# Enoch's Notebook: An Introduction to Linear Algebra

# Enoch's Notebook: An Introduction to Linear Algebra

A Beginner's Guide to Linear Algebra

*Enoch Yu*

ISBN: 9798192810682

# Contents

# Preface

Linear algebra is a foundational branch of mathematics that primarily investigates in linear equations, vectors, and their relations. The application of linear algebra is so immense that numerous fields require foundational understanding of linear algebra. For instance, it would not be possible to gain profound understandings in machine learning and data science without the basics of linear algebra. This collection of notes discuss the essence of linear algebra in perspective of teaching.

## §Motivation

One of the most effective way to learn, at least for me, is to teach. During the summer of 2026, I was enrolled in the course XM511 from Stanford Pre-Collegiate University-Level Online Math, instructed by Dr. Margarita Kanarsky [Kan26]. During the lectures, I thought it would be a great idea to write notes on linear algebra just as I wrote the notes for machine learning. Writing the teaching-style notes for machine learning allowed me to differentiate what I truly know.

That being said, my notes do not represent the course. I have included my understandings the supporting textbook [Ric07] from numerous external sources and restructured the notes itself to adhere to the Permission to Use Materials. Therefore, any errors in these notes are my responsibility.

## §About the Author

Enoch is an ardent student who retains keen interest in pure math, machine learning, and quantitative finance. His previous accomplishments includes honorable mention from Korean Mathematical Olympiad ( 300 in South Korea) and numerous awards from AMC. He currently dedicates his time on making YouTube videos on various math topics and Nanum Learn. More on Enoch can be found in his personal website.

# Note 1

# Vectors and Vector Spaces

If you have backgrounds on physics or computer science, you are probably familiar with what vectors are. In physics classes, we learn that vectors are quantities with magnitude and direction. In computer science, vectors are introduced as lists or arrays of data. Both definitions capture different aspects of what vectors are. Throughout the notes, we will implement both understandings. In this note, we will discuss fundamental terms for vectors and vector spaces before implementing matrices in vector spaces. Also, most of the heavy abstractions are done in modern algebra, so it will be an introduction.

## §1.1 Basic Terms and Notations

First, let's get started by building from definitions.

**Definition 1.1.1**

**Vectors** are mathematical objects, or quantities, with direction and magnitude. **Scalars** are real or complex numbers that do not have direction.

A Vector is a more general term that includes different objects like matrices and polynomials. Before we discuss about vector spaces, let's clarify common notations.

**Definition 1.1.2**

$\mathbb{R}$ and $\mathbb{C}$ denote the set of real and complex numbers respectively. Epsilon symbols $\in$ and $\notin$ are used to denote whether an element is in the set or not.

Examples of using such notations would be $\pi \in \mathbb{C}$ and $i \notin \mathbb{R}$.

**Definition 1.1.3**

A **vector space** is a set of vectors $\mathbb{V}$ and scalars with vector addition and scalar multiplication that adheres to the following ten properties, or **axioms for a vector space**.

Vector Addition

A1. **Closure:** $\forall\, u, v \in \mathbb{V}, u \oplus v \in \mathbb{V}$.

A2. **Commutativity:** $\forall\, u, v \in \mathbb{V}, u \oplus v = v \oplus u$.

A3. **Associativity:** $\forall\, u, v, w \in \mathbb{V}, u \oplus (v \oplus w) = (u \oplus v) \oplus w$.

A4. **Additive Identity Element:** $\exists\, 0 \in \mathbb{V}$ such that $\forall\, u \in \mathbb{V}, u \oplus 0 = u$, also known as the **zero vector**.

A5. **Additive Inverses:** $\forall\, u \in \mathbb{V}, \exists -u \in \mathbb{V}$ such that $u \oplus (-u) = 0$.

Scalar Multiplication

S1. **Closure:** $\forall\, u \in \mathbb{V}$ and scalars $\alpha$, $\alpha \odot u \in \mathbb{V}$.

S2. **Associativity:** $\forall\, u \in \mathbb{V}$ and scalars $\alpha, \beta$, $\alpha \odot (\beta \odot u) = (\alpha \cdot \beta) \odot u$.

S3. **Identity Element:** $\forall\, u \in \mathbb{V}, 1 \odot u = u$.

S4. **Distributivity I:** $\forall\, u \in \mathbb{V}$ and scalars $\alpha, \beta$, $(\alpha + \beta) \odot u = \alpha \odot u \oplus \beta \odot u$.

S5. **Distributivity II:** $\forall\, u, v \in \mathbb{V}$ and scalars $\alpha$, $\alpha \odot (u \oplus v) = \alpha \odot u \oplus \alpha \odot v$.

Side note here is that for this section of the note, we will be using the binary operation symbols $\oplus$ and $\odot$ to denote vector addition and scalar multiplication. However, from the next section, we will be using more standard notations like $v + u$ and $\lambda v$ for vector addition and scalar multiplication. The reason I decided to use this is that this is the way I first learned and it helped me to emphasize axioms for vector spaces.

Based on the scalars that we have here, we could indicate vector spaces with the following terms.

**Definition 1.1.4**

If the set of scalars is in $\mathbb{R}$, then the vector space is known as a **vector space over** $\mathbb{R}$, or a **real vector space**. Similarly, if the scalars are in $\mathbb{C}$, then the vector space is denoted as a **vector space over** $\mathbb{C}$, or a **complex vector space**.

One thing to note is that there is a concept called **scalar fields**; however, we will not go too deep into such fields. Before we move on to different concepts in vector spaces, I wanted to discuss about tuples and dimensions.

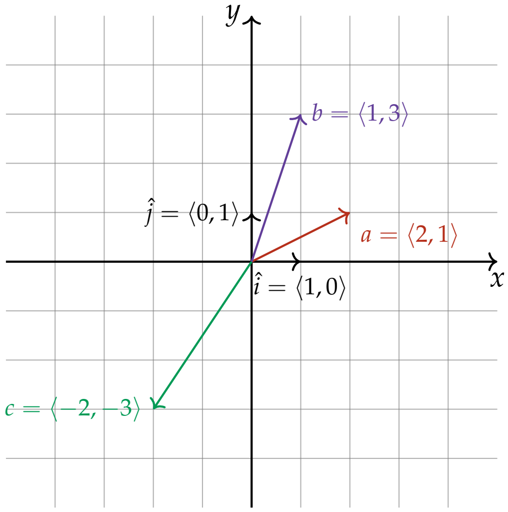
**Definition 1.1.5**

An $n$**-tuple** is an ordered array, or list, of $n$ real numbers. For an $n$-tuple, its dimension is said to be $n$-dimensional and the set of such real tuples is $n$**-space**, denoted as $\mathbb{R}^n$.

For instance, the following 4-tuple is in 4-space.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \in \mathbb{R}^4$$

With tuples in mind, let's take a look at how we can represent 2-tuples, 3-tuples, and vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$. The vector space $\mathbb{R}^2$ can simply be conceived as 2-dimensional space, or plane. Also, in linear algebra, we generally don't move the tail of a vector from the origin as we do in introductory physics classes.




From the graph above, there are two vectors $\hat{i}$ and $\hat{j}$. They are called "$i$ hat" and "$j$ hat" respectively.

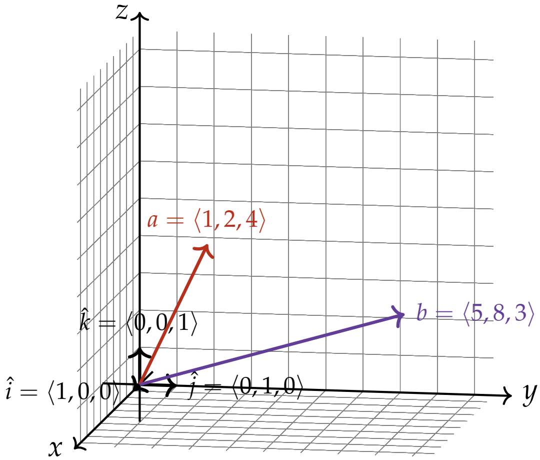
**Definition 1.1.6**

**Unit Vectors** are vectors with magnitude 1 that are used to specify direction.

Here, $\hat{i}$ and $\hat{j}$ denote the unit vectors, and they are the standard symbol. If you wondered if unit vectors are used to construct vectors, you are right! In our examples above, we can write each vector as the following.

$$\begin{aligned}
a &= \langle 2, 1\rangle = 2\hat{i} + \hat{j} \\
b &= \langle 1, 3\rangle = 1\hat{i} + 3\hat{j} \\
c &= \langle -2, -3\rangle = -2\hat{i} - 3\hat{j}
\end{aligned}$$

Now, let's take a look at vector space $\mathbb{R}^3$. There are two ways of graphing 3-dimensional space: right-handed system and left-handed system. However, we will be using right-hand system throughout the notes.




Here, $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ are the unit vectors. Going back to the definition of vector spaces, we can write the following equations for $u = \langle x_1, y_1, z_1 \rangle, v = \langle x_2, y_2, z_2 \rangle \in \mathbb{R}^3$ and $\alpha \in \mathbb{R}$.

$$u \oplus v = \langle x_1, y_1, z_1 \rangle + \langle x_2, y_2, z_2 \rangle = \langle x_1 + x_2, y_1 + y_2, z_1 + z_2 \rangle$$
$$\alpha \odot u = \alpha \odot \langle x_1, y_1, z_1 \rangle = \langle \alpha x_1, \alpha y_1, \alpha z_1 \rangle$$

Now that we saw some geometrical interpretation of vector spaces, let's discuss more information with some problems.

## §1.2 Fundamental Properties

Let's first start by solving some problems.

**Exercise 1.2.1**

Identify if the set $S = \{(x, y) \in \mathbb{R}^2 \mid x + y = 5\}$ is a vector space.

**Solution.**

We could see if the set satisfies all axioms for a vector space. However, just by looking at the condition $x + y = 5$, we know that there cannot exist a zero vector since $0 + 0 \neq 5$. Because the set violates

A4, the set is not a vector space.

**Exercise 1.2.2**

Identify if the set $\mathbb{S} = \{\langle x, y\rangle \in \mathbb{R}^2 \mid x + y = 0\}$ is a vector space.

**Solution.**

Unlike the previous problem, we know that there exists a zero vector. For this one, let's go through each axiom. Let $u = \langle x_1, -x_1\rangle$ and $v = \langle x_2, -x_2\rangle$. Notice that $u \oplus v = \langle x_1 + x_2, -x_1 - x_2\rangle \in \mathbb{R}^2$ and satisfies the condition $x_1 + x_2 + (-x_1 - x_2) = 0$. Therefore, A1 is satisfied. Moreover, consider the following equations for A2 and A3 where $w = \langle x_3, -x_3\rangle$.

$$\begin{aligned}
u \oplus v &= \langle x_1, -x_1\rangle + \langle x_2, -x_2\rangle = \langle x_1 + x_2, -x_1 - x_2\rangle \\
&= \langle x_2 + x_1, -x_2 - x_1\rangle = \langle x_2, -x_2\rangle + \langle x_1, -x_1\rangle \\
&= v \oplus u \\
u \oplus (v \oplus w) &= \langle x_1, -x_1\rangle + (\langle x_2, -x_2\rangle + \langle x_3, -x_3\rangle) \\
&= \langle x_1, -x_1\rangle + \langle x_2 + x_3, -x_2 - x_3\rangle \\
&= \langle x_1 + x_2 + x_3, -x_1 - x_2 - x_3\rangle \\
&= \langle x_1 + x_2, -x_1 - x_2\rangle + \langle x_3, -x_3\rangle \\
&= (u \oplus v) \oplus w
\end{aligned}$$

Therefore, A2 and A3 are satisfied. Furthermore, notice that $\langle 0, 0\rangle \in \mathbb{R}^2$ and $0 + 0 = 0$, satisfying A4. Lastly for vector addition A5 holds as $\langle -x, x\rangle$ is the additive inverse of $\langle x, -x\rangle$ that satisfies the provided conditions. Thus, all axioms for vector addition are met.

Continuing with scalar multiplication, notice that S1 and S3 satisfy as $\langle x, y\rangle \in \mathbb{R}^2$ and $1 \odot \langle x, y\rangle = \langle x, y\rangle$. Moreover, consider the following equations for $\alpha, \beta \in \mathbb{R}$.

$$\begin{aligned}
\alpha \odot (\beta \odot u) &= \alpha \odot \langle \beta x_1, -\beta x_1 \rangle = \langle \alpha\beta x_1, -\alpha\beta x_1 \rangle = \alpha\beta \odot \langle x_1, -x_1 \rangle \\
&= (\alpha\beta) \odot u \\
(\alpha + \beta) \odot u &= \langle (\alpha+\beta)x_1, -(\alpha+\beta)x_1 \rangle = \langle \alpha x_1 + \beta x_1, -\alpha x_1 - \beta x_1 \rangle \\
&= \langle \alpha x_1, -\alpha x_1 \rangle + \langle \beta x_1, -\beta x_1 \rangle \\
&= \alpha \langle x_1, -x_1 \rangle + \beta \langle x_1, -x_1 \rangle \\
&= \alpha \odot u \oplus \beta \odot u \\
\alpha \odot (u \oplus v) &= \alpha \langle x_1 + x_2, -x_1 - x_2 \rangle = \langle \alpha x_1 + \alpha x_2, -\alpha x_1 - \alpha x_2 \rangle \\
&= \langle \alpha x_1, -\alpha x_1 \rangle + \langle \alpha x_2, -\alpha x_2 \rangle \\
&= \alpha \langle x_1, -x_1 \rangle + \alpha \langle x_2, -x_2 \rangle \\
&= \alpha \odot u \oplus \alpha \odot v
\end{aligned}$$

Thus, S2, S4, and S5 are all satisfied. Because all ten axioms for a vector space are met, the set $\mathbb{S} = \{\langle x, y \rangle \in \mathbb{R}^2 \mid x + y = 0\}$ is a vector space.

**Exercise 1.2.3**

Consider vectors in the form $\langle x, y, z \rangle$ where the vector addition is defined as $\langle x, y, z \rangle \oplus \langle x', y', z' \rangle = \langle x + x', y + y', z + z' \rangle$ and scalar multiplication as $\alpha \langle x, y, z \rangle = \langle x^\alpha, y^\alpha, z^\alpha \rangle$. Identify if the vectors form a vector space.

