11} & B_{12} \\ O & B_{22} \end{bmatrix}$$

where A_{11} and B_{11} are $n \times n$ matrices, A_{12} and B_{12} are $1 \times n$ matrices, and A_{22} and B_{22} are scalars. Since A and B are upper triangular, A_{11} and B_{11} are also upper triangular matrices.

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}O & A_{11}B_{12} + A_{12}B_{22} \\ OB_{11} + A_{22}O & OB_{12} + A_{22}B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} & A_{11}B_{12} + A_{12}B_{22} \\ O & A_{22}B_{22} \end{bmatrix}$$

Since $A_{11}B_{11}$ is an upper triangular matrix, AB is also upper triangular.
By (i) and (ii), for all upper triangular matrices A and B, the product AB is also upper triangular.

□

Extra Theorem : Properties of Symmetric Matrices a. If A and B are symmetric, then $A + B$ is also symmetric.

b. If A is symmetric, then kA is also symmetric for any scalar k.

c. If A and B are symmetric, then AB is symmetric if and only if $AB = BA$.

Proof. Let A and B be symmetric matrices that the operations are defined, and let k be a scalar.

a. (Exercise 3.2 35)

$$(A+B)^T = A^T + B^T = A + B$$

Therefore $A + B$ is symmetric.

b. (Exercise 3.2 35)

$$(kA)^T = k(A^T) = kA$$

Therefore kA is symmetric.

c. (Exercise 3.2 36)

$$(AB)^T = B^T A^T = BA$$

Therefore AB is symmetric if and only if $AB = BA$.

□

Definition. A square matrix is **skew-symmetric** if $A^T = -A$. In other words, $A_{ij} = -A_{ji}$.

Extra Theorem : Properties of Skew-Symmetric Matrices a. If A and B are skew-symmetric, $A + B$ is also skew-symmetric.

b. If A is a square matrix, then $A - A^T$ is skew-symmetric.

c. Any square matrix A can be represented as the sum of a symmetric matrix and a skew-symmetric matrix.

Proof. Let A and B be square matrices that the operations are defined.

a. (Exercise 3.2 40) Suppose that A and B are skew-symmetric, so that $A^T = -A$ and $B^T = -B$. Then

$$(A+B)^T = A^T + B^T = (-A) + (-B) = -(A+B)$$

Therefore $A+B$ is also skew-symmetric.

b. (Exercise 3.2 42)

$$(A - A^T)^T = A^T - (A^T)^T = A^T - A = -(A - A^T)$$

Therefore $A - A^T$ is also skew-symmetric.

c. (Exercise 3.2 43) By Theorem : Properties of Symmetric Matrices and Theorem : Properties of Skew-Symmetric Matrices, $A + A^T$ is symmetric and $A - A^T$ is skew-symmetric. Then representing A as

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$$

which is a sum of symmetric matrix and skew-symmetric matrix.

□

Definition. A **trace** of a $n \times n$ matrix A is denoted as $\mathrm{tr}(A)$ and defined as

$$\mathrm{tr}(A) = A_{11} + A_{22} + \cdots + A_{nn}$$

Extra Theorem : Properties of the Trace of Matrices
Let A and B be $n \times n$ matrices, and let k be a scalar.

a. $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$

b. $\mathrm{tr}(kA) = k\mathrm{tr}(A)$

c. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

d. $\mathrm{tr}(AA^T) = \sum_{1 \le i,j \le n} A_{ij}^2$

Proof. Let A and B be $n \times n$ matrices, and let k be a scalar.

a. (Exercise 3.2 44)

$$\begin{aligned} \mathrm{tr}(A+B) &= (A+B)_{11} + (A+B)_{22} + \cdots + (A+B)_{nn} \\ &= (A_{11} + B_{11}) + (A_{22} + B_{22}) + \cdots + (A_{nn} + B_{nn}) \\ &= (A_{11} + A_{22} + \cdots + A_{nn}) + (B_{11} + B_{22} + \cdots + B_{nn}) = \mathrm{tr}(A) + \mathrm{tr}(B) \end{aligned}$$

b. (Exercise 3.2 44)

$$\begin{aligned}\operatorname{tr}(kA) &= (kA)_{11} + (kA)_{22} + \cdots + (kA)_{nn} \\ &= k(A_{11} + A_{22} + \cdots + A_{nn}) = k\operatorname{tr}(A)\end{aligned}$$

c. (Exercise 3.2 45)

$$\begin{aligned}\operatorname{tr}(AB) &= (AB)_{11} + (AB)_{22} + \cdots + (AB)_{nn} \\ &= \sum_{1\le i,j\le n} A_{ij}B_{ji} \\ &= \sum_{1\le i,j\le n} B_{ij}A_{ji} \\ &= (BA)_{11} + (BA)_{22} + \cdots + (BA)_{nn} = \operatorname{tr}(BA)\end{aligned}$$

d. (Exercise 3.2 46)

$$\begin{aligned}\operatorname{tr}(AA^T) &= (AA^T)_{11} + (AA^T)_{22} + \cdots + (AA^T)_{nn} \\ &= \sum_{1\le i,j\le n} A_{ij}(A^T)_{ji} \\ &= \sum_{1\le i,j\le n} A_{ij}^2\end{aligned}$$

□

3.4 The Inverse of a Matrix

Definition. If A is an $n \times n$ matrix, an **inverse** of A is an $n \times n$ matrix, denoted as A^{-1}, such that

$$AA^{-1} = I_n \text{ and } A^{-1}A = I_n$$

If such A^{-1} exists, A is **invertible**.

Theorem 3.6
If A is invertible, then its inverse is unique.

Proof. Suppose that A' and A'' are both inveses of A. Then

$$AA' = I = A'A \text{ and } AA'' = I = A''A$$

Thus,

$$A' = A'I = A'(AA'') = (A'A)A'' = IA'' = A''$$

Therefore, the inverse of A is unique. □

Theorem 3.7
If A is an invertible $n \times n$ matrix, then the system of linear equations given by $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof. (i)

$$A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$$

Therefore the system is consistent.
(ii) Suppose that $\mathbf{x}'$ is another solution of $A\mathbf{x} = \mathbf{b}$. Then

$$A\mathbf{x}' = \mathbf{b} \Rightarrow A^{-1}(A\mathbf{x}') = A^{-1}\mathbf{b} \qquad \Rightarrow (A^{-1}A)\mathbf{x}' = \mathbf{x}' = A^{-1}\mathbf{b}$$

Thus, $\mathbf{x}'$ is the same solution as before. Therefore, the solution is unique. □

Theorem 3.8
If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then A is invertible if and only if $ad - bc \neq 0$ (a **determinant** of A, denoted as det A), and in that case

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

Proof. ($\Leftarrow$) If $ad - bc \neq 0$,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\left(\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\right) = \frac{1}{ad-bc}\begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\left(\frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\right)\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Therefore, A is invertible and $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$
($\Rightarrow$) Suppose that $ad - bc = 0$ and scalars x, y, z, w exist such that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} x & y \\ z & w \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Then

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} ax+bz & ay+bw \\ cx+dz & cy+dw \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Since $ad = bc$,

$$\begin{aligned} 1 = (ax+bz)(cy+dw) &= acxy + adxw + bcyz + bdwz \\ &= acxy + bcxw + adyz + bdwz = (ay+bw)(cx+dz) = 0 \end{aligned}$$

Thus, such x, y, z, w does not exist, and therefore A is not invertible. □

Theorem 3.9 : Properties of Invertible Matrices
If A, B are invertible matrices of the same size and c is a nonzero scalar,

a. A^{-1} is invertible and

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

b. cA is invertible and

$$(cA)^{-1} = \frac{1}{c}A^{-1}$$

c. AB is invertible and

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

d. A^T is invertible and

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

e. A^n is invertible for all nonnegative integers n and

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

Proof. Let A and B be invertible matrices of the same size and let c be a nonzero scalar.

a.

$$A^{-1}A = AA^{-1} = I$$

thus A is an inverse of A^{-1}.

b. (Exercise 3.3 14)

$$(cA)(\frac{1}{c}A^{-1}) = (c\frac{1}{c})(AA^{-1}) = I$$
$$(\frac{1}{c}A^{-1})(cA) = (\frac{1}{c}c)(A^{-1}A) = I$$

thus $\frac{1}{c}A^{-1}$ is an inverse of cA.

c.

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$$
$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I$$

thus $B^{-1}A^{-1}$ is an inverse of AB.

d. (Exercise 3.3 15)

$$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$$
$$(A^{-1})^TA^T = (AA^{-1})^T = I^T = I$$

thus $(A^{-1})^T$ is an inverse of A^T.

e. (i) $(A^0)^{-1} = I^{-1} = I = (A^{-1})^0$

(ii) Suppose that for arbitrary nonnegative integer n, $(A^n)^{-1} = (A^{-1})^n$. Then

$$(A^{n+1})^{-1} = (A^n A)^{-1} = A^{-1}(A^n)^{-1} = A^{-1}(A^{-1})^n = (A^{-1})^{n+1}$$

By (i) and (ii), $(A^n)^{-1} = (A^{-1})^n$ for any nonnegative integer n.

$\square$

Definition. An **elementary matrix** is any matrix that can be obatined by performing *single* elementary row operation on an identity matrix.

Types of Elementary Matrices

Let E be an $n \times n$ matrix. Then $E = \begin{bmatrix} \mathbf{E}_1^R \\ \mathbf{E}_2^R \\ \vdots \\ \mathbf{E}_n^R \end{bmatrix}$ is an elementary matrix if and only if E satisfies at least one of the followings:

a. There exist integers p, q such that $1 \leq p, q \leq n$ and

$$\mathbf{E}_i^R = \begin{cases} \mathbf{e}_q & \text{if } i = p \\ \mathbf{e}_p & \text{if } i = q \\ \mathbf{e}_i & \text{otherwise} \end{cases}$$

b. There exists an integer p and a nonzero scalar k such that $1 \leq p \leq n$ and

$$\mathbf{E}_i^R = \begin{cases} k\mathbf{e}_p & \text{if } i = p \\ \mathbf{e}_i & \text{otherwise} \end{cases}$$

c. There exists integers p, q and a scalar k such that $1 \leq p, q \leq n$ and

$$\mathbf{E}_i^R = \begin{cases} \mathbf{e}_p + k\mathbf{e}_q & \text{if } i = p \\ \mathbf{e}_i & \text{otherwise} \end{cases}$$

Theorem 3.10
If a elementary row operation converts I_n to E, then the same operation converts $n \times r$ matrix A to EA.

Proof. Let A be a $n \times r$ matrix, k be a nonzero scalar, p, q be integers between 1 and n, and E be an elementary matrix which corresponds to each type of elementary row operation.

$$A = \begin{bmatrix} \mathbf{A}_1^R \\ \vdots \\ \mathbf{A}_p^R \\ \vdots \\ \mathbf{A}_q^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix} \xrightarrow{R_p \leftrightarrow R_q} \begin{bmatrix} \mathbf{A}_1^R \\ \vdots \\ \mathbf{A}_q^R \\ \vdots \\ \mathbf{A}_p^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 A \\ \vdots \\ \mathbf{e}_q A \\ \vdots \\ \mathbf{e}_p A \\ \vdots \\ \mathbf{e}_n A \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_q \\ \vdots \\ \mathbf{e}_p \\ \vdots \\ \mathbf{e}_n \end{bmatrix} A = EA$$

$$A = \begin{bmatrix} \mathbf{A}_1^R \\ \vdots \\ \mathbf{A}_p^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix} \xrightarrow{kR_p} \begin{bmatrix} \mathbf{A}_1^R \\ \vdots \\ k\mathbf{A}_p^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 A \\ \vdots \\ k(\mathbf{e}_p A) \\ \vdots \\ \mathbf{e}_n A \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ k\mathbf{e}_p \\ \vdots \\ \mathbf{e}_n \end{bmatrix} A = EA$$

$$A = \begin{bmatrix} \mathbf{A}_1^R \\ \vdots \\ \mathbf{A}_p^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix} \xrightarrow{R_p + kR_q} \begin{bmatrix} \mathbf{A}_1^R \\ \vdots \\ \mathbf{A}_p^R + k\mathbf{A}_q^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 A \\ \vdots \\ \mathbf{e}_p A + k(\mathbf{e}_q A) \\ \vdots \\ \mathbf{e}_n A \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_p + k\mathbf{e}_q \\ \vdots \\ \mathbf{e}_n \end{bmatrix} A = EA$$

□

Theorem 3.11
All elementary matrices are invertible, and the inverse of an elementary matrix is also an elementary matrix.