**Solution.**

First, notice that the vector addition adheres to the standard vector addition. Therefore, we only have to look at scalar multiplication if any axioms are violated. From a glance, S1, S2, and S3 appears to be satisfied. Therefore, let's take a look at S4.

$$(\alpha + \beta) \odot u = \left( x^{\alpha+\beta}, y^{\alpha+\beta}, z^{\alpha+\beta} \right)$$

Notice that in general, $x^{\alpha+\beta} \neq x^\alpha + x^\beta$. Therefore, the set of vectors does not form a vector space.

Just as we have vector spaces $\mathbb{R}^n$, we could also define spaces depending on the objects.

**Definition 1.2.4**

The symbol $\mathbb{P}^n$ denotes the vector space composed of polynomials of degree at most $n$ and adheres to the ten properties. $\mathbb{M}_{m\times n}$ represents the vector space with matrices of order $m \times n$ that adheres to the ten properties.

Therefore, we can say that $\mathbb{R}^n$ is more or less the same as $\mathbb{M}_{1\times n}$. With the understandings above, we can derive some important theorems. Although most of them are intuitive, it's always nice to formally prove them so let's do that now.

**Theorem 1.2.5**

For all vector spaces, zero vector, or the additive identity element, is unique.

*Proof.*

For the sake of contradiction, let $0_1, 0_2 \in V$ be distinct additive identity element for a vector space $V$. By definition, the following equations hold.

$$
\begin{aligned}
0_1 \oplus 0_2 &= 0_1 \\
0_1 \oplus 0_2 &= 0_2 \\
\therefore 0_1 \oplus 0_2 &= 0_1 = 0_2
\end{aligned}
$$

Because $0_1 = 0_2$ is contradictory from the fact that $0_1$ and $0_2$ are distinct, the initial assumption that there exists distinct additive identity element in a vector space is false.

Continuing from the uniqueness of the zero vector in a vector space, we can assert the following theorem.

**Theorem 1.2.6**

For all $u \in V$ for any vector space $V$, $v \in V$ such that $u \oplus v = 0$ is unique.

*Proof.*

For the sake of contradiction, assume $v, w \in V$ are distinct additive inverses of $u$. By definition and properties of vector spaces, the following equations hold.

$$v = v \oplus 0 = v \oplus (u \oplus w) = (v \oplus u) \oplus w = 0 \oplus w = w$$

Because $v = w$ contradicts the definition that $v$ and $w$ are unique, the initial assumption that there exists more than one additive inverse of a vector in vector spaces is false.

Just as $x = 0$ if $x + x = x$, we could do similar things in vector spaces.

**Theorem 1.2.7**

For all $v \in V$ and any vector space $V$, $v \oplus v = v$ if and only if $v = 0$.

*Proof.*

First, the first half of the statement can be proven. Consider the following equations.

$$\begin{aligned}(v \oplus v) \oplus (-v) &= v \oplus (-v) = 0 \\ &= v \oplus (v \oplus (-v)) = v \oplus 0 = v\end{aligned}$$

Therefore, if $v \oplus v = v$, then $v = 0$. Moreover, if $v = 0$, then $v \oplus v = 0 \oplus 0 = 0$. Therefore, the premise holds.

Now let's take a look at theorems for scalar multiplications.

**Theorem 1.2.8**

The following properties hold for all vector space $V$ over $\mathbb{F}$, $u, \mathbf{0} \in V$, and any scalars $\alpha$.

1. $0 \odot u = \mathbf{0}$
2. $\alpha \odot \mathbf{0} = \mathbf{0}$.
3. $(-1) \odot u = -u$
4. $u = -(-u)$
5. $\alpha \odot u = \mathbf{0}$ if and only if $\alpha = 0$ or $u = \mathbf{0}$.

Let's start with the first proof.

*Proof.*

Consider the following equation.

$$0 \odot u = (0 + 0) \odot u = 0 \odot u \oplus 0 \odot u$$

By Theorem 1.2.7, $0 \odot u = \mathbf{0}$.

Below is the proof for the second property.

*Proof.*

Consider the following equation.

$$\alpha \odot \mathbf{0} = \alpha \odot (\mathbf{0} \oplus \mathbf{0}) = \alpha \odot \mathbf{0} \oplus \alpha \odot \mathbf{0}$$

By Theorem 1.2.7, $\alpha \odot \mathbf{0} = \mathbf{0}$

Continuing, here's the proof for the third one.

*Proof.*

By the properties of vector spaces, $\mathbf{0} = 0 \odot u = (1 - 1) \odot u = 1 \odot u \oplus (-1) \odot u = u \oplus (-1) \odot u$. Because $(-1) \odot u$ is the additive inverse of $u$, $(-1) \odot u = -u$.

Below is the proof for the fourth one.

*Proof.*

Notice that $-(-u)$ is the additive inverse of $-u$. However, by definition, $u$ is the additive inverse of $-u$. By Theorem 1.2.6, $u = -(-u)$.

Finally, let's prove the last property.

*Proof.*

It suffices to show the latter half of the statement as the first half is shown by the first two properties. Consider the following equations for $\alpha \neq 0$.

$$\begin{aligned}\frac{1}{\alpha} \odot (\alpha \odot u) &= \frac{1}{\alpha} \odot \mathbf{0} = \mathbf{0} \\ &= \left(\frac{1}{\alpha} \cdot \alpha\right) \odot u = 1 \odot u = u\end{aligned}$$

Therefore, if $\alpha \neq 0$, then $u = \mathbf{0}$. Similarly, if $u \neq \mathbf{0}$, then $\alpha = 0$ as the product of nonzero numbers is a nonzero number.

We have covered the fundamental concepts and properties in vector spaces. Before we move on with the next topic in vector spaces, I wanted to discuss briefly on dot product as it will frequently appear when you deal with vectors.

**Definition 1.2.9**

**Dot product** is an algebraic operation defined as the following for vectors $a = [a_1, \ldots, a_n]$ and $b = [b_1, \ldots, b_n]$

$$a \cdot b = \sum_{k=1}^{n} a_k b_k$$

Here notice that the number of elements of the vectors in dot product must be equal for the product to be defined. Below are some examples of dot products.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -2 \\ -3 \end{bmatrix} = 1(-1) + 2(-2) + 3(-3) = -14$$

$$\begin{bmatrix} 0 & 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} = 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1$$

Note that this is very different from the matrix multiplication that we will discuss in the next note and we must write the notation $\cdot$ to ensure that we are not doing matrix multiplication. With that in mind, let's continue our discussion with subspaces.

## §1.3 Subspaces

In the previous section, we discussed that all ten axioms must be satisfied for a set to be a vector space. However, testing and proving that a set satisfies all ten axioms, to be honest, is quite tedious. The good news is that we only need to prove three conditions to show that a set is a vector space thanks to **subspaces**!

**Definition 1.3.1**

A **subspace** of a vector space $V$ is a subset of $V$ that uses the same definition of vector addition and scalar multiplication from $V$.

Now for a set to be a subspace, it must satisfy the following conditions.

**Theorem 1.3.2**

A non-empty subset $S$ of a vector space $V$ is a subspace of $V$ if and only if the following conditions are satisfied.

1. **Closure for Addition:** $\forall u, v \in S, u + v \in S$
2. **Closure for Scalar Multiplication:** $\forall u \in S$ and scalars $\alpha$, $\alpha u \in S$

*Proof.*

Notice that it suffices to show that $S$ is a vector space, or adhere to the then axioms for a vector space. Moreover, A1 and S1 are assumed to be true by the initial condition.

By the second condition, $\mathbf{0} = 0u \in S$ and $-u = (-1)u \in S$. Therefore, A4 and A5 are satisfied. For A2 and A3, $u, v, w \in V$ is true for $u, v, w \in S$ by definition. Moreover, because $V$ is a vector space, A2 and A3 hold. Similarly, because $V$ is a vector space that satisfies S1, S2, S3, S4, and S5, its subset $S$ also satisfies the conditions. Therefore, if the two closure conditions are met and $S$ is a subset of $V$, then $S$ is a subspace of $V$.

Now the good thing is that subspaces are vector spaces, so it suffices to show that a set is a subspace of a known vector space with the two conditions shown above to show that it is a vector space. Let's take a look at an example.

**Exercise 1.3.3**

Determine whether $S = \{(x, y, z) \in \mathbb{R}^3 \mid y = 0\}$ is a vector space.

**Solution.**

Notice that $S$ is a non-empty subset of $\mathbb{R}^3$, which is a known vector space. Therefore, it suffices to show that $u = (x_1, 0, z_1), v = (x_2, 0, z_2) \in S$ and scalars $\alpha$ satisfy the closure conditions for addition and scalar multiplication.

Notice that $u + v = (x_1 + x_2, 0, z_1 + z_2) \in S$. Moreover, $\alpha u = (\alpha x_1, 0, \alpha z_1) \in S$. Therefore, $S$ is a vector space, more specifically a subset of $\mathbb{R}^3$.

As you can see from the solution above, subspaces really made our lives easier. However, to make it even better, we can boil it down to a single condition.

**Corollary 1.3.4**

For scalars $\alpha$ and $\beta$ and $u, v \in S$ where $S$ is a non-empty subset of a vector space $V$ that has the same operations for vector addition and scalar multiplication, $S$ is a subspace if and only if the following equation is satisfied.

$$\alpha u + \beta v \in S$$

*Proof.*

First and foremost, the first half of the statement can be proven. If $\alpha u + \beta v \in S$, then $u + v \in S$ where $\alpha = \beta = 1$ since $\alpha$ and $\beta$ are all scalars. Similarly, $\alpha u + \beta v \in S$ implies that $\alpha u \in S$ where $\beta = 0$. By Theorem 1.3.2, $S$ is a subspace of $V$.

Continuing with the second half of the statement, if $S$ is a subset of $V$, then $\alpha u, \beta v \in S$. Therefore, $\alpha u + \beta v \in S$, proving the corollary.

More examples on this can be found in the practice problems below. Now let's get into the topics of **linear combination** and the **span** of a set.

### 1.3.1 Linear Combinations and Span

As always, let's start with definitions.

**Definition 1.3.5**

For scalars $\alpha_i$ and $v_i$ in a vector space $V$, the vector $u$ defined as the following is a **linear combination of the vectors** $v_i$.

$$u = \sum_{k=1}^{n} \alpha_k v_k$$

For instance, $[1, 2, 3]$ is a linear combination of $[2, -2, 1]$, $[1, 0, 2]$, and $[1, -1, 1]$ as it can be represented as the following.

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

Now we have a special term for the collection of such vectors.

**Definition 1.3.6**

The **span of the set of vectors** $\{v_1, \dots v_n\}$ is the set of all possible linear combinations of the vectors.

$$\text{span}\{v_1, \dots, v_n\} = \left\{ \sum_{k=1}^{n} \alpha_k v_k \;\middle|\; \alpha_k \in \mathbb{R} \right\}$$

With the definition, we could notice an important theorem.

**Theorem 1.3.7**

For a vector space $V$ and its subset $S = \{v_1, \dots, v_n\}$, $\text{span}(S)$ is a subspace of $V$.

*Proof.*

First and foremost, notice that $\mathbf{0} \in \text{span}(S)$ since all the scalars in linear combination can equal zero. Moreover, let $u = \sum_{k=1}^{n} \alpha_k v_k$ and $v = \sum_{k=1}^{n} \alpha_k v_k$ for any scalars $\alpha_k$ and $\beta_k$. By definition, $\alpha u + \beta v \in \text{span}(S)$ is true. Therefore, by Corollary 1.3.4, $\text{span}(S)$ is a subspace of $V$.

Side note here is that there are spans of infinite sets and they adhere to the theorem above. Continuing with the definition of the span of a set and its relation with subspace, we can show the two following theorems.

**Theorem 1.3.8**

Consider the following statements for a vector space $V$ and its subspace $W$ and non-empty subset $S$ such that $S \subset W$.

1. $\text{span}(S) \subset W$
2. The intersection of all subspaces containing $S$ is $\text{span}(S)$.

Let's start by proving the first statement.

*Proof.*

Let $u = \sum_{k=1}^{n} \alpha_k v_k \in \text{span}(S)$. By definition, $v_k \in W$ for integer $k \in [1, n]$. Moreover, $u \in W$ by Corollary 1.3.4 and $\text{span}(S) \subset W$.

Below is the proof for the second one.

*Proof.*

First and foremost, notice that by the previous statement, it was shown that $\text{span}(S)$ is included in all subspaces that contains $S$. Let $A$ be the intersection of all subspaces that contain $S$. Therefore, $\text{span}(S) \subset A$. Moreover, by Theorem 1.3.7 $\text{span}(S)$ is a subspace of $V$ that contains $S$. Thus, $A \subset \text{span}(S)$. Because $\text{span}(S) \subset A$ and $A \subset \text{span}(S)$, $\text{span}(S) = A$.

Next, let's conclude our discussion on linear combination and span of a set of vectors by proving the following statements.

**Theorem 1.3.9**

For a vector space $V$ and its subsets $S, T \subset V$, consider the following statements.

1. If $T \subset S$, then $\text{span}(T) \subset \text{span}(S)$.
2. $\text{span}(\text{span}(S)) = \text{span}(S)$.
3. If $T \subset \text{span}(S)$, then $\text{span}(T) \subset \text{span}(S)$.
4. For nonzero vectors $v_1, \ldots, v_n \in V$, scalar $\lambda \in \mathbb{R}$, and $1 \leq j, k \leq n$ such that $\lambda \neq 0$ and $j \neq k$, the following equations hold.

$$\text{span}\{v_1, \ldots, v_k, \ldots, v_n\} = \text{span}\{v_1, \ldots, v_k + \lambda v_j, \ldots, v_n\}$$
$$\text{span}\{v_1, \ldots, v_k, \ldots, v_n\} = \text{span}\{v_1, \ldots, \lambda v_k, \ldots, v_n\}$$

There's a lot to prove, but let's do them one by one starting with the first one.