Proof. Let E be an elementary matrix. Since there exists a reverse operation for the elementary row operations corresponds to E, the elementary matrix E' of the reverse operation will satisfy $EE' = E'E = I$ by Theorem 3.10. Therefore, the inverse of E exists and the inverse is also an elementary matrix. □

Theorem 3.12 : The Fundamental Theorem of Invertible Matrices - Version 1
Let A be an $n \times n$ matrix. Then the following propositions are equivalent;

a. A is invertible.

b. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.

c. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

d. The RREF of A is I_n.

e. A is a product of elementary matrices.

Proof. We give the proof for this theorem at Version 2. □

Theorem 3.13
Let A be a square matrix. If B is a square matrix such that either $AB = I$ or $BA = I$, then A is invertible and $A^{-1} = B$.

Proof. (i) Suppose that $BA = I$. Then the linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, since

$$\mathbf{x} = I\mathbf{x} = (BA)\mathbf{x} = B(A\mathbf{x}) = \mathbf{0}$$

Thus A is invertible by F.T.I.M. Then

$$BA = I \Rightarrow (BA)A^{-1} = A^{-1} \Rightarrow B(AA^{-1}) = B = A^{-1}$$

Therefore B is an inverse of A.
(ii) (Exercise 3.3 41) Suppose that $AB = I$. By (i), B is invertible and $B^{-1} = A$. Therefore, A is also invertible and $A^{-1} = (B^{-1})^{-1} = B$. □

Theorem 3.14
Let A be a square matrix. Then a sequence of elementary row operations which converts A to I also converts I to A^{-1}.

Proof. Let $E_1, E_2, \cdots, E_k$ be the elementary matrices which correspond to each step of elementary row operation converting A to I. Then by Theorem 3.10,

$$I = (E_k \cdots E_2 E_1)A$$

Then by Theorem 3.13,

$$A^{-1} = E_k \cdots E_2 E_1 = (E_k \cdots E_2 E_1)I$$

Therefore, the same steps of elementary row operation also converts I to A^{-1}. □

3.5 Subspaces, Basis, Dimension, and Rank

Definition. Let S be a set of vectors in $\mathbb{R}^n$. Then S is a **subspace** of $\mathbb{R}^n$if

a. $\mathbf{0} \in S$

b. $\mathbf{u}, \mathbf{v} \in S \Rightarrow \mathbf{u} + \mathbf{v} \in S$

c. $\mathbf{u} \in S \Rightarrow c\mathbf{u} \in S$ for any scalar c

Definition. Let A be an $n \times m$ matrix.
The **row space** of A, denoted by row(A), is a subspace of $\mathbb{R}^m$spanned by the rows of A.
The **column space** of A, denoted by col(A), is a subspace of $\mathbb{R}^n$spanned by the columns of A.
The **null space** of A, denoted by null(A), is a subspace of $\mathbb{R}^m$consisting of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$.

Theorem 3.15
For vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$in $\mathbb{R}^n$, span($\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$) is a subspace of $\mathbb{R}^n$.

Proof. Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$be vectors in $\mathbb{R}^n$, and let $S = \text{span}(\ \mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$.
(i) Since $0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_k = \mathbf{0}$, $\mathbf{0} \in S$.
(ii) Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in S. Then there exist scalars $c_1, c_2 \cdots, c_k$ and $d_1, d_2, \cdots, d_k$ such that

$$\begin{aligned}\mathbf{u} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k \\ \mathbf{v} &= d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_k\mathbf{v}_k\end{aligned}$$

Then

$$\begin{aligned}\mathbf{u} + \mathbf{v} &= (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) + (d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_k\mathbf{v}_k) \\ &= (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_k + d_k)\mathbf{v}_k\end{aligned}$$

Therefore $\mathbf{u} + \mathbf{v} \in S$.
(iii) Let $\mathbf{u}$ be a vector in S. Then there exist scalars $c_1, c_2, \cdots, c_k$ such that

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

Then for all scalar c,

$$\begin{aligned}c\mathbf{u} &= c(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) \\ &= cc_1\mathbf{v}_1 + cc_2\mathbf{v}_2 + \cdots + cc_k\mathbf{v}_k\end{aligned}$$

Therefore $c\mathbf{u} \in S$.
By (i), (ii), and (iii), S is a subspace of $\mathbb{R}^n$. $\square$

Theorem 3.16
If matrices A and B are row equivalent, then $\text{row}(A) = \text{row}(B)$.

Proof. Let A and B be $n \times m$ matrices. Since A and B are row equivalent, there exist $n \times n$ elementary matrices $E_1, E_2, \cdots, E_k$ which satisfies

$$B = (E_k \cdots E_2E_1)A$$

by Theorem 3.10. Let $C = E_k \cdots E_2E_1$.
Since $B = CA$,

$$B = \begin{bmatrix}\mathbf{C}_1^R \\ \vdots \\ \mathbf{C}_n^R\end{bmatrix} A = \begin{bmatrix}\mathbf{C}_1^R A \\ \vdots \\ \mathbf{C}_n^R A\end{bmatrix} = \begin{bmatrix}C_{11}\mathbf{A}_1^R + \cdots + C_{1n}\mathbf{A}_n^R \\ \vdots \\ C_{n1}\mathbf{A}_1^R + \cdots + C_{nn}\mathbf{A}_n^R\end{bmatrix}$$

Thus, rows of B are linear combinations of rows of A. Also, since C is invertible by Theorem 3.11, $A = C^{-1}B$, so rows of A are linear combinations of rows of B. Therefore, by the proposition at Exercise 2.3 21(b), $\text{span}(\mathbf{A}_1^R, \cdots, \mathbf{A}_n^R) = \text{span}(\mathbf{B}_1^R, \cdots, \mathbf{B}_n^R)$, that is $\text{row}(A) = \text{row}(B)$. $\square$

Theorem 3.17
Let A be an $n \times m$ matrix and let N be the set of solutions of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. Then N is a subspace of $\mathbb{R}^m$.