*Proof.*

Notice that since all $t_k \in T$ satisfy $t_k \in S$, all $\sum_{k=1}^{n} c_k t_k \in \text{span}(T)$ for some appropriate $n$ and constant $c_k$ also satisfy $\sum_{k=1}^{n} c_k t_k \in \text{span}(S)$. Therefore, $\text{span}(T) \subset \text{span}(S)$.

Below is the proof for the second statement, which is a corollary of the first statement.

*Proof.*

First and foremost, notice that $S \subset \text{span}(S)$. By the first statement, $\text{span}(S) \subset \text{span}(\text{span}(S))$ holds. Moreover, notice that because $\text{span}(S)$ contains all linear combinations of $S$, making more linear combinations of the linear combinations of $S$ will not add more elements. Therefore, $\text{span}(\text{span}(S)) \subset \text{span}(S)$ and $\text{span}(\text{span}(S)) = \text{span}(S)$.

Continuing, let's prove the third statement.

*Proof.*

Notice that by the first statement, $\text{span}(T) \subset \text{span}(\text{span}(S))$. Continuing with the second statement shows that $\text{span}(T) \subset \text{span}(S)$.

Finally, here is the proof for the last statement.

*Proof.*

First, it is self-evident that $\text{span}\{v_1, \ldots, v_k + \lambda v_j, \ldots, v_n\} \subset \text{span}\{v_1, \ldots, v_k, \ldots, v_n\}$ since all linear combination of the set of vectors on the left can be represented as the linear combination of the set of vectors on the right. Moreover notice that $v_k$ can be represented as $v_k = (v_k + \lambda v_j) - \lambda v_j$ even if $k = j$. Therefore, $v_k \in \text{span}\{v_1, \ldots, v_k + \lambda v_j, \ldots, v_n\}$ and $\text{span}\{v_1, \ldots, v_k, \ldots, v_n\} \subset \text{span}\{v_1, \ldots, v_k + \lambda v_j, \ldots, v_n\}$ by statement 3. Therefore, the equations in the statement hold.

With the ideas of linear combinations and the span of a set of vectors in mind, let's discuss about linear independence.

### 1.3.2 Linear Independence

One of the most important reasons why we use Linear Algebra is due to efficiency. As you could have guessed, most vector spaces have infinitely many vectors. Of course, there must be an efficient way of representing the vector space instead of manually listing. Determining such representative vectors involves numerous factors, and that's what we will discuss with linear independence, basis, and dimension.

**Definition 1.3.10**

Consider the linear combination of $V = \{v_1, \ldots, v_n\}$ in a vector space and scalars $c_1, \ldots, c_n$ that satisfy the following equation.

$$\sum_{k=1}^{n} c_k v_k = \mathbf{0}$$

The set of vectors $V$ is **linearly dependent** if at least one of the scalars in the equation is nonzero. The set of vectors is **linearly independent** if all scalars are zero in the equation.

Let's take a look at a quick example.

**Exercise 1.3.11**

Determine whether $V = \{\langle 1, 1\rangle, \langle 2, 3\rangle \in \mathbb{R}^2\}$ is linearly independent.

**Solution.**

By definition, for the set to be linearly independent, all scalars $c_1$ and $c_2$ must be zero for the following equation to hold.

$$c_1\langle 1, 1\rangle + c_2\langle 2, 3\rangle = \langle 0, 0\rangle$$

Writing in system of linear equations, the equation above can be written as the following.

$$c_1 + 2c_2 = 0$$
$$c_1 + 3c_2 = 0$$

Solving the equation, $c_1 = c_2 = 0$ and no other set of scalars can lead to the equation. Therefore, the set of vectors is linearly independent.

One interesting observation that we can make is that for $n \in \mathbb{Z}$, if the set of vectors in $\mathbb{R}^n$ contains more than $n$ elements, then the set is linearly dependent. We will discuss more on this later when we discuss matrix and its application in linear systems, but for such cases, we have more variables than linear equations, which implies that there exists infinitely many solutions. Therefore, such a set of vectors must be linearly dependent. Continuing, below is a theorem that we can deduce from the definition.

**Theorem 1.3.12**

A set with a finite number of nonzero vectors is linearly dependent if and only if a vector in the ordering of the set can be represented as a linear combination of vectors preceding it.

*Proof.*

First and foremost, the first half of the theorem can be shown. Consider the set of vectors $\{v_1, \ldots, v_{j-1}, v_j, \ldots, v_n\}$. Assume for scalars $c_k$, the following equation is true.

$$v_j = \sum_{k=1}^{j-1} c_k v_k$$

Therefore, $\mathbf{0} = \sum_{k=1}^{j-1} c_k v_k - v_j$. Letting $c_k = 0$ for integer $k \in (j, n]$, the following equation holds.

$$\sum_{k=1}^{j-1} c_k v_k - v_j + \sum_{k=j+1}^{n} c_k v_k = \mathbf{0}$$

Because at least one scalar is nonzero, the set of vectors $\{v_1, \ldots, v_n\}$ is linearly dependent.

Continuing, let the set of vectors $\{v_1, \ldots, v_n\}$ be linearly dependent. For the corresponding scalars $c_k$, let $c_j$ be the last scalar that is nonzero. Thus, the following equation holds.

$$\sum_{k=1}^{j} c_k v_k + \sum_{k=j+1}^{n} 0 v_k = \mathbf{0}$$

Rearranging the equation, $v_k = -\sum_{k=1}^{j-1} \frac{c_k}{v_j} v_k$, which is a linear combination. Because the converse holds, the theorem is true.

From this theorem, we can deduce a corollary.

**Corollary 1.3.13**

For a vector space $V = \text{span}(S)$ and linearly dependent set $S = \{v_1, \ldots, v_n\} \subset V$, then there exists a set $T \subset S$ such that $|T| = n - 1$ and $V = \text{span}(T)$.

*Proof.*

By Theorem 1.3.12, there exists element $v_j \in S$ such that $v_j$ is the linear combination of the previous elements as listed in $S$.

Let $T = \{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_n\}$. Consider the following equations for scalars $d_k$ and $c_k$.

$$\begin{aligned}
V = \sum_{k=1}^{n} d_k v_k &= \sum_{k=1}^{j-1} d_k v_k + d_j v_j + \sum_{k=j+1}^{n} d_k v_k \\
&= \sum_{k=1}^{j-1} d_k v_k + d_j \left( \sum_{k=1}^{j-1} c_k v_k \right) + \sum_{k=j+1}^{n} d_k v_k \\
&= \sum_{k=1}^{j-1} (d_j c_k + d_k) v_k + \sum_{k=j+1}^{n} d_k v_k = \operatorname{span}(T)
\end{aligned}$$

Therefore, the corollary holds.

Continuing from the corollary, the following theorems could be asserted.

**Theorem 1.3.14**

Consider the following statements for set $S$.

1. If $S$ contains a zero vector, then it is linearly dependent.
2. If $|S| = 1$, then it is linearly dependent if and only if its element is the zero vector.
3. If $|S| = 2$, then it is linearly dependent if and only if a vector is a scalar multiple of the other.
4. If $S$ is linearly dependent, then every set that contains $S$ is linearly dependent.
5. If $S$ is linearly independent, then all its subsets are linearly independent.

There's quite a bit to prove! Let's start with the first one.

*Proof.*

Let $S = \{\mathbf{0}, v_1, \ldots, v_n\}$. Notice that $\alpha(\mathbf{0}) + \sum_{k=1}^{n} 0 v_k = \mathbf{0}$ for any scalar $\alpha$. Therefore, $S$ is linearly dependent.

Below is the proof for the second one.

*Proof.*

Let $S = \{v\}$. Notice that if $v \neq \mathbf{0}$, it is not possible for a nonzero scalar $\alpha$ to make $\alpha v = \mathbf{0}$. Therefore, if $|S| = 1$, then $v = \mathbf{0}$. For the converse, if $v = \mathbf{0}$, then any scalar $\alpha$ will let $\alpha v = \mathbf{0}$. Therefore the statement holds.

Here is the proof for the third statement.

*Proof.*

Let $S = \{v_1, v_2\}$. First, if $v_2 = \alpha v_1$ for any scalar $\alpha$, then $\mathbf{0} = \alpha v_1 - \alpha v_1 = \alpha v_1 - v_2$. Therefore, $S$ is linearly dependent. For the converse, $c_1 v_1 + c_2 v_2 = 0$ for scalars $c_1$ and $c_2$. Therefore, assuming that the scalars are nonzero without loss of generality, $v_2 = -\frac{c_1}{c_2} v_1$. Thus, the statement is true.

Below is the proof for the fourth one.

*Proof.*

Let $S = \{v_1, \ldots, v_n\}$ and $T = \{u_1, \ldots, u_n, v_1, \ldots, v_n\}$. By definition, there exists $c_k$ that are not all zero such that $\sum_{k=1}^{n} c_k v_k = \mathbf{0}$. Because $\sum_{k=1}^{n} 0 u_k + \sum_{k=1}^{n} c_k v_k = \mathbf{0}$, $T$ is also linearly dependent.

Finally, here is the proof for the last statement.

*Proof.*

Let $T \subset S$ where $S$ is linearly independent. Assume for the sake of contradiction that there exists a subset $T$ that is linearly dependent. However, if $T$ is linearly dependent, then $S$ is linearly dependent by the previous property, which is a contradiction. Therefore, by contradiction, if $S$ is linearly independent then all its subsets $T$ are also linearly independent.

With this, we have concluded our discussion on linear dependence and independence. Before concluding this note, let's discuss basis and dimension of a set of vectors.

### 1.3.3 Basis and Dimension

As always, let's start with definition.

**Definition 1.3.15**

For a vector space $V$ and its subset $S$, if $\text{span}(S) = V$, then $S$ is a spanning set for $V$.

When we say aloud, we say that $S$ spans $V$. For instance, the set $\{\langle 1,0\rangle, \langle 0,1\rangle\}$ spans $\mathbb{R}^2$ since all vectors $\langle a,b\rangle \in \mathbb{R}^2$ can be represented as the sum $a\langle 1,0\rangle + b\langle 0,1\rangle$.

**Definition 1.3.16**

A linearly independent subset that spans $V$ is a **basis** for $V$.

Continuing from the previous example, we can see that $\{\langle 1,0\rangle, \langle 0,1\rangle\}$ not only spans $\mathbb{R}^2$, but is also a basis for $\mathbb{R}^2$ since it is linearly independent. As you can see, the same will apply if we generalize it to $\mathbb{R}^n$.

**Definition 1.3.17**

The **standard basis for** $\mathbb{R}^n$ is the set $S = \{e_1, \ldots e_n\}$ where $e_i$ is the $i$th column of $I_n$.

We could do the same for polynomials. First, notice that the set $\{1, x, x^2\}$ spans $\mathbb{P}^2$. Moreover, the set is linearly independent. Therefore, $\{1, x, x^2\}$ is a basis for $\mathbb{P}^2$.

**Definition 1.3.18**

The **standard basis for** $\mathbb{P}^n$ is the set $S = \{1, x, x^2, \ldots, x^n\}$.

From the two definitions above, we can notice that there are finite elements for the two standard bases for $\mathbb{R}^n$ and $\mathbb{P}^n$. We call such vector spaces **finite-dimensional**.

**Definition 1.3.19**

A vector space with a basis of finite elements is called **finite-dimensional**. Whereas, **infinite dimensional** vector spaces do not contain a basis with finite elements.

An example of infinite dimensional vector space would be $\mathbb{P}$ since the basis is required to be an infinite set. However, we will not be discussing infinite dimensional vector spaces. From the definitions above, we could establish a theorem.

**Theorem 1.3.20**

For a basis $S$ of a vector space $V$ such that $|S| = n$, any set in $V$ with more than $n$ elements is linearly dependent.

*Proof.*

Let $S = \{v_1, \ldots, v_n\}$ and $T = \{u_1, \ldots, u_m\}$ such that $S, T \in V$ and $m > n$. Notice that because $\text{span}(S) = V$ and $T \in V$, there exists a set of constants $a_{ij}$ that satisfies the following equation (note that we will be using this indices in the next note when we discuss matrices.).

$$\begin{aligned} u_1 &= \sum_{j=1}^{n} a_{1j} v_j \\ u_2 &= \sum_{j=1}^{n} a_{2j} v_j \\ &\vdots \\ u_m &= \sum_{j=1}^{n} a_{mj} v_j \end{aligned}$$

Therefore, it suffices to show that there exists constants $c_1, \ldots, c_m$

such that not all are zero and that satisfies the following equation.

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_i a_{ij} v_j = \mathbf{0}$$

Continuing, it suffices to show that the variables $c_i$ and constants $a_{ij}$ satisfy the following equations where not all $c_i$ are zero.

$$\begin{aligned}
\sum_{i=1}^{m} c_i a_{i1} &= 0 \\
\sum_{i=1}^{m} c_i a_{i2} &= 0 \\
&\vdots \\
\sum_{i=1}^{m} c_i a_{in} &= 0
\end{aligned}$$

Notice that there are $m$ variables and $n$ equations. It is known that for a linear system of equations, there exists an infinite number of solutions if there are more variables than the equations (don't worry we will prove this in future notes.). Because $m > n$, there exists $c_1, \ldots c_m$ that are not all zero and that satisfy the equations for linear dependence, the theorem holds.

Continuing, we can derive three corollaries from the theorem.

**Corollary 1.3.21**

If a basis for the vector space $V$ contains $n$ elements, then any linearly independent subsets of $V$ contain at most $n$ elements.

*Proof.*

By definition, a basis of a vector space $V$ spans $V$. Because any linearly dependent subsets of $V$ contain more than $n$ elements by Theorem 1.3.20, any linearly independent subsets of $V$ must not contain more than $n$ elements.

Below is the second corollary.

**Corollary 1.3.22**

If $V$ is a finite vector space, every basis of $V$ will contain the same number of elements.

*Proof.*

Let $S$ and $T$ be bases of $V$ with $p$ and $q$ elements respectively. By definition, the bases of a vector space are linearly independent. By the previous corollary, it is evident that $q \leq p$ and $p \leq q$ are true. Therefore, $p = q$ and the corollary holds.