Proof. Let A be an $n \times m$ matrix and let N be the set of all solutions of $A\mathbf{x} = \mathbf{0}$.
(i) Since $A\mathbf{0} = \mathbf{0}$, $\mathbf{0} \in N$.
(ii) Let $\mathbf{u}, \mathbf{v}$ be vectors in N. Since $A\mathbf{u} = A\mathbf{v} = \mathbf{0}$, $A(\mathbf{u}+\mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0}$. Therefore, $\mathbf{u}+\mathbf{v} \in N$.
(iii) Let $\mathbf{u}$ be a vector in N. Since $A\mathbf{u} = \mathbf{0}$, for all scalar c, $A(c\mathbf{u}) = c(A\mathbf{u}) = c\mathbf{0} = \mathbf{0}$. Therefore, $c\mathbf{u} \in N$.
By (i), (ii), and (iii), N is a subspace of $\mathbb{R}^n$. □

Theorem 3.18
For any system of linear equations $A\mathbf{x} = \mathbf{b}$ with real coefficients, exactly one of the following is true;

a. The system has no solution.

b. The system has a unique solution.

c. The system has infinitely many solutions.

Proof. Consider the case when the system of linear equations has more than two solutions. Suppose that $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$are solutions of the linear system $A\mathbf{x} = \mathbf{b}$. Then $A(\mathbf{x}_2 - \mathbf{x}_1) = \mathbf{0}$, so $\mathbf{x}_2 - \mathbf{x}_1 \in$ null(A). By the definition of subspace, for any scalar c, $c(\mathbf{x}_2 - \mathbf{x}_1) \in$ null(A), so $A(c(\mathbf{x}_2 - \mathbf{x}_1)) = \mathbf{0}$. Since $A(\mathbf{x}_1 + c(\mathbf{x}_2 - \mathbf{x}_1)) = A\mathbf{x}_1 + A(c(\mathbf{x}_2 - \mathbf{x}_1)) = \mathbf{b} + \mathbf{0} = \mathbf{b}$, there exist infinitely many solutions. □

Definition. A **basis** for a subspace S in $\mathbb{R}^n$is a set of vectors in S that

a. spans S

b. is linearly independent.

How to Find Bases of Subspaces Related to Matrix

a. The basis of row(A) consists of the nonzero rows of RREF of the matrix.

b. The basis of col(A) consists of the columns of A where the leading 1s of A are located.

c. The basis of null(A) consists of the vectors which spans the solution of $A\mathbf{x} = \mathbf{0}$.

How to Find Subspaces Related to Matrix : Example
Find bases of row(A), col(A), and null(A) where

$$A = \begin{bmatrix} 1 & 1 & 3 & 1 & 6 \\ 2 & -1 & 0 & 1 & -1 \\ -3 & 2 & 1 & -2 & 1 \\ 4 & 1 & 6 & 1 & 3 \end{bmatrix}$$

Solution. The RREF of A is

$$R = \begin{bmatrix} 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The basis of row(A) is

$$\left\{\begin{bmatrix}1 & 0 & 1 & 0 & -1\end{bmatrix}, \begin{bmatrix}0 & 1 & 2 & 0 & 3\end{bmatrix}, \begin{bmatrix}0 & 0 & 0 & 1 & 4\end{bmatrix}\right\}$$

The basis of col(A) consists of the first, second, and fourth columns of A, so the basis of col(A) is

$$\left\{\begin{bmatrix}1\\2\\-3\\4\end{bmatrix}, \begin{bmatrix}1\\-1\\2\\1\end{bmatrix}, \begin{bmatrix}1\\1\\-2\\1\end{bmatrix}\right\}$$

The solution of $A\mathbf{x} = \mathbf{0}$ is

$$\mathbf{x} = s\begin{bmatrix}-1\\-2\\1\\0\\0\end{bmatrix} + t\begin{bmatrix}1\\-3\\0\\-4\\1\end{bmatrix} \; (s, t \in \mathbb{R})$$

thus the basis of null(A) is

$$\left\{\begin{bmatrix}-1\\-2\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\-3\\0\\-4\\1\end{bmatrix}\right\}$$

Theorem 3.19
Let S be a subspace of $\mathbb{R}^n$. Then any two bases for S have the same number of vectors.

Proof. Let $\mathcal{B}_1 = \{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ and $\mathcal{B}_2 = \{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ be bases for subspace S of $\mathbb{R}^n$.
Suppose that $k > m$. Since $\mathcal{B}_1 \subset S$ and $\mathcal{B}_2$ is a basis for S, there exist scalars a_{ij} such that

$$\begin{aligned}\mathbf{u}_1 &= a_{11}\mathbf{v}_1 + \cdots + a_{1m}\mathbf{v}_m\\ &\vdots\\ \mathbf{u}_k &= a_{k1}\mathbf{v}_1 + \cdots + a_{km}\mathbf{v}_m\end{aligned}$$

Since $\mathcal{B}_2$ is a basis for S, $\mathbf{v}_1, \cdots, \mathbf{v}_m$ are linearly independent. So the linear system

$$\begin{aligned}&c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k\\ &= (c_1a_{11} + \cdots + c_ka_{k1})\mathbf{v}_1 + (c_1a_{12} + \cdots + c_ka_{k2})\mathbf{v}_2 + \cdots + (c_1a_{1m} + \cdots + c_ka_{km})\mathbf{v}_m = \mathbf{0}\end{aligned}$$

has only the trivial solution. Thus, the homogeneous linear system with augmented matrix

$$\left[\begin{array}{ccc|c}a_{11} & \cdots & a_{k1} & 0\\ \vdots & & \vdots & \vdots\\ a_{1m} & \cdots & a_{km} & 0\end{array}\right]$$

is consistent, and the system has infinitely many solutions because the size of the matrix is $m \times k$. There exists a nontrivial solution of this system, so there exist scalars $c_1, \cdots, c_k$ such that at least one of $c_1, \cdots, c_k$ are nonzero, and $c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$. Therefore, $\mathbf{u}_1, \cdots, \mathbf{u}_k$ are linearly dependent, which contradicts with the assumption.
Similarly, in case of $k < m$ also comes up to contradiction. Therefore, $k = m$. □

Definition. If S is a subspace of $\mathbb{R}^n$, then the number of vectors in a basis for S is the **dimension** of S, denoted by dim S.

Theorem 3.20
For any matrix A,

$$\dim \operatorname{row}(A) = \dim \operatorname{col}(A)$$

Proof. The number of nonzero rows and the number of leading 1s are the same. Since the nonzero rows form the basis for row(A), and the columns of A with the leading 1s form the basis for col(A), dim row(A) = dim col(A). □

Definition. The **rank** of a matrix A, denoted by rank(A), is the dimension of its row and column spaces.

Theorem 3.21
For any matrix A,

$$\operatorname{rank}(A) = \operatorname{rank}(A^T)$$

Proof.

$$\operatorname{rank}(A) = \dim \operatorname{row}(A) = \dim \operatorname{col}(A) = \dim \operatorname{row}(A^T) = \operatorname{rank}(A^T)$$

□

Definition. The **nullity** of a matrix A, denoted by nullity(A), is the dimension of its null space.

Theorem 3.22 : The Rank Theorem
If A is an $m \times n$ matrix, then

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n$$

Proof. By Theorem 2.2. □

Theorem 3.23 : The Fundamental Theorem of Invertible Matrices : Version 2
Let A be an $n \times n$ matrix. The following propositions are equivalent;

a. A is invertible.

b. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.

c. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

d. The RREF of A is I_n.

e. A is a product of elementary matrices.

f. $\operatorname{rank}(A) = n$

g. $\operatorname{nullity}(A) = 0$

h. The columns of A are linearly independent.

i. The columns of A span $\mathbb{R}^n$.

j. The columns of A form a basis for $\mathbb{R}^n$.

k. The rows of A are linearly independent.

l. The rows of A span $\mathbb{R}^n$.

m. The rows of A form a basis for $\mathbb{R}^n$.

Proof. Let A be an $n \times n$ matrix.
(a $\Rightarrow$ b) Theorem 3.7

(b $\Rightarrow$ c) Since $\mathbf{x} = \mathbf{0}$ is a solution of $A\mathbf{x} = \mathbf{0}$, $\mathbf{0}$ is the unique solution of the system. Therefore, the system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(c $\Rightarrow$ d) The corresponding augmented martrix of the homogeneous system with unique solution $\mathbf{0}$ is $[I_n|\mathbf{0}]$. Since $A\mathbf{x} = \mathbf{0}$ has only the unique solution $\mathbf{0}$, A is row equivalent with I_n, therefore RREF of A is I_n.

(d $\Rightarrow$ e) Since A and I_n are row equivalent, there exists a sequence of elementary row operations which converts I_n to A. Let $E_1, E_2, \cdots, E_k$ be the corresponding elementary matrices to each steps fo elementary row operations. Then by Theorem 3.10,

$$A = E_k \cdots E_2 E_1 I_n = E_k \cdots E_2 E_1$$

Therefore A is a product of elementary matrices.

(e $\Rightarrow$ a) Since all elementary matrices are invertible, A, which is the product of elementary matrices, is also invertible.

(d $\Leftrightarrow$ f) If the RREF of A is I_n, since $\text{rank}(I_n) = n$, by Theorem 3.16 $\text{rank}(A) = \text{rank}(I_n) = n$. Conversely, if $\text{rank}(A) = n$, the RREF of A has n leading 1s so the RREF should be I_n.

(f $\Leftrightarrow$ g) Rank Theorem.

(c $\Leftrightarrow$ h) Let A be an $n \times m$ matrix. Since $A\mathbf{x} = \mathbf{0}$ has only the trivial solution,

$$c_1\mathbf{A}_1^C + \cdots + c_m\mathbf{A}_m^C = \mathbf{0}$$

is true only if $c_1 = \cdots = c_m = 0$. Therefore, columns of A are linearly independent.
Conversely, if columns of A are linearly independent, the system $c_1\mathbf{A}_1^C + \cdots + c_m\mathbf{A}_m^C = \mathbf{0}$ does not have a nontrivial solution, so $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(b $\Rightarrow$ i) Since for any $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ is consistent, $\mathbf{b}$ is a linear combination of colmuns of A. Therefore, the columns of A span $\mathbb{R}^n$.

(i $\Rightarrow$ f) $\text{col}(A) = \mathbb{R}^n$, so $\text{rank}(A) = \dim \text{col}(A) = \dim \mathbb{R}^n = n$.

(h and i $\Leftrightarrow$ j) Definition of basis for a subspace of $\mathbb{R}^n$.

By Theorem 3.21, $\text{rank}(A^T) = \text{rank}(A)$. Therefore, the proposition related to columns of A can be applied to rows of A, so (k), (l), and (m) are also equivalent.

$\square$

Theorem 3.24
Let A be an $n \times m$ matrix. Then

a. $\text{rank}(A^T A) = \text{rank}(A)$

b. The $n \times n$ matrix $A^T A$ is invertible if and only if $\text{rank}(A) = n$.

Proof. a. Let A be an $n \times m$ matrix.

(i) Suppose $\mathbf{x}$ is a solution for $A\mathbf{x} = \mathbf{0}$. Then $A^T A\mathbf{x} = \mathbf{0}$. Therefore $\text{null}(A) \subset \text{null}(A^T A)$.

(ii) Suppose $\mathbf{x}$ is a solution for $A^T A\mathbf{x} = \mathbf{0}$. Then

$$(A\mathbf{x}) \cdot (A\mathbf{x}) = (A\mathbf{x})^T (A\mathbf{x}) = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T \mathbf{0} = 0$$

Therefore, $A\mathbf{x} = \mathbf{0}$, so $\text{null}(A^T A) \subset \text{null}(A)$.