Now before continuing with our third corollary, let's first define what the dimension of a vector space is.

**Definition 1.3.23**

For a finite dimensional vector space $V$, the number of elements in any basis of $V$ is called the **dimension** of $V$, denoted as $\dim(V)$.

For instance, recall the standard basis of $\mathbb{R}^n$. From the number of elements in the basis, we can write that $\dim(\mathbb{R}^n) = n$. Similarly, we can also notice that $\dim(\mathbb{P}^n) = n + 1$. Now let's continue with our third corollary.

**Corollary 1.3.24**

Every set $S$ in an $n$-dimensional vector space such that $|S| = n + 1$ is linearly dependent.

*Proof.*

By definition, a basis of an $n$-dimensional vector space $V$ contains $n$ elements. By Theorem 1.3.20, since $n + 1 > n$, $S$ must be linearly dependent.

One fun fact to note is that for the vector space $\{0\}$, we say $\dim(\{0\}) = 0$

by convention. Moreover, consider the following theorem.

**Theorem 1.3.25**

For a basis $S = \{v_1, \ldots, v_n\}$ of a vector space $V$, any vector $u \in V$ can be written as a unique linear combination of $S$.

*Proof.*

For the sake of contradiction, assume that $u$ can be written as $u = \sum_{k=1}^{n} c_k v_k = \sum_{k=1}^{n} d_k v_k$ where $c_k \neq d_k$ for some choice of $k$. Therefore, the following equations can be obtained.

$$\sum_{k=1}^{n} c_k v_k = \sum_{k=1}^{n} d_k v_k$$

$$\sum_{k=1}^{n} (c_k - d_k) v_k = 0$$

Note that by definition, $S$ is linearly independent and $c_k - d_k = 0$ for all integer $k \in [1, n]$. Because it contradicts with the assumption that $c_k \neq d_k$ for some choice of $k$, the theorem holds.

Now with the theorem in mind, because we know that the linear combination of a basis to represent a vector is unique, we can establish the following definition.

**Definition 1.3.26**

For a basis $S = \{v_1, \ldots, v_n\}$ of a vector space $V$ and any vector $u = \sum_{k=1}^{n} c_k v_k \in V$, the **coordinates of** $u$ **with respect to** $S$, denoted as $(u)_S$ is the list of coefficients $c_1, \ldots, c_n$.

We often represent the coordinates with $n$-tuple. To see that, let's take a look at the following example.

**Exercise 1.3.27**

Find the coordinate of $f(x) = x^2 + 2x + 3$ with respect to the basis $S = \{x^2, 2x^2 + 1, x\}$ in the vector space $\mathbb{P}^2$.

**Solution.**

Consider the following equation.

$$f(x) = 3\left(2x^2 + 1\right) - 5\left(x^2\right) + 2(x)$$

Therefore, the coordinate of $f(x)$ can be represented as the following 3-tuple.

$$\begin{bmatrix} -5 \\ 3 \\ 2 \end{bmatrix}_S$$

Lastly, let's conclude our discussion on basis and dimension with a theorem.

**Theorem 1.3.28**

Consider the following statements for a vector space $V$ and its subset $S \subset V$.

1. There exists a basis for $V$ that is a subset of $S$ if $\operatorname{span}(S) = V$.
2. There exists a basis for $V$ that includes all elements in $S$ if $S$ is linearly independent.

Let's start by proving the first statement.

*Proof.*

Notice that by definition, a basis of a vector space $V$ is a linearly independent set that spans $V$. Therefore, it suffices to show that there exists a linearly independent set that spans $V$ that is also a subset of $S$.

By Theorem 1.3.12, a subset of $S$ is linearly dependent if any only

> if a vector can be represented as the linear combination of the preceding vectors. Let $T$ be a subset of $S$ without the vectors that can be represented as the linear combination of the preceding vectors. Notice that $T$ spans $V$ as the vector removed can be represented as the linear combination of the other vectors. Moreover, because $T$ is linearly independent, there exists a basis for $V$ that is a subset of $S$.

Below is the proof for the second statement.

> *Proof.*
>
> Let $T$ be a basis for $V$ and $A := S \cup T$. By definition, $\text{span}(A) = V$. Construct $B \in A$ by removing all vectors that can be represented as a linear combination of the preceding vectors. Notice that no elements from $S$ will be removed as $S$ is linearly independent. By construction, $B$ spans $V$ and is also linearly independent. Because $S \subset B$, the statement holds.

We discussed fundamental topics of vector spaces. Of course there are a lot more to discuss in vector spaces; however, to do so, we need some understanding of matrices. So let's pause our discussions on vector spaces for now, and move on to matrices and their applications in linear system of equations before looking at their implications in vector spaces.

## §1.4 Practice Problems

1. Consider the set $S = \{(x, y, z) \in \mathbb{R}^3 \mid x = 0\}$. Identify whether the set is a vector space under the standard vector addition and scalar multiplication.

2. Determine whether $S = \{f \in \mathbb{R} \mid f'(0) = 0\}$ is a vector space for differentiable functions $f$.

3. For all $n \in \mathbb{Z}$ and $a_i \in \mathbb{R}$, determine whether

$$S = \left\{ (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \;\middle|\; \sum_{k=1}^{n} a_k x_k = 0 \right\}$$

is a vector space.

4. Show that the intersection of subspaces of a vector space $V$ is also a subspace of $V$.

5. Show that for distinct real numbers $a_1, \ldots, a_n$, the set $\{e^{a_1 x}, \ldots, e^{a_n x}\}$ is linearly independent.

6. Let $A, B \in \mathbb{R}^3$ be lines defined as $x = y = z$ and $3x = 4y = 5z$ respectively. If $C = \{a + b \mid a \in A, b \in B\}$, show that $C = A \oplus B$ where $\oplus$ represents the direct sum [Erd20].

7. If $\{a, b, c\}$ is linearly independent in a vector space $V$, show that $\{a + b, b + c, c + a\}$ is also linearly independent [Erd20].

8. Show that the set of all polynomials in $\mathbb{P}^3$ that includes the origin is a subspace of $\mathbb{P}^3$.

# Note 2

# Matrices

Matrices are one of the fundamental components that build linear algebra. This note discusses the fundamental concepts and properties of matrices.

## §2.1 Basic Terms and Notations

As always, let's get started with the definition.

**Definition 2.1.1**

A **matrix** is a 2-dimensional array composed of entries called **elements**. A matrix is said to have **order** "$m$ by $n$" and written as $m \times n$ if there are $m$ rows and $n$ columns.

Technically, the elements of a matrix can be anything including functions. However, the majority of the matrices discussed in linear algebra will have numeric elements. Let's take a look at some examples.

$$A = \underbrace{\begin{bmatrix} 1 & 2 & 3 & 4 \\ -1 & -2 & -3 & -4 \\ 0 & -1 & 0 & 1 \end{bmatrix}}_{4 \text{ columns}} \left.\vphantom{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}}\right\} 3 \text{ rows}, \qquad B = \underbrace{\begin{bmatrix} \pi \\ e \\ 0 \\ 1 \end{bmatrix}}_{1 \text{ column}} \left.\vphantom{\begin{bmatrix} \pi \\ e \\ 0 \\ 1 \end{bmatrix}}\right\} 4 \text{ rows}$$

Here, matrix $A$ has order $3 \times 4$ and matrix $B$ has order $4 \times 1$. The form of a matrix that has order $m \times n$ can be generalized as the following.

$$A = \underbrace{\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{bmatrix}}_{m \text{ columns}} \left.\vphantom{\begin{bmatrix} a \\ a \\ a \\ \vdots \\ a \end{bmatrix}}\right\} n \text{ rows}$$

It is also worth noting that matrix $A$ is often represented as $[a_{ij}]$ and the elements as $a_{ij}$ like $a_{12}$ for $a_{1,2}$.

**Definition 2.1.2**

An element $a_{ij}$ is said to have row index of $i$ and column index of $j$.

As you probably have guessed, we assign special names to some matrices depending on $m$ and $n$.

**Definition 2.1.3**

**Square matrices** are matrices with an equal number of rows and columns. The elements with the same row and column index are called **diagonal elements** of the matrix. The collection of such elements is called **main diagonal**, or **principal diagonal**.

What if either $m$ or $n$ is equal to 1? Yes, we do have special terms for them.

**Definition 2.1.4**

**Row matrices** are matrices that are composed of only one row. A **column matrix** is an array composed of one column.

If you have a machine learning background, you likely have heard of the term "high dimensional arrays or vectors." Here, the term "dimension" works the same way.

**Definition 2.1.5**

The **dimension** of a row or column matrix is the number of **components**, or the elements. An **n-tuple** is an $n$-dimensional row or column vector.

Let's take a look at an example from the previous matrix $B$ reproduced below.

$$B = \underbrace{\begin{bmatrix} \pi \\ e \\ 0 \\ 1 \end{bmatrix}}_{\text{1 column}} \left.\vphantom{\begin{bmatrix} \pi \\ e \\ 0 \\ 1 \end{bmatrix}}\right\} \text{4 rows}$$

Here, matrix $B$ is a 4-dimensional matrix with four components. This column matrix can also be called as 4-tuple.

We could also define matrices by their elements.

**Definition 2.1.6**

A **zero matrix** is a matrix with all its elements as 0.

With the fundamental terms in mind, let's discuss the properties of matrices.

## §2.2 Fundamental Properties

First, when are two matrices "equal" to each other? Consider the following definition for a formal explanation.

**Theorem 2.2.1**

Two matrices $A$ and $B$ with orders $m \times n$ and $p \times q$ respectively are said to be equal to each other if they retain the same order and corresponding elements, i.e. they satisfy the following properties.

1. $m = p$
2. $n = q$
3. For elements $a_{ij}$ in $A$ and $b_{ij}$ in $B$, the following equation satisfies.
$$a_{ij} = b_{ij} \, \forall \, i \in [1, m] \text{ and } j \in [1, n]$$

Although this is very trivial and intuitive, it is always nice to formally define concepts!

**Theorem 2.2.2**

For matrices $A$ and $B$ with same order, the element in the $(i, j)$-entry of their sum $A + B$ is equivalent to the sum of elements in the $(i, j)$-entry of $A$ and $B$.

Note that the orders of matrices being added must be identical. We do not consider additions with matrices of different orders. Below is an example of addition.

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & \pi & 4 \end{bmatrix} + \begin{bmatrix} -2 & 1 & 3 \\ 3 & \pi & -4 \end{bmatrix} = \begin{bmatrix} -1 & 4 & 8 \\ 5 & 2\pi & 0 \end{bmatrix}$$

Similarly, we could see that subtractions are defined in similar way. Subtracting the matrices above, we get the following results.

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & \pi & 4 \end{bmatrix} - \begin{bmatrix} -2 & 1 & 3 \\ 3 & \pi & -4 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 2 \\ -1 & 0 & 8 \end{bmatrix}$$

Now, as we look at the additions and subtractions of matrices, you probably wondered about the application of additive properties. Intuitively, the fundamental properties of addition can be applied to matrices addition.

**Theorem 2.2.3**

For matrices $A$ and $B$ with order $m \times n$, $A + B = B + A$.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{m\times n}$. By definition, the sum $A + B = [a_{ij} + b_{ij}]$. Because addition is commutative, the following equation holds.

$$A + B = [a_{ij} + b_{ij}] = [b_{ij} + a_{ij}] = B + A$$

Thus, the commutative property of matrix addition holds.

In addition to the commutative property, the associative property is also applicable.

**Theorem 2.2.4**

For matrices $A$, $B$, and $C$ with order $m \times n$, $A + (B + C) = (A + B) + C$.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$, $B = [b_{ij}]_{m\times n}$, and $C = [c_{ij}]_{m\times n}$. Consider the following equation.

$$\begin{aligned} A + (B + C) &= [a_{ij}] + [b_{ij} + c_{ij}] = [a_{ij} + (b_{ij} + c_{ij})] = [a_{ij} + b_{ij} + c_{ij}] \\ &= [(a_{ij} + b_{ij}) + c_{ij}] = [a_{ij} + b_{ij}] + [c_{ij}] \\ &= (A + B) + C \end{aligned}$$

Thus, the associative property of addition applies to the matrix addition.

One obvious property in addition is that when you add 0 to a number, you get the number itself. The same applies in matrix addition.

**Theorem 2.2.5**

The sum of matrices $A_{m\times n}$ and $0_{m\times n}$ yields the matrix $A$, i.e. the zero matrix is the identity elements of matrix addition.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$. By definition, the sum $A + 0 = [a_{ij} + 0] = [a_{ij}] = A$. Therefore, the zero matrix is the identity element of matrix addition.

Now what if we want to add the same matrix multiple times? Fortunately, we already resolved the issue. For instance, we write $x + x + x$ as $3x$. We can multiply matrices with scalars.

**Theorem 2.2.6**

For a complex number $\lambda$ and $A = [a_{ij}]$, $\lambda A = [\lambda a_{ij}]$.

Let's take a look at an example.

**Exercise 2.2.7**

Compute $2A - 3B$ for $A$ and $B$ defined as the following.

$$A = \begin{bmatrix} 2 & 3 & -1 \\ 1 & 7 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 1 & 0 \\ 5 & 3 & 8 \end{bmatrix}$$

**Solution.**

Before subtracting, we could perform scalar multiplication first.

$$\begin{aligned} 2A - 3B &= 2\begin{bmatrix} 2 & 3 & -1 \\ 1 & 7 & 3 \end{bmatrix} - 3\begin{bmatrix} -2 & 1 & 0 \\ 5 & 3 & 8 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 6 & -2 \\ 2 & 14 & 6 \end{bmatrix} - \begin{bmatrix} -6 & 3 & 0 \\ 15 & 9 & 24 \end{bmatrix} \end{aligned}$$

From here we could subtract two matrices. Therefore, $2A - 3B$ can be represented as following.

$$2A - 3B = \begin{bmatrix} 10 & 3 & -2 \\ -13 & 5 & -18 \end{bmatrix}$$

With the scalar multiplication in mind, we can expand the application of foundational concepts here with matrices.

**Theorem 2.2.8**

For $A = [a_{ij}]_{m \times n}$, $B = [b_{ij}]_{m \times n}$, and $\lambda_1, \lambda_2 \in \mathbb{C}$, the following properties hold.