By (i), (ii), $\text{null}(A) = \text{null}(A^T A)$, so $\text{nullity}(A) = \text{nullity}(A^T A)$, thus $\text{rank}(A) = \text{rank}(A^T A)$ by Rank Theorem.

b. Let A be an $n \times n$ matrix. Since $\text{rank}(A) = \text{rank}(A^T A)$, by F.T.I.M, $A^T A$ is invertible if and only if $\text{rank}(A^T A) = \text{rank}(A) = n$.

$\square$

Theorem 3.25
Let S be a subspace of $\mathbb{R}^n$and let $\mathcal{B} = \{\ \mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be a basis for S. For every vector $\mathbf{v}$ in S, there is exactly one way to represent $\mathbf{v}$ as a linear combination of the basis vectors in $\mathcal{B}$.

Proof. (Existence) Since $\mathcal{B}$ is a basis of S, any vector $\mathbf{v} \in S$ is a linear combination of vectors in $\mathcal{B}$.

(Uniqueness) Suppose that there are more than two ways to represent $\mathbf{v} \in S$ as a linear combination of $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$. That is, the linear system with augmented matrix $[A|\mathbf{v}] = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_k & | & \mathbf{v}\end{bmatrix}$ has more than two solutions. Let those two solutions be $\mathbf{x}_1$ and $\mathbf{x}_2$, then $A\mathbf{x}_1 = \mathbf{v}$ and $A\mathbf{x}_2 = \mathbf{v}_2$. Since $A(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$, the homogeneous system $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, which contradicts with F.T.I.M. Therefore, representation of $\mathbf{v}$ as a linear combination of $\mathcal{B}$ is unique. $\square$

Definition. Let S be a subspace of $\mathbb{R}^n$and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ be a basis for S. Let $\mathbf{v}$ be a vector in S, and $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$. Then $c_1, c_2, \cdots, c_k$ are the **coordinates of v with respect to $\mathcal{B}$**, and the column vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix}$$

is called the **coordinate vector of v with respect ot $\mathcal{B}$**.

Extra Theorem : Inequality on subspaces' dimension
Note. This lemma will be used in this subsection as 'Lemma'.
Let S_1, S_2 be subspaces of $\mathbb{R}^n$. If $S_1 \subset S_2$, then $\dim(S_1) \leq \dim(S_2)$

Proof. Let $\mathcal{B}_1 = \{\mathbf{u}_1, \mathbf{u}_2, \cdots \mathbf{u}_k\}$ be a basis for S_1, $\mathcal{B}_2 = \{\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_m\}$ be a basis for S_2 . Suppose $k > m$. Since $\mathcal{B} \subset S_1 \subset S_2$, there exist a_{ij} $(1 \leq i \leq k,\ 1 \leq j \leq m)$ such that

$$\begin{aligned} \mathbf{u}_1 &= a_{11}\mathbf{v}_1 + \cdots + a_{1m}\mathbf{v}_m \\ &\vdots \\ \mathbf{u}_k &= a_{k1}\mathbf{v}_1 + \cdots + a_{km}\mathbf{v}_m \end{aligned}$$

Then $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots c_k\mathbf{u}_k = (c_1a_{11} + \cdots + c_ka_{k1})\,\mathbf{v}_1 + \cdots (c_1a_{1m} + \cdots + c_ka_{km})\,\mathbf{v}_m = \mathbf{0}$ has only the trivial solution $c_1a_{11} + \cdots + c_ka_{k1} \cdots c_1a_{1m} + \cdots + c_ka_{km} = 0$ since $\mathcal{B}_2$ is linearly independent. $\left[\begin{array}{ccc|c} a_{11} & \cdots & a_{k1} & 0 \\ \vdots & & \vdots & \vdots \\ a_{1m} & \cdots & a_{km} & 0 \end{array}\right]$ has a nontrivial solution since rank $\left(\left[\begin{array}{ccc|c} a_{11} & \cdots & a_{k1} & 0 \\ \vdots & & \vdots & \vdots \\ a_{1m} & \cdots & a_{km} & 0 \end{array}\right]\right) \leq m < k$.
The number of free variables is $k - \text{rank}\left(\left[\begin{array}{ccc|c} a_{11} & \cdots & a_{k1} & 0 \\ \vdots & & \vdots & \vdots \\ a_{1m} & \cdots & a_{km} & 0 \end{array}\right]\right) > 0$.
Therefore, there exist scalars $c_1, \cdots, c_k$ such that at least one of $c_1, \cdots, c_k$ are nonzero and $c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$, so $\mathbf{u}_1, \cdots, \mathbf{u}_k$ are linearly dependent.
However, since $\mathcal{B}_1$ is a basis, $\mathbf{u}_1, \cdots, \mathbf{u}_k$ should be linearly independent : contradiction. $\therefore k \leq m$, that is, $\dim S_1 \leq \dim S_2$. □

3.6 Solutions of Exercises Worthy to Solve

a. **Exercise 3.5 57**

Proof. Let A be an $m \times n$ matrix and let R be the RREF of A. Then for any $\mathbf{v} \in \text{null}(A)$, $R\mathbf{v} = \begin{bmatrix} \mathbf{R}_1^R \cdot \mathbf{v} \\ \vdots \\ \mathbf{R}_n^R \cdot \mathbf{v} \end{bmatrix} = \mathbf{0}$. $\therefore \mathbf{R}_i^R \cdot \mathbf{v} = \mathbf{0}$ for $1 \leq i \leq n$. For every $\mathbf{n} \in \text{row}(A)$, $\exists c_1, c_2, \cdots c_n$ such that $\mathbf{u} = c_1\mathbf{R}_1^R + c_2\mathbf{R}_2^R + \cdots c_n\mathbf{R}_n^R$ since $\{\mathbf{R}_1^R, \cdots \mathbf{R}_n^R\}$ form a basis for row(A), then