1. $\lambda_1(A \pm B) = \lambda_1 A \pm \lambda_1 B$
2. $(\lambda_1 \pm \lambda_2)A = \lambda_1 A \pm \lambda_2 A$
3. $\lambda_1(\lambda_2 A) = (\lambda_1 \lambda_2)A$

Let's prove them! Below is the proof for the first property.

*Proof.*

Consider the following equation.

$$\lambda_1(A \pm B) = \lambda_1([a_{ij}] \pm [b_{ij}]) = \lambda_1([a_{ij} \pm b_{ij}])$$

Note that individual elements are multiplied in scalar multiplication. Moreover, because $\lambda_1(a_{ij} \pm b_{ij}) = \lambda_1 a_{ij} \pm \lambda_1 b_{ij}$, the following equation holds.

$$\lambda_1([a_{ij} \pm b_{ij}]) = [\lambda_1 a_{ij} \pm \lambda_1 b_{ij}] = [\lambda_1 a_{ij}] \pm [\lambda_1 b_{ij}] = \lambda_1 A \pm \lambda_1 B$$

Therefore, $\lambda_1(A \pm B) = \lambda_1 A \pm \lambda_1 B$.

Now let's prove the second property.

*Proof.*

By definition of scalar multiplication, $(\lambda_1 \pm \lambda_2)A$ can be represented as the following.

$$(\lambda_1 \pm \lambda_2)A = (\lambda_1 \pm \lambda_2)[a_{ij}] = [(\lambda_1 \pm \lambda_2)a_{ij}]$$

Using the distributive property, the following equation holds.

$$[(\lambda_1 \pm \lambda_2)a_{ij}] = [\lambda_1 a_{ij} \pm \lambda_2 a_{ij}] = [\lambda_1 a_{ij}] \pm [\lambda_2 a_{ij}] = \lambda_1 A \pm \lambda_2 A$$

Thus, the equation $(\lambda_1 \pm \lambda_2)A = \lambda_1 A \pm \lambda_2 A$ holds.

Finally, here is the proof to the third property.

*Proof.*

Consider the following equation.

$$\begin{aligned}\lambda_1(\lambda_2[a_{ij}]) = \lambda_1[\lambda_2 a_{ij}] = [\lambda_1\lambda_2 a_{ij}] = [(\lambda_1\lambda_2)a_{ij}] \\ = (\lambda_1\lambda_2)[a_{ij}] = (\lambda_1\lambda_2)A\end{aligned}$$

Therefore, the property holds.

We have discussed fundamental properties so far. In the next section of the note, let's discuss what matrix multiplication with intuitions and computations.

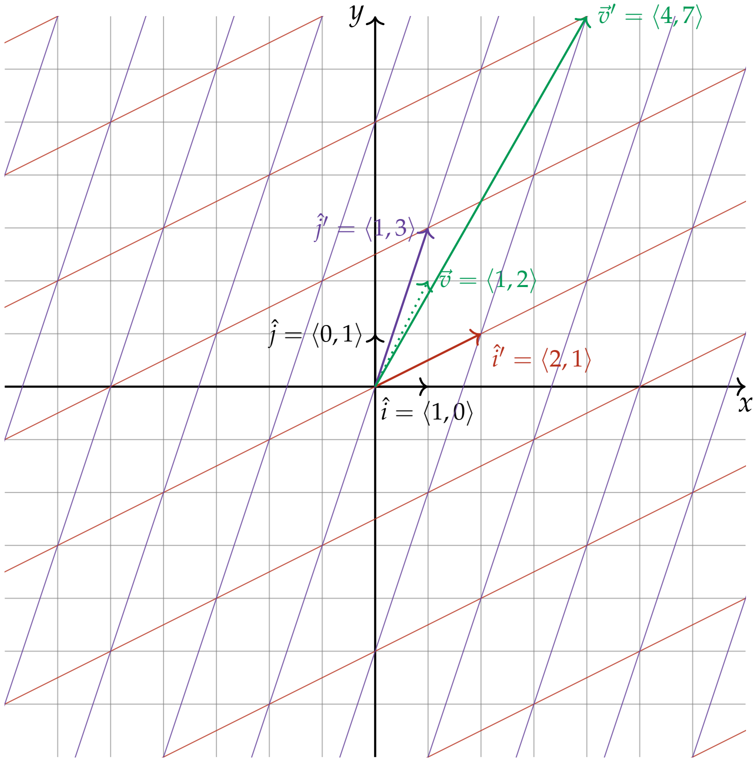
## §2.3 Matrix Multiplication

Before we get into formulas and computations, let's build our intuition on what a matrix and its multiplication is supposed to represent. We will discuss **linear transformation** in depth in latter notes, but we can think of it as a function that transforms the vector space. Matrix instructs the function how to transform the space. Let's take a look at a quick example.

Let $A$ be a matrix that has order $2 \times 2$. The first column tells how to map the $\hat{\imath}$ and the second column represents where our $\hat{\jmath}$ is mapped.

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Matrix $A$ is telling us to map the standard basis $\hat{\imath} = \langle 1, 0 \rangle$ to $\hat{\imath}' = \langle 2, 1 \rangle$ and from $\hat{\jmath} = \langle 0, 1 \rangle$ to $\hat{\jmath}' = \langle 1, 3 \rangle$. Consider the diagram below.




After the transformation, you can imagine mapping everything in the standard vector space with respect to the new colored space. That, is what matrix multiplication does. See how a vector $\vec{v} = \langle 1, 2 \rangle$ is now on $\langle 4, 7 \rangle$. Using matrix multiplication, we can represent the mapping as the following.

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

Because we cannot draw such planes every time, let's take a look at how the vector $\vec{v} = \langle 1, 2 \rangle$ is mapped to $\langle 4, 7 \rangle$. First, from the previous note, we have seen that the first row represents the $x$ coordinate and the second row represents the $y$ coordinate. Now, to find where the point $(1, 2)$ is located in the "scaled" version, we could simply look at the basis vectors in the "scaled" plane. Because the vector that we are looking for is the sum of the two "scaled" basis vector, we can represent our multiplication as the following.

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$

Nice! We get the same result. Now that we have intuitions on what a matrix and its multiplication represents, let's take a look at algebraic properties and resulting formulas from the intuition.

### 2.3.1 The Method and Properties

As you can see above, matrix multiplication is not defined element-wise unlike addition and scalar multiplication. First, take a look at the following formula for individual elements.

**Theorem 2.3.1**

For matrices $[a_{ij}]_{m \times n}$ and $[b_{ij}]_{n \times p}$, the elements $c_{ij}$ of the product matrix $[c_{ij}]_{m \times p}$ are defined as the following expression.

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

Let's take a look at an example to see how this works.

**Exercise 2.3.2**

Compute the following matrix multiplication.

$$\begin{bmatrix} 2 & 3 & 4 \\ 1 & 5 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 0 \end{bmatrix}$$

**Solution.**

Consider the following computation.

$$\begin{bmatrix} 2 & 3 & 4 \\ 1 & 5 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 4 & 0 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 + 3 \cdot 2 + 4 \cdot 4 & 2 \cdot 2 + 3 \cdot 3 + 4 \cdot 0 \\ 1 \cdot 1 + 5 \cdot 2 - 1 \cdot 4 & 1 \cdot 2 + 5 \cdot 3 - 1 \cdot 0 \end{bmatrix}$$

$$= \begin{bmatrix} 2 + 6 + 16 & 4 + 9 + 0 \\ 1 + 10 - 4 & 2 + 15 \end{bmatrix} = \begin{bmatrix} 24 & 13 \\ 7 & 17 \end{bmatrix}$$

It is important to note that matrix multiplication applies under a specific condition.

**Theorem 2.3.3**

For matrices $A_{m \times n}$ and $B_{p \times q}$, the matrices can be multiplied if and only if $n = p$. The resulting matrix will have order $m \times q$.

I believe the condition above is somewhat self-evident. When the condition above is not met, we say the product is **undefined**. Continuing from this condition, we could get an interesting property.

**Theorem 2.3.4**

Matrix multiplication is not commutative, i.e. for matrices $A$ and $B$, it is not necessarily true that $AB = BA$.

*Proof.*

I will leave a more formal proof for the readers. However, I believe

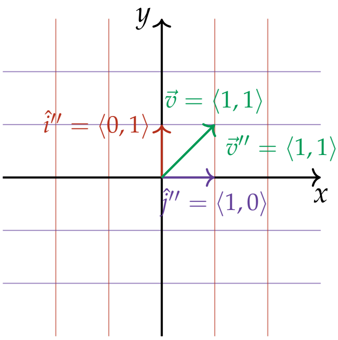
> this is self-evident as for matrices $A_{m\times n}$ and $B_{n\times p}$ where $m \neq p$, $AB$ is defined while $BA$ is not.

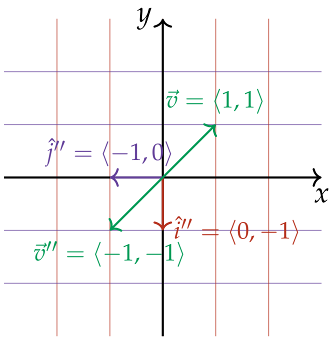
A fun way of understanding commutativity is to think about transformation questions that we see in elementary or middle school competition math. You might be familiar with questions asking orders of transformation of an object. For instance, let $A$ be defined as the $90^\circ$ rotation counterclockwise and let transformation $B$ be reflection across the $y$ axis. Is performing $AB$ equal to $BA$? Obviously, the answer is false when you visually think about it. What if we apply the same idea here with matrices?

Notice the matrices $A$ and $B$ can be represented as following using matrices.

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

Taking a look at the visualization, the left diagram represents $AB$ and the right diagram visualizes $BA$ with $\vec{v} = \langle 1, 1 \rangle$ as the reference.




Let's take a look at what's happening here. $\vec{v}''$ in our left diagram can be represented as the following.

$$\vec{v}'' = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \left( \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$$

Notice that we have to write it in the order $B(A\vec{v})$ as we would be performing $A$ first. We can think of it as reading it from left to right just as we would read from left to right when computing composite functions.

$$\vec{v}'' = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \left( \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \left( \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

This is exactly what we have for the left diagram. We could do the same for the right side.

$$\vec{v}'' = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \left( \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \left( \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

This is also supported! Alternatively, we could just compute $AB$ and $BA$ to see what our "final transformation is".

$$BA = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
$$AB = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$$

Clearly, we could see that the mapping is different.

Another interesting property for matrices $A$, $B$, and $C$ is that even if $AC = BC$, it does not necessarily mean that $A = B$. Think about the case when $C$ is a zero matrix. Although matrix multiplication is not commutative, the associative and distributive property still applies.

**Theorem 2.3.5**

For matrices $A$, $B$, and $C$, the following equations are true given that the conditions for the addition and multiplication are met.

1. $A(BC) = (AB)C$
2. $A(B + C) = AB + AC$
3. $(B + C)A = BA + CA$
4. $\lambda(AB) = A(\lambda B)$

Let's start by proving the associative property.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$, $B = [b_{ij}]_{n\times q}$, and $C = [c_{ij}]_{q\times r}$. First, notice that the elements of the product $BC$, or $D = [d_{ij}]_{n\times r}$ can be represented as $d_{ij} = \sum_{k=1}^{q} b_{ik}c_{kj}$. Multiplying it again with $[a_{ij}]$, the elements of the product $E = [e_{ij}]_{m\times r}$ can be represented as the following.

$$e_{ij} = \sum_{s=1}^{n} a_{is}d_{sj} = \sum_{s=1}^{n}\left(a_{is}\sum_{k=1}^{q} b_{sk}c_{kj}\right) = \sum_{s=1}^{n}\sum_{k=1}^{q} a_{is}b_{sk}c_{kj}$$

Similarly, we could do the same for the right-hand side of the equation. Note that the elements of the product $AB$, or $D' = [d'_{ij}]_{m\times q}$ can be represented as $d'_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. Continuing, the elements in the final product $E' = [e'_{ij}]_{m\times r}$ can be written as the following.

$$\begin{aligned} e'_{ij} &= \sum_{s=1}^{q} d'_{is}c_{sj} = \sum_{s=1}^{q}\left(\left(\sum_{k=1}^{n} a_{ik}b_{ks}\right)c_{sj}\right) = \sum_{s=1}^{q}\sum_{k=1}^{n} a_{ik}b_{ks}c_{sj} \\ &= \sum_{k=1}^{q}\sum_{s=1}^{n} a_{is}b_{sk}c_{kj} = \sum_{s=1}^{n}\sum_{k=1}^{q} a_{is}b_{sk}c_{kj} = e_{ij} \end{aligned}$$

Because $e_{ij} = e'_{ij}$ for all possible combinations, $A(BC) = (AB)C$.

**In my head...**

You might wonder if we are allowed to say that the following equation is true.

$$\sum_{k=1}^{q}\sum_{s=1}^{n} a_{is}b_{sk}c_{kj} = \sum_{s=1}^{n}\sum_{k=1}^{q} a_{is}b_{sk}c_{kj}$$

When proving this property, I initially thought that my proof contained an error and spent my time reading the proof over and over again. This actually applies and visualizing it helped me. This part is not rigorous, so I didn't include it in the proof above, but you can think of rotating the entire summation by 90°.

$$\sum_{k=1}^{q}\sum_{s=1}^{n} a_{is}b_{sk}c_{kj} = \begin{matrix} a_{i1}b_{11}c_{1j} \\ \vdots \\ a_{in}b_{n1}c_{1j} \end{matrix} + \cdots + \begin{matrix} a_{i1}b_{1q}c_{qj} \\ \vdots \\ a_{in}b_{nq}c_{qj} \end{matrix}$$

Similarly, the second sum can be represented as the following.

$$\sum_{s=1}^{n}\sum_{k=1}^{q} a_{is}b_{sk}c_{kj} = \begin{matrix} a_{i1}b_{11}c_{1j} \\ \vdots \\ a_{i1}b_{1q}c_{qj} \end{matrix} + \cdots + \begin{matrix} a_{in}b_{n1}c_{1j} \\ \vdots \\ a_{in}b_{nq}c_{qj} \end{matrix}$$

Now imagine you write the entire terms out, and rotate it 90°. They are identical!

Next, let's prove the first distributive law.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$, $B = [b_{ij}]_{n\times q}$, and $C = [c_{ij}]_{n\times q}$. Consider the following equation.

$$\begin{aligned}
[a_{ij}]\left([b_{ij}] + [c_{ij}]\right) = [a_{ij}][b_{ij} + c_{ij}] &= \left[\sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj})\right] \\
&= \left[\sum_{k=1}^{n} a_{ik}b_{kj} + \sum_{k=1}^{n} a_{ik}c_{kj}\right] \\
&= \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right] + \left[\sum_{k=1}^{n} a_{ik}c_{kj}\right] \\
&= [a_{ij}][b_{ij}] + [a_{ij}][c_{ij}]
\end{aligned}$$

Therefore, the distributive property $A(B+C) = AB + AC$ applies.

We could prove the right distributive law in similar fashion.

*Proof.*

Let $A = [a_{ij}]_{n\times q}$, $B = [b_{ij}]_{m\times n}$, and $C = [c_{ij}]_{m\times n}$. Consider the following equation.

$$\begin{aligned}\left([b_{ij}] + [c_{ij}]\right)[a_{ij}] = [b_{ij} + c_{ij}][a_{ij}] &= \left[\sum_{k=1}^{n}(b_{ik} + c_{ik})a_{kj}\right] \\ &= \left[\sum_{k=1}^{n} b_{ik}a_{kj} + \sum_{k=1}^{n} c_{ik}a_{kj}\right] \\ &= \left[\sum_{k=1}^{n} b_{ik}a_{kj}\right] + \left[\sum_{k=1}^{n} c_{ik}a_{kj}\right] \\ &= [b_{ij}][a_{ij}] + [c_{ij}][a_{ij}]\end{aligned}$$

Thus, the right distributive property $(B + C)A = BA + CA$ holds.

Finally, let's prove the fourth property.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{n\times p}$. Consider the following equation.

$$\lambda(AB) = \lambda\left[\sum_{k=1}^{n} a_{ik}b_{kj}\right] = \left[\sum_{k=1}^{n} \lambda a_{ik}b_{kj}\right] = \left[\sum_{k=1}^{n} a_{ik}(\lambda b_{kj})\right] = A(\lambda B)$$

Thus, the fourth property holds.

In addition to the visual implications of a matrix and its multiplication above, matrix multiplication can be extended to systems of equations. For instance, consider the following system of equations.

$$\begin{aligned}2x + 3y &= 5 \\ -x + 7y &= 3\end{aligned}$$

Using **coefficient matrix**, we could represent the system above as the following matrix multiplication.

$$\begin{bmatrix}2 & 3 \\ -1 & 7\end{bmatrix}\begin{bmatrix}x \\ y\end{bmatrix} = \begin{bmatrix}5 \\ 3\end{bmatrix}$$

Let's take a look at more interesting facts on matrix multiplication.

**Definition 2.3.6**

A square matrix with 1 in all entries of its main diagonal and 0 in the other entries is an **identity matrix**, denoted as $I_n$ for an identity matrix with order $n \times n$. Moreover, all identity matrices satisfy the following equation for square matrix $A$ that can be multiplied.

$$AI = IA = A$$

This is called an identity matrix because when you multiply any square matrix by the identity matrix with the corresponding order, you get the square matrix itself. Consider the following examples.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

From this, we can obtain an important fact about the uniqueness of the identity matrix.

**Theorem 2.3.7**

The identity matrix $I$ for square matrices $A$ is unique.

*Proof.*

For the sake of contradiction, let matrix $A_{m\times m}$ have two unique identity matrices $I_{m\times m}$ and $J_{m\times m}$. By definition, $AI = IA = A$ and $AJ = JA = A$. Therefore, $AI = IA = AJ = JA$ and $AI = AJ$. Because $A$ is any square matrix with order $m \times m$, let $A = I$. Substituting, $II = IJ$, and $I = J$ is obtained by definition. By contradiction from the assumption that $I \neq J$, the identity matrix for a square matrix $A$ is unique.

Another interesting insight on matrix multiplication is dissecting the matrices being multiplied. Let's take a look at a simple example.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11}+a_{12}b_{21}+a_{13}b_{31} & a_{11}b_{12}+a_{12}b_{22}+a_{13}b_{32} \\ a_{21}b_{11}+a_{22}b_{21}+a_{23}b_{31} & a_{21}b_{12}+a_{22}b_{22}+a_{23}b_{32} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} b_{11} & 0 \\ b_{21} & 0 \\ b_{31} & 0 \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} 0 & b_{12} \\ 0 & b_{22} \\ 0 & b_{32} \end{bmatrix}$$

In the matrix multiplication above, we could see that to find the element in the entry $(i, j)$, it is sufficient for us to see row $i$ from the left matrix and column $j$ from the right. In the next section of the note, we will discuss different topics about matrix.

## §2.4 Additional Fundamental Topics

This section discusses frequently appearing concepts in matrices that they have their own names!

### 2.4.1 Transpose

Let's begin with transpose of a matrix.

**Definition 2.4.1**

The **transpose** of a matrix $A$ is defined by a matrix that has switched rows and columns of $A$, i.e. each element is defined as following.

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

Let's take a look at an example. You could think of it as holding the left top and bottom right of a matrix and rotating to see its back. For matrix

$A$ defined as the following,

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$$

its transpose, or $A$ transpose is defined as the following.

$$A^\intercal = \begin{bmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{bmatrix}$$

As you can see, if $A$ has the order $m \times n$, its transpose will have the order $n \times m$. Another observation that we can make is that a transpose of a row matrix is a column matrix and vice versa. Here is a quick example. Assume matrix $A$ defined as the following.

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

The transpose of $A$ can be represented as the following.

$$A^\intercal = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Notice that the transpose of transpose of $A$ is simply $A$.

$$(A^\intercal)^\intercal = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = A$$

Before formally stating our observations, here are few definitions to note.

**Definition 2.4.2**

A positive integral powers of $A$, or multiplying the matrix $n$ times can be represented as $A^n$. $A^0$ is defined to be the identity matrix $I_n$ with corresponding order $n$. A square matrix is **nilpotent** if there exists $n \in \mathbb{Z}$ such that $A^n = 0$.

Other definitions from transpose is symmetry and skew-symmetry.

**Definition 2.4.3**

For matrix $A$ and its transpose $A^\intercal$, matrix $A$ is **symmetric** if $A^\intercal = A$ and **skew-symmetric** if $A^\intercal = -A$.

An example of a symmetric matrix is the identity matrix. For more general example, matrix $A$ and $B$ below are symmetric and skew-symmetric respectively.

$$A = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}, \quad B = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

Notice that all symmetric and skew-symmetric matrices are square matrices. Moreover, the diagonal elements must be zero for a square matrix to be skew-symmetric.

Now let's discuss the properties of transpose of a matrix.

**Theorem 2.4.4**

For matrix $A$, $B$, scalar $\lambda$ and positive integer $n$, the following equations are true assuming that the addition and multiplication conditions are satisfied.

1. $(A^\intercal)^\intercal = A$
2. $(A + B)^\intercal = A^\intercal + B^\intercal$
3. $(\lambda A)^\intercal = \lambda A^\intercal$
4. $(AB)^\intercal = B^\intercal A^\intercal$
5. $(A^n)^\intercal = (A^\intercal)^n$

There are lots of properties to prove, but let's do them one by one. Below is the proof for the first property.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$. By definition, its transpose $A^\intercal := [a_{ji}]_{n\times m}$. Continuing, $\left([a_{ji}]_{n\times m}\right)^\intercal = [a_{ij}]_{m\times n} = A$. Therefore, $(A^\intercal)^\intercal = A$.

Below is the proof for the second property.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{m\times n}$. Consider the following equation.

$$\begin{aligned}(A+B)^\intercal &= \left([a_{ij}]_{m\times n} + [b_{ij}]_{m\times n}\right)^\intercal = \left([a_{ij}+b_{ij}]_{m\times n}\right)^\intercal \\ &= [a_{ji}+b_{ji}]_{n\times m} = [a_{ji}]_{n\times m} + [b_{ji}]_{n\times m} \\ &= A^\intercal + B^\intercal\end{aligned}$$

Thus, the property holds.

Let's continue our proof for $(\lambda A)^\intercal = \lambda A^\intercal$.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$. Because $(\lambda A)^\intercal = [\lambda a_{ij}]^\intercal_{m\times n} = [\lambda a_{ji}]_{n\times m} = \lambda[a_{ji}]_{n\times m} = \lambda A^\intercal$, the property holds.

We have two more left! Here is the proof for the fourth property.

*Proof.*

Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{n\times p}$. The following equation is satisfied by definitions of matrix multiplication and transpose.

$$\begin{aligned}(AB)^\intercal &= \left([a_{ij}]_{m\times n}[b_{ij}]_{n\times p}\right)^\intercal = \left(\left[\sum_{k=1}^{n} a_{ik}b_{kj}\right]_{m\times p}\right)^\intercal \\ &= \left[\sum_{k=1}^{n} a_{jk}b_{ki}\right]_{p\times m}\end{aligned}$$

Continuing from the right hand side, the following equation is obtained.

$$\begin{aligned} B^\intercal A^\intercal &= \left[b_{ij}^\intercal\right]_{p\times n}\left[a_{ij}^\intercal\right]_{n\times m} = \left[\sum_{k=1}^{n} b_{ik}^\intercal a_{kj}^\intercal\right]_{p\times m} \\ &= \left[\sum_{k=1}^{n} b_{ki}a_{jk}\right]_{p\times m} = \left[\sum_{k=1}^{n} a_{jk}b_{ki}\right]_{p\times m} \end{aligned}$$

Because both the left-hand side and the right-hand side of the equation lead to the same matrix, the property holds.

Finally, below is the proof for the corollary of the fourth property.

*Proof.*

First, notice that by the fourth property, the following equation holds.

$$\left(A^2\right)^\intercal = A^\intercal A^\intercal = (A^\intercal)^2$$

Therefore, it suffices to show that $(A^n)^\intercal = (A^\intercal)^n$ holds for $n = k+1$ if it is true for $n = k$. Assume $(A^k)^\intercal = (A^\intercal)^k$. By the fourth property, the following equation is obtained.

$$\left(A^{k+1}\right)^\intercal (A^kA)^\intercal = (A^k)^\intercal A^\intercal = (A^\intercal)^k A^\intercal = (A^\intercal)^{k+1}$$

By induction, the property holds for $k \geq 2$. For $k = 1$, it is self evident that the property holds true. Therefore, the property holds for all positive integers $n$.

For the next part of the section, we will discuss partitions.

### 2.4.2 Partitions

As always, let's start with key definitions.

**Definition 2.4.5**

As the name suggests a **submatrix** is a matrix of $A$ that is obtained by omitting certain rows or columns.

Consider matrix $A$ defined as the following.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Below are a few example of submatrices of $A$.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ \cancel{4} & \cancel{5} & \cancel{6} \\ 7 & 8 & 9 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \end{bmatrix}$$

$$A = \begin{bmatrix} \cancel{1} & \cancel{2} & \cancel{3} \\ 4 & \cancel{5} & 6 \\ \cancel{7} & \cancel{8} & \cancel{9} \end{bmatrix} \to \begin{bmatrix} 4 & 6 \end{bmatrix}$$

**Definition 2.4.6**

We say the matrix is **partitioned** if horizontal and vertical lines are used to divide the matrix into submatrices.

Here are a few examples of different partitions of $A$.

$$\left[\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ \hline 7 & 8 & 9 \end{array}\right], \quad \left[\begin{array}{c|cc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ \hline 7 & 8 & 9 \end{array}\right], \quad \left[\begin{array}{c|c|c} 1 & 2 & 3 \\ \hline 4 & 5 & 6 \\ 7 & 8 & 9 \end{array}\right]$$

From different partitions, we can obtain different **blocks**.

**Definition 2.4.7**

A **block** of a partition of $A$ is a submatrix of $A$ divided by the horizontal and vertical partition lines.

The highlighted portion of the second partition above is an example of a block. Now, why should we even care about partitions and blocks? Partitions of a matrix are important because it could be a very useful tool in computationally heavy matrix multiplication. Let's take a look at a generalized example $AB$.

$$AB = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \\ b_{41} & b_{42} & b_{43} \end{bmatrix}$$

At least for me, this multiplication looks very painful. Fortunately, using partitions, we can turn this monster into a baby monster.

$$\begin{aligned}
AB &= \left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right] \left[\begin{array}{ccc} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ \hline b_{31} & b_{32} & b_{33} \\ b_{41} & b_{42} & b_{43} \end{array}\right] \\
&= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} + \begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \\ a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} b_{31} & b_{32} & b_{33} \\ b_{41} & b_{42} & b_{43} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}b_{11}+a_{12}b_{21} & a_{11}b_{12}+a_{12}b_{22} & a_{11}b_{13}+a_{12}b_{23} \\ a_{21}b_{11}+a_{22}b_{21} & a_{21}b_{12}+a_{22}b_{22} & a_{21}b_{13}+a_{22}b_{23} \\ a_{31}b_{11}+a_{32}b_{21} & a_{31}b_{12}+a_{32}b_{22} & a_{31}b_{13}+a_{32}b_{23} \\ a_{41}b_{11}+a_{42}b_{21} & a_{41}b_{12}+a_{42}b_{22} & a_{41}b_{13}+a_{42}b_{23} \end{bmatrix} \\
&\qquad + \begin{bmatrix} a_{13}b_{31}+a_{14}b_{41} & a_{13}b_{32}+a_{14}b_{42} & a_{13}b_{33}+a_{14}b_{43} \\ a_{23}b_{31}+a_{24}b_{41} & a_{23}b_{32}+a_{24}b_{42} & a_{23}b_{33}+a_{24}b_{43} \\ a_{33}b_{31}+a_{34}b_{41} & a_{33}b_{32}+a_{34}b_{42} & a_{33}b_{33}+a_{34}b_{43} \\ a_{43}b_{31}+a_{44}b_{41} & a_{43}b_{32}+a_{44}b_{42} & a_{43}b_{33}+a_{44}b_{43} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}b_{11}+a_{12}b_{21}+a_{13}b_{31}+a_{14}b_{41} & a_{11}b_{12}+a_{12}b_{22}+a_{13}b_{32}+a_{14}b_{42} & a_{11}b_{13}+a_{12}b_{23}+a_{13}b_{33}+a_{14}b_{43} \\ a_{21}b_{11}+a_{22}b_{21}+a_{23}b_{31}+a_{24}b_{41} & a_{21}b_{12}+a_{22}b_{22}+a_{23}b_{32}+a_{24}b_{42} & a_{21}b_{13}+a_{22}b_{23}+a_{23}b_{33}+a_{24}b_{43} \\ a_{31}b_{11}+a_{32}b_{21}+a_{33}b_{31}+a_{34}b_{41} & a_{31}b_{12}+a_{32}b_{22}+a_{33}b_{32}+a_{34}b_{42} & a_{31}b_{13}+a_{32}b_{23}+a_{33}b_{33}+a_{34}b_{43} \\ a_{41}b_{11}+a_{42}b_{21}+a_{43}b_{31}+a_{44}b_{41} & a_{41}b_{12}+a_{42}b_{22}+a_{43}b_{32}+a_{44}b_{42} & a_{41}b_{13}+a_{42}b_{23}+a_{43}b_{33}+a_{44}b_{43} \end{bmatrix}
\end{aligned}$$

Well, this looks good to me! If the matrices being multiplied have higher dimension, partitioning such matrices could help in computation and reducing mistakes. Now that we have seen partitions of matrices, we could derive special forms from the ideas.