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= \left(c_1\mathbf{R}_1^R + c_2\mathbf{R}_2^R + \cdots c_n\mathbf{R}_n^R\right) \cdot \mathbf{v} \\ &= c_1\mathbf{R}_1^R \cdot \mathbf{v} + c_2\mathbf{R}_2^R \cdot \mathbf{v} + \cdots c_n\mathbf{R}_n^R \cdot \mathbf{v} = 0 \end{aligned}$$

Therefore, for any $\mathbf{u} \in \text{row}(A)$ and $\mathbf{v} \in \text{null}(A)$, $\mathbf{u} \cdot \mathbf{v} = 0$ so $\mathbf{u}$ and $\mathbf{v}$ are orthogonal. □

b. **Exercise 3.5 58**

Proof. Let A, B be $n \times n$ matrices with rank n. Then by F.T.I.M, $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ has only the trivial solution, and $\exists A^{-1}$, $\exists B^{-1}$.
Suppose that $AB\mathbf{x} = \mathbf{0}$ has a nontrivial solution $\mathbf{x}_1 \neq \mathbf{0}$. Then $AB\mathbf{x}_1 = \mathbf{0}$, so $A^{-1}AB\mathbf{x}_1 = B\mathbf{x}_1 = \mathbf{0}$, so the system $B\mathbf{x} = \mathbf{0}$ has a nontrivial solution. Since this contradict with our assumption, $(AB)\mathbf{x} = \mathbf{0}$ has only the trivial solution.
By F.T.I.M, rank of AB is n. □

Another solution :

Proof. By F.T.I.M, $\exists A^{-1}$ and $\exists B^{-1}$. Therefore, $AB^{-1} = B^{-1}A^{-1}$. By F.T.I.M, rank$(AB) = n$. □

c. **Exercise 3.5 59**

Proof. **(a)** For $n \times m$ matrix A and $m \times r$ matrix B,

$$AB = \begin{bmatrix} \mathbf{A}_1^R \\ \vdots \\ \mathbf{A}_n^R \end{bmatrix} B = \begin{bmatrix} \mathbf{A}_1^R B \\ \vdots \\ \mathbf{A}_n^R B \end{bmatrix} = \left[A_{11}\mathbf{B}_1^R + \cdots + A_{1m}\mathbf{B}_m^R \vdots A_{n1}\mathbf{B}_1^R + \cdots + A_{nm}\mathbf{B}_m^R \right].$$

Since rows of AB are linear combinations of $\mathbf{B}_1^R, \cdots, \mathbf{B}_m^R$, row$(AB) \subset$ row(B).
$\therefore$ By Lemma,rank$(AB) \leq$ rank(B). **(b)** In case when $A = O$. □

Another solution of **(a)** :

Proof. For $n \times m$ matrix A and $m \times r$ matrix B,

$$\text{rank}(B) = r - \text{nullity}(B)$$
$$\text{rank}(AB) = r - \text{nullity}(AB)$$

$\forall \mathbf{x} \in \text{null}(B)$, $(AB)\mathbf{x} = A(B\mathbf{x}) = A\mathbf{0} = \mathbf{0}$. *i.e.* $\mathbf{x} \in \text{null}(AB)$. $\therefore$ null$(B) \subset$ null(AB).
By Lemma, nullity$(B) \leq$ nullity(AB). $\therefore$ rank$(B) \geq$ rank(AB). □

d. **Exercise 3.5 60**

Proof. **(a)** For $n \times m$ matrix A and $m \times r$ matrix B, $AB = A\begin{bmatrix} \mathbf{B}_1^C & \mathbf{B}_2^C & \cdots & \mathbf{B}_r^C \end{bmatrix} = \begin{bmatrix} A\mathbf{B}_1^C & A\mathbf{B}_2^C & \cdots & A\mathbf{B}_r^C \end{bmatrix} \left[\left(B_{11}\mathbf{A}_1^C + \cdots + B_{m1}\mathbf{A}_m^C \right) \quad \cdots \quad \left(B_{1r}\mathbf{A}_1^C + \cdots + B_{mr}\mathbf{A}_m^C \right) \right]$ Since columns of AB are linear combinations of $\mathbf{A}_1^C \cdots \mathbf{A}_m^C$, col$(AB) \subset$ col(A).
$\therefore$ By Lemma, rank$(AB) \leq$ rank(B). **(b)** In case when $B = O$ □

Another solution of **(a)** :

Proof.

$$\begin{aligned}\operatorname{rank}(AB) &= \operatorname{rank}(B^TA^T) && \text{(Theorem 3.21)}\\ &\le \operatorname{rank}(A^T) && \text{(Exercise 3.5 59)}\\ &= \operatorname{rank}(A) && \text{(Theorem 3.21)}\end{aligned}$$

□

e. **Exercise 3.5 61**

Proof. **(a)** By Exercise 3.5 59, $\operatorname{rank}(A) = \operatorname{rank}\left(U^{-1}(UA)\right) \le \operatorname{rank}(UA) \le \operatorname{rank}(A)$. $\therefore \operatorname{rank}(UA) = \operatorname{rank}(A)$.

(b) By Exercise 3.5 60, $\operatorname{rank}(A) = \operatorname{rank}\left((AV)V^{-1}\right) \le \operatorname{rank}(AV) \le \operatorname{rank}(A)$. $\therefore \operatorname{rank}(AV) = \operatorname{rank}(A)$. □

f. **Exercise 3.5 62**

Proof. ($\Rightarrow$) Let A be an $m\times n$ matrix with rank 1. Then $\dim(\operatorname{row}(A)) = \dim(\operatorname{col}(A)) = 1$. For $\mathbf{v} \in \mathbb{R}^m$, suppose that $\{\mathbf{v}^T\}$ is a basis for row(A). Then $\exists c_1, c_2, \cdots, c_m$ such that $\mathbf{A}_i^R = c_i\mathbf{u}^T$.
Then $A = \begin{bmatrix}\mathbf{A}_1^R\\ \vdots \\ \mathbf{A}_m^R\end{bmatrix} = \begin{bmatrix}c_1\mathbf{v}^T\\ \vdots \\ c_m\mathbf{v}^T\end{bmatrix} = \begin{bmatrix}c_1\\ \vdots \\ c_m\end{bmatrix}\mathbf{v}^T$. Let $\mathbf{u} = \begin{bmatrix}c_1\\ \vdots \\ c_m\end{bmatrix}$, then $\mathbf{u}\mathbf{v}^T = A$