### 2.4.3 Special Forms of Matrix

As always, let's build from definitions.

**Definition 2.4.8**

A **Symmetrically partitioned matrix** refers to a partitioned matrix such that its arrangements of the blocks are diagonally symmetric. The **diagonal blocks** are the blocks that pass through the diagonal symmetry.

We could also define matrices with diagonal blocks that are not necessarily symmetrically partitioned.

**Definition 2.4.9**

A **block diagonal matrix** is a matrix with square diagonal blocks and zero on the other entries.

Below is the general form for square matrices $A_i$.

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Continuing, we could define rows and columns based on its elements.

**Definition 2.4.10**

**Zero rows** and **zero columns** are rows and columns composed of zeros respectively.

This definition is quite straightforward, and now that we have basic terms in mind, let's discuss two forms from the definitions above.

**Definition 2.4.11**

Matrix in **row echelon form (REF)** satisfies the conditions below.

- All nonzero leading entry is 1. (Side note here is that some textbooks and lectures do not require the leading entry to be 1, but I actually would like to keep it this way as this is how I initially learned it.)
- All nonzero leading entries in a nonzero row in a later row are to the right of nonzero leading entries in previous rows.
- If there are both nonzero and zero rows, all zero rows must appear after the nonzero rows above.

The leading entry of matrices in REF is known as a **pivot**. Similarly, matrices in **reduced row echelon form (RREF)** are matrices that satisfy the following conditions.

- All nonzero leading entries in a nonzero row is 1.
- If there are both nonzero and zero rows, all zero rows must appear after the nonzero rows above.
- All leading 1s in the succeeding rows appear in a column after the columns with previous leading 1s.
- The leading 1 must be the only nonzero element present in each column if there exists such 1.

Below are examples of a matrices in REF and RREF respectively.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Matrices that follow these forms will be very useful when solving linear systems of equations, but we will save that for the latter notes. Now let's discuss about triangular matrices.

Triangular matrices are a special type of square matrices that is determined by elements below the main diagonal. If all its elements below the main diagonal are zeros, we call it an **upper triangular** matrix. Similarly, if all elements above the main diagonal are zeros, we call them **lower triangular** matrices. For more formal statement, consider the following definition.

**Definition 2.4.12**

A square matrix $A = [a_{ij}]$ is **upper triangular** if $a_{ij} = 0\,\forall\, i > j$. Similarly, the matrix is **lower triangular** if $a_{ij} = 0\,\forall\, i < j$.

Let's take a look at an example. The left and right matrices below are upper triangular and lower triangular matrices respectively.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 2 & 1 & 0 \\ 0 & 3 & 2 & 0 \end{bmatrix}$$

Now that we have definitions of triangular matrices in mind, we can discuss a theorem from such matrices.

**Theorem 2.4.13**

The product of two lower triangular matrices is a lower triangular matrix, and the product of two upper triangular matrices is an upper triangular matrix.

*Proof.*

Let $A = [a_{ij}]_{n\times n}$ and $B = [b_{ij}]_{n\times n}$ be lower triangular matrices. Consider the following equation for their product matrix $C = [c_{ij}]_{n\times n}$.

$$C = \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right] = \left[\sum_{k=1}^{j-1} a_{ik}b_{kj} + \sum_{k=j}^{n} a_{ik}b_{kj}\right]$$

From the expression, it suffices to show that $\sum_{k=1}^{j-1} a_{ik}b_{kj} + \sum_{k=j}^{n} a_{ik}b_{kj} = 0$ for $i < j$. Notice that by definition, $b_{kj} = 0$ for all $k = 1$ to $k = j-1$. Similarly, $a_{ik} = 0$ for all integer $k \in [j, n]$. Therefore,

$c_{ij} = 0$ for all $i < j$, proving the first half of the theorem.

Similarly, let $A' = [a'_{ij}]_{n\times n}$ and $B' = [b'_{ij}]_{n\times n}$ be upper triangular matrices. The product matrix $C' = [c'_{ij}]_{n\times n}$ can be represented as following.

$$C' = \left[\sum_{k=1}^{n} a'_{ik}b'_{kj}\right] = \left[\sum_{k=1}^{j-1} a'_{ik}b'_{kj} + \sum_{k=j}^{n} a'_{ik}b'_{kj}\right]$$

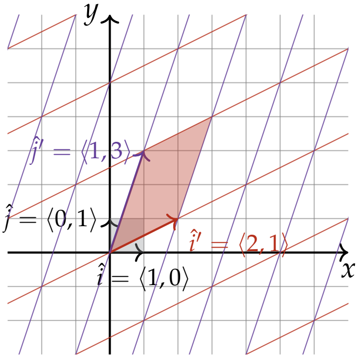
By definition, it is evident that $a'_{ik} = 0$ for integer $k \in [1, j-1]$ and $b'_{kj} = 0$ for integer $k \in [j, n]$. Therefore, for $i > j$, $c'_{ij} = 0$ and the theorem is true.

This is it for different topics on fundamentals of matrices! I know this note is quite long, but let's conclude our notes with the discussion on determinants.

## §2.5 Determinants

Finally determinants! One of the most foundational concepts in linear algebra is the determinant. Being one of the foundational concepts, I also think it is a hard concept if the intuition is excluded in the discussion due to its formula heavy nature.

As you could have noticed, a linear transformation does not only change the position of the vectors. Consider the following diagram.



Notice that from the diagram above, the gray area is scaled to the new red area. Notice that the area of the new parallelogram is $3 \cdot 4 - 1 \cdot 2 \cdot \frac{1}{2} \cdot 2 - 1 \cdot 3 \cdot \frac{1}{2} \cdot 2 - 1 \cdot 2 = 12 - 2 - 3 - 2 = 5$. Because the area which was originally 1 is scaled to 5, we can say that

$$\begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} = 5$$

holds. The determinant is only defined for square matrices, and for each square matrix, the determinants tell how much the area has changed after the linear transformation. Determinants can surely be negative, and they represent the change in orientation. With this intuition in mind, let's discuss more computational perspective of determinants.

Before we discuss the formula for determinants, let's define a notation.

**Definition 2.5.1**

The **minor** of matrix $A$, denoted as $M_{ij}$, is a submatrix of $A$ without the $i^{\text{th}}$ row and $j^{\text{th}}$ column of $A$. The **cofactor** of $a_{ij}$ is defined as $(-1)^{i+j}\det(M_{ij})$.

Consider the following examples for $A$ defined as the following.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

$M_{11}$ and $M_{24}$ are defined as the following.

$$M_{11} = \begin{bmatrix} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{bmatrix}, \quad M_{24} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}$$

Now let's define what determinants are with a recursive formula.

**Definition 2.5.2**

The **determinant** of a square matrix $A$, denoted as $\det(A)$ or $|A|$, is a value defined with the following rules.

1. $\det(A) = a_{11}$ if $A = [a_{11}]$.
2. The following equation can be used to determine $\det(A)$ for $A$ with order $n \times n$ for $n > 1$.
$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(M_{1j})$$

Let's take a look at an example with a $2 \times 2$ matrix $A = [a_{ij}]$.

$$\begin{aligned}
\det(A) &= \sum_{j=1}^{2} (-1)^{1+j} a_{1j} \det(M_{1j}) \\
&= (-1)^{1+1} a_{11} \det(M_{11}) + (-1)^{1+2} a_{12} \det(M_{12}) \\
&= (-1)^2 a_{11} a_{22} + (-1)^3 a_{12} a_{21} \\
&= a_{11} a_{22} - a_{12} a_{21}
\end{aligned}$$

Returning to our first example, we can verify that the formula holds.

$$\begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} = 2 \cdot 3 - 1 \cdot 1 = 5$$

We could also derive the formula for higher orders with many computations but I will leave them for you! I also think it is just better to use the recursive formula and the formula for $2 \times 2$ matrices instead of memorizing complex formulas, but I don't know. I think it's up to you. Below is another example.

$$\begin{aligned}
\det\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) &= \sum_{j=1}^{3} (-1)^{1+j} a_{1j} \det(M_{1j}) \\
&= a_{11} \det(M_{11}) = 1 \cdot 1 - 0 \cdot 0 = 1
\end{aligned}$$

With these in mind, we can establish a very important theorem in computing determinants.

**Theorem 2.5.3**

For any square matrix $A$, the cofactor expansion along the $i^{\text{th}}$ row and the cofactor expansion along the $j^{\text{th}}$ column equal. In other words, the following equation holds.

$$\det(A) = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det(M_{ij}) = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det(M_{ij})$$

Let's save the proof for the theorem for later as we need to show different properties of determinants before proving the theorem above. For now, let's discuss a special property of triangular matrices.

**Theorem 2.5.4**

For a triangular (upper or lower) matrix $A_{n\times n} = [a_{ij}]$, the determinant is defined as the following.

$$\det(A) = \prod_{i=1}^{n} a_{ii}$$

*Proof.*

First, the case for upper triangular matrix can be proven. Notice that $\det(A) = \sum_{i=1}^{n}(-1)^{i+1}a_{i1}\det(M_{i1}) = a_{11}\det(M_{11})$ by Theorem 2.5.3. Continuing, notice that $M_{11}$ is another upper triangular matrix by definition.

If such calculations are continued, it is evident that the last term would be $a_{nn}$. Similarly, the second to the last term would be $a_{nn}a_{n-1,n-1}$. By inductive hypothesis, it is evident that

$$\det(A) = a_{11}\det(M_{11}) = a_{11}\prod_{i=2}^{n} a_{ii} = \prod_{i=1}^{n} a_{ii}$$

and the theorem holds for upper triangular matrix.

The case for lower triangular matrix can be proven in similar manner. Note that $\det(A) = \sum_{j=1}^{n}(-1)^{1+j}a_{1j}\det(M_{1j}) = a_{11}\det(M_{11})$. With the same inductive hypothesis above, it can be concluded that

$$\det(A) = a_{11}\det(M_{11}) = a_{11}\prod_{i=2}^{n} a_{ii} = \prod_{i=1}^{n} a_{ii}$$

and the theorem holds.

As you could have guessed, we get a direct corollary from the theorem.

**Corollary 2.5.5**

The determinant of any diagonal matrix is the product of the diagonal entries.

*Proof.*

First and foremost, notice that a diagonal matrix is both an upper and lower triangular matrix. Thus by Theorem 2.5.4, the corollary holds.

Continuing from triangular matrices, we can also apply our understanding of a transpose of a matrix. Consider the following theorem.

**Theorem 2.5.6**

$\det(A) = \det(A^\intercal)$ holds for all square matrix $A$.

*Proof.*

First and foremost, notice that $A = A^\intercal$ if $A$ is a $1 \times 1$ matrix. Thus, the theorem holds for any $A_{1\times 1}$. Continuing, note that for $A_{n\times n}$ with $n > 1$, $M_{ij}$ equals $M'_{ij}$ for $A^T$. Moreover by definition, $M'_{ij} = M_{ji}$

and the following equation holds $A^\intercal = [a'_{ij}]$.

$$\begin{aligned}\det(A) &= \sum_{i=1}^{n}(-1)^{i+1}a_{i1}\det(M_{i1})\\ &= \sum_{j=1}^{n}(-1)^{1+j}a'_{1j}\det(M'_{1j})\\ &= \det\left(A^\intercal\right)\end{aligned}$$

Thus, the theorem holds.

Now that we discussed the relationships between matrices and determinants, let's conclude our discussion with products and determinants.

### 2.5.1 Determinant and Products

Before we conclude our discussion on determinant and this note, I wished to introduce one of the most important results of determinants. In this part of the section, we will try to prove that the product of the determinants is the same as the determinant of the product. To prove this result, we would need a lot of lemmas, so let's get started with the first one.

**Lemma 2.5.7**

For a matrix $A_{n\times n}$ and an elementary matrix $E_{n\times n}$ for interchanging two rows, $\det(E) = -1$ and $\det(EA) = \det(E)\det(A)$.