($\Leftarrow$) Suppose that for $\mathbf{u} = \begin{bmatrix}u_1\\ \vdots \\ u_m\end{bmatrix} \in \mathbb{R}^m$ and $\mathbf{v} = \begin{bmatrix}v_1\\ \vdots \\ v_m\end{bmatrix} \in \mathbb{R}^m$, $\mathbf{u}\mathbf{v}^T = A$. [1]
$A = \mathbf{u}\mathbf{v}^T = \mathbf{u}\begin{bmatrix}v_1 & \cdots & v_n\end{bmatrix} = \begin{bmatrix}v_1\mathbf{u} & \cdots & v_n\mathbf{u}\end{bmatrix}$
Then $\{\mathbf{u}\}$ forms a basis for col(A), since the columns of A are linear combination of $\mathbf{u}$.
$\therefore \dim(\operatorname{col}(A)) = \operatorname{rank}(A) = 1$ □

g. **Exercise 3.5 63**

Proof. Let A be an $m\times n$ matrix. Then $A = AI_n = \begin{bmatrix}\mathbf{A}_1^C & \cdots & \mathbf{A}_n^C\end{bmatrix}\begin{bmatrix}\mathbf{e}_1\\ \vdots \\ \mathbf{e}_n\end{bmatrix} = \mathbf{A}_1^C\mathbf{e}_1 + \cdots + \mathbf{A}_n^C\mathbf{e}_n$. Since $\mathbf{A}_1^C \in \mathbb{R}^m$ and $\mathbf{e}_i \in \mathbb{R}^n$, ranks of $\mathbf{A}_i^C\mathbf{e}_i$ are all 1. □

h. **Exercise 3.5 64**

Proof. For $m\times n$ matrices A, B, let $\mathcal{B}_1$, $\mathcal{B}_2$ be bases for row(A),row(B). Then $(\mathbf{A}{+}\mathbf{B})_i^R = \mathbf{A}_i^R + \mathbf{B}_i^R$, so $(\mathbf{A}{+}\mathbf{B})_i^R$ are linear combinations of vectors in $\mathcal{B}_1 \cup \mathcal{B}_2$.
Thus, there exists a subset $C \subset B$ which forms a basis for row$(A+B)$.
$\therefore \operatorname{rank}(A+B) = |C| \le |\mathcal{B}_1 \cup \mathcal{B}_2| \le |\mathcal{B}_1| + |\mathcal{B}_2| = \operatorname{rank}(A) + \operatorname{rank}(B)$ □

[1] 사실 문제가 틀림. Nonzero vectors **u**, **v** 여야 함.

i. **Exercise 3.5 65**

Proof. $AA = A\left[\mathbf{A}_1^C \cdots \mathbf{A}_n^C\right] = \left[A\mathbf{A}_1^C \cdots A\mathbf{A}_n^C\right] = 0$. $\therefore$ $\mathbf{A}_1^C \cdots \mathbf{A}_n^C$ are solutions of the system $A\mathbf{x} = \mathbf{0}$. $\mathbf{A}_1^C \cdots \mathbf{A}_n^C \in \text{null}(A)$. and by definition of subspaces, $\text{col}(A) \subset \text{null}(A)$.
$\therefore$ By Lemma, $\dim(col(A)) = \text{rank}(A) \leq \text{null}(A)$. $\therefore$ $\text{null}(A) \geq n/2$ ☐

j. **Exercise 3.5 66**

Proof. **(a)** $\left(\mathbf{x}^T A\mathbf{x}\right)^T = \mathbf{x}^T A\mathbf{x} = -\mathbf{x}^T A\mathbf{x}$. $\therefore$ $\mathbf{x}^T A\mathbf{x}$ is also skew-symmetric. Since $\mathbf{x}^T A\mathbf{x}$ is a 1×1 matrix and all main diagonal entries are zero in skew-symmetric matrices, $\mathbf{x}^T A\mathbf{x} = 0$. **(b)** Let $\mathbf{x} \in \mathbb{R}^n$ such that $(I_n + A)\,\mathbf{x} = \mathbf{0}$. Then $\mathbf{x}^T\,(I_n + A)\,\mathbf{x} = \mathbf{x}^T I_n\mathbf{x} + \mathbf{x}^T A\mathbf{x} = \mathbf{x}^T\mathbf{x} + 0 = \mathbf{x}^T\mathbf{x} = 0$. Since $\mathbf{x}^T\mathbf{x} = 0$, $\mathbf{x} = \mathbf{0}$.
$\therefore$ $\text{null}(I_n + A) = \{\mathbf{0}\}$. Since $\text{rank}(I_n + A) = n$, by F.T.I.M, $A + I_n$ is invertible. ☐

Appendix A

Cautions on Exam

A.1 Notations

Wrong	Right	Explanation
$2 \cdot 3 = 6$	$2 \times 3 = 6$	$\cdot$ for dot product is only valid for vectors.
$2 \cdot \mathbf{v}$	$2\mathbf{v}$	
$\begin{bmatrix} 1 & 3 & 4 \end{bmatrix}$	$\begin{bmatrix} 1, & 3, & 4 \end{bmatrix}$	When writing row vectors, commas are necessary.

A.2 Description

- Free variables like s, t and u must be indicated that they are arbitrary real number. *e.g.* $s, t \in \mathbb{R}$
- When proving span $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \mathbb{R}^3$, you must prove each side, not only one.
- $A = B$ means not only same entries, but also same size.
- Write as '적어도 하나는 0이 아닌' instead of '모두 0은 아닌' or else.
- Abbreviation : only 4 things are allowed : 'REF', 'RREF', 'EMO', 'F.T.I.M.'. Each of them stands for (Reduced) Row Echelon Form, Elementary Matrix Operation, Fundamental Theorems of Invertible Matrices.