*Proof.*

First and foremost, the case when $E$ exchanges two consecutive rows can be proven. Let $B = FA$ where $F$ is the $n \times n$ elementary matrix that interchanges the $k^{\text{th}}$ and $k+1^{\text{th}}$ rows of $A$. Therefore, for $A = [a_{ij}]$ and $B = [b_{ij}]$, $b_{k+1,j} = a_{kj}$ holds and for the minor of $B$ as $M'_{ij}$, $M'_{i+1,j} = M_{ij}$. Moreover, notice that $F = -1$ for $n = 2$ and by inductive hypothesis with exchanging the use of $\det(A) = (-1)^2 a_{11}\det(M_{11})$ and $\det(A) = (-1)^{2n}a_{nn}\det(M_{nn})$, it is evident that $\det(F) = -1$. Thus, the following equations hold.

$$\begin{aligned}
\det(B) &= \sum_{j=1}^{n} (-1)^{k+1+j} b_{k+1,j} \det(M'_{k+1,j}) \\
&= \sum_{j=1}^{n} (-1)^{k+1+j} a_{kj} \det(M_{kj}) \\
&= -\sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det(M_{kj}) \\
&= -\det(A)
\end{aligned}$$

Thus, $\det(FA) = \det(F) \cdot \det(A)$. Continuing, notice that any exchange between two consecutive rows requires one $F$. Moreover, any exchange between two rows with a row in between requires three $F$s. Since any exchange can be performed with the combination of such two operations, it is evident that every $E$ can be represented as odd number of $F$s. Therefore, the following equations hold for odd $m$.

$$\begin{aligned}
\det(EA) = \det(F_1 F_2 \cdots F_m A) &= \det(F_1)\det(F_2)\cdots\det(F_m)\det(A) \\
&= (-1)^m \det(A) = -\det(A) \\
&= \det(E)\det(A)
\end{aligned}$$

Thus, the lemma holds.

Continuing from this lemma, we can establish an interesting corollary.

**Corollary 2.5.8**

The equation $\det(A) = 0$ holds for all square matrices $A$ with either two identical rows or columns.

*Proof.*

Let $k_1$ and $k_2$ be the identical rows of the matrix $A$. For an elementary matrix $E$ that interchanges the rows $k_1$ and $k_2$, the following equations hold by Lemma 2.5.7.

$$\det(A) = \det(EA) = \det(E)\det(A) = -\det(A)$$

Thus, $\det(A) = 0$. Similarly, if $A$ has two identical columns $k_1$ and $k_2$, the same argument applies by Theorem 2.5.6 as $\det(A) = \det(A^\intercal)$.

For our second lemma, we have another elementary row operation.

**Lemma 2.5.9**

For a matrix $A_{n\times n}$ and an elementary matrix $E_{n\times n}$ that scales $k^{\text{th}}$ row of $A$ by some nonzero scalar $\lambda$, $\det(E) = \lambda$ and $\det(EA) = \det(E)\det(A)$.

*Proof.*

First and foremost, because $E_{n\times n} = k$, it is evident that utilizing $\det(A) = (-1)^2 a_{11}\det(M_{11})$ and $\det(A) = (-1)^{2n} a_{nn}\det(M_{nn})$, appropriately will lead to $\det(E) = \lambda$ with inductive hypothesis. Moreover, consider the following equations.

$$\begin{aligned}\det(EA) &= \sum_{i=1}^{n}(-1)^{1+k}\lambda a_{ik}\det(M_{ik})\\ &= \lambda\sum_{i=1}^{n}(-1)^{1+k}a_{ik}\det(M_{ik})\\ &= \lambda\det(A) = \det(E)\det(A)\end{aligned}$$

Thus, the lemma holds.

Just like our first lemma, we can expand this second lemma with an interesting corollary.

**Corollary 2.5.10**

For any matrix $A_{n\times n}$ and scalars $\lambda$, $\det(\lambda A) = \lambda^n\det(A)$.

*Proof.*

Let $E_k$ be the elementary row operation that scales the $k^{\text{th}}$ row of $A$ by $\lambda$. Then, the following equations hold by Lemma 2.5.9.

$$\begin{aligned}\det(\lambda A) &= \det(E_n E_{n-1} \cdots E_1 A) \\ &= \det(E_n)\det(E_{n-1}) \cdots \det(E_1)\det(A) \\ &= \lambda^n \det(A)\end{aligned}$$

Thus, the corollary holds.

Below is the third lemma on determinants and elementary matrices.

**Lemma 2.5.11**

For a matrix $A_{n\times n}$ and an elementary matrix $E$ of the same order that adds $\lambda$ times row $k_1$ to row $k_2$ of $A$, then $\det(E) = 1$ and $\det(EA) = \det(E)\det(A)$.

*Proof.*

First and foremost, notice that $E$ is a triangular matrix and the elements in its diagonal are all 1. Thus by Theorem 2.5.4, $\det(E) = 1$.

Consider a new matrix $B$ that is obtained by replacing $k_2$ with $k_1$ from $A$. Then, $b_{k_2 j} = a_{k_1 j}$ and the minor $M'_{ij}$ of $B$ satisfy $M'_{k_2 j} = M_{k_2 j}$. Moreover by Corollary 2.5.8, $\det(B) = 0$. Therefore, the following equations hold.

$$\begin{aligned}&\det(EA) \\ &= \sum_{j=1}^{n} (-1)^{k_2+j}(a_{k_2 j} + \lambda a_{k_1 j})\det(M_{k_2 j}) \\ &= \sum_{j=1}^{n} (-1)^{k_2+j}(a_{k_2 j})\det(M_{k_2 j}) + \lambda \sum_{j=1}^{n} (-1)^{k_2+j}(a_{k_1 j})\det(M_{k_2 j}) \\ &= \det(A) + \lambda \det(B) = \det(A)\end{aligned}$$

Thus $\det(EA) = \det(A) = \det(E)\det(A)$.

With the three lemmas above, we can establish an important theorem that links the determinants of a matrix and an elementary matrix.

**Theorem 2.5.12**

For a matrix $A_{n\times n}$ and any elementary matrix $E$ of the same order, $\det(EA) = \det(E)\det(A)$.

*Proof.*

By definition, there exists three different elementary matrices that interchanges rows, scales a row, or add a scaled row to the other. By Lemmas 2.5.7, 2.5.9, and 2.5.11, $\det(EA) = \det(E)\det(A)$ for all elementary matrices $E$.

Now, I know it's lots of statements, but I promise we will get to the main result. Let's take a look at one last important result to get to our desired theorem! This theorem and its result include the topic of invertibility, which has not been introduced yet. Please feel free to read Section 3.3 and return or vice versa as it shouldn't be something too hard.

**Theorem 2.5.13**

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

*Proof.*

First, notice that $A$ is invertible if and only if there exists a sequence of elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 A = U$ holds for some upper triangular matrix $A$. In other words,

$$\det(E_k)\cdots\det(E_1)\det(A) = \det(U)$$

where $\det(E_i) \neq 0$ and $\det(U) \neq 0$. Moreover, there exists such elementary matrices if and only if $\det(A) \neq 0$. Thus, the theorem holds.

Now finally! Let's prove the theorem!

**Theorem 2.5.14**

For square matrices $A$ and $B$ with the same order, $\det(AB) = \det(A)\det(B)$.

*Proof.*

First, the theorem can be shown for the case where $A$ is not invertible. Notice that if $A$ is not invertible, then $AB$ must also be not invertible as if $AB$ is invertible, then there exists a matrix $C$ such that $(AB)C = I$ and $A(BC) = I$, which is contradictory. Thus by Theorem 2.5.13, $\det(AB) = 0 = 0 \cdot \det(B) = \det(A)\det(B)$ and the theorem holds for not invertible $A$.

Continuing, the case for invertible $A$ can be proven. By definition, if $A$ is invertible, then it can be represented as a product of sequence of elementary matrices $E_1, \ldots, E_k$. Therefore, $\det(AB) = \det(E_1 \cdots E_k B) = \det(E_1 \cdots E_k)\det(B) = \det(A)\det(B)$ by Theorem 2.5.12. Thus, the theorem holds.

With this important theorem in mind, let's wrap up our notes with two corollaries.

**Corollary 2.5.15**

For an invertible matrix $A$, $\det(A^{-1}) = \frac{1}{\det(A)}$.

*Proof.*

Notice that $1 = \det(AA^{-1}) = \det(A)\det(A^{-1})$ by Theorem 2.5.14. Therefore, $\det(A^{-1}) = \frac{1}{\det(A)}$.

Continuing, below is the second corollary.

**Corollary 2.5.16**

If two matrices $A$ and $B$ are similar, then $\det(A) = \det(B)$.

*Proof.*

By definition, if $A$ and $B$ are similar, then there exists an invertible matrix $P$ that satisfy $A = P^{-1}BP$. Therefore by Theorem 2.5.14 and Corollary 2.5.15, $\det(A) = \det(P^{-1})\det(B)\det(P) = \frac{1}{\det(P)} \cdot \det(B)\det(P) = \det(B)$ and the corollary holds.

This is it for determinants! I think it is an interesting way to connect the relationship between linear transformations, which we will discuss in future notes and matrix representation of their effects. In the next notes, we will discuss different applications of matrices.

## §2.6 Practice Problems

1. Consider matrices $A$ and $B$ defined as the following.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix}$$

   Compute the following matrix operations.

   (a) $A - 2B$
   (b) $AB$
   (c) $(AB - BA)^\intercal$

2. Find $a_2 B$ if $AB$ is defined as the following.

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

3. Find matrix $A$ such that the following equation holds.

$$2A - 3A^\intercal = \begin{bmatrix} -1 & -5 \\ 0 & -4 \end{bmatrix}$$

4. Show that all square matrices can be represented as the sum of symmetric and skew-symmetric matrices.

5. Show that if $AB = BA$ for nilpotent matrices $A$ and $B$, then the product $AB$ is nilpotent.

6. Show that for square matrices $A$ and $B$, $AB + (AB)^\intercal$ is symmetric.

7. For square matrices $A$ and $B$ of order $n \times n$, show that the following equation holds [Erd20].

$$\det\left(\begin{bmatrix} A & B \\ B & A \end{bmatrix}\right) = \det(A+B)\det(A-B)$$

8. Show that if a square matrix $A$ is skew-symmetric, then $A^3$ is also skew-symmetric.

Note 3

# System of Linear Equations

One of the most important applications of matrices would be in linear systems of equations. With the ideas on matrices, we can open our third eye when viewing systems of linear equations. This note is dedicated to matrices in linear equations, Gaussian Elimination, and LU decompositions.

## §3.1 Basics and Intuitions

As always, let's get started with fundamental definitions, concepts, and intuitions that will be used throughout.

First, I am pretty sure most of you are aware of what a system of equations is. However, because it is always nice to formally state our understanding, here is the definition of a linear system of equations.

**Definition 3.1.1**

A **system of linear equations** is composed of linear equations, or equations whose highest degree is 1. A system of $m$ equations and $n$ variables adhere to the following forms where $a_{ij}$ and $b_i$ are known quantities.

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\
a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\
&\vdots \\
a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= b_m
\end{aligned}$$

Below is an example of a system of linear equations.

$$\begin{aligned}
x + 2y + 3z &= 4 \\
-x + 3y + 2z &= 5
\end{aligned}$$

Notice that all equations above are linear, meaning, their degrees are 1. Now, what does it mean to find a solution to such systems?

**Definition 3.1.2**

A **solution to a system of linear equations** is a set of scalar values for unknown variables such that all linear equations are satisfied when substituted.

For instance, $(2, 1)$ is the solution to the following system of linear equations.

$$\begin{aligned}
x - 2y &= 0 \\
x + y &= 3
\end{aligned}$$

To check it, you can simply substitute to see if the equations are satisfied.

Now, if you recall from the previous note where we discussed matrix multiplication, system of linear equations can be written as a matrix multiplication.

**Definition 3.1.3**

A system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

can be written as $Ax = b$ where $A$, $x$, and $b$ are matrices defined as the following.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}, \quad b \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

When you actually perform matrix multiplication, you can see that $Ax = b$ is identical to the system of linear equations. Now that we have a basic understanding of the relationship between matrices and linear systems of equations, let's take a look at a theorem and an important corollary.

**Theorem 3.1.4**

For two distinct solutions $x_1$ and $x_2$ of $Ax = b$, $\alpha x_1 + \beta x_2$ is also a solution to the linear system given that $\alpha, \beta \in \mathbb{R}$ and $\alpha + \beta = 1$.

*Proof.*

From Theorem 2.3.5, it was shown that $\lambda(AB) = A(\lambda B)$ for scalar $\lambda$

and matrices $A$ and $B$ that can be multiplied. Substituting $\alpha x_1 + \beta x_2$, the following system is obtained.

$$\begin{aligned} A(\alpha x_1 + \beta x_2) &= A(\alpha x_1) + A(\beta x_2) = \alpha(Ax_1) + \beta(Ax_2) \\ &= \alpha b + \beta b = b \end{aligned}$$

Therefore, $(\alpha x_1 + \beta x_2)$ is another solution to the system.

From this theorem, we can establish an important corollary about the number of solutions to a system of linear equations.

**Corollary 3.1.5**

A system of linear equations can retain 0, 1, or infinitely many solutions.

*Proof.*

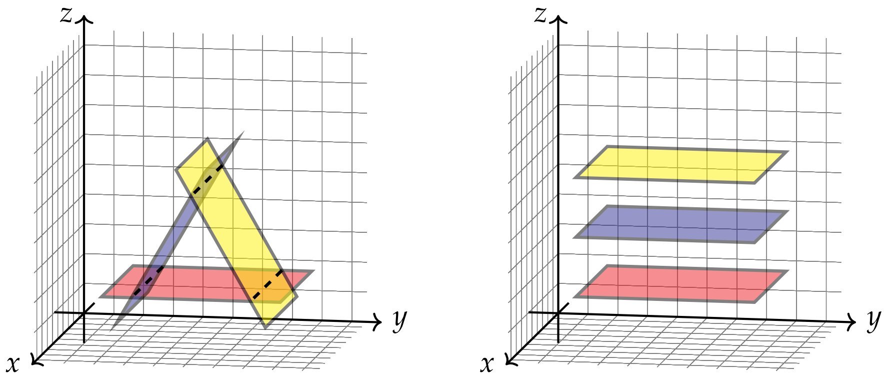
First, it is self-evident that a system can retain 0 or 1 solution. Therefore, the fact that there exist infinite solutions if there are more than 1 solution must be proven.

Let $f(\alpha) = \alpha x_1 + (1 - \alpha)x_2$ be a function that returns possible solutions where $x_1 \neq x_2$ are fixed values. It suffices to show that the function $f$ is injective as there exists infinitely many values of $\alpha$. Consider the following equations.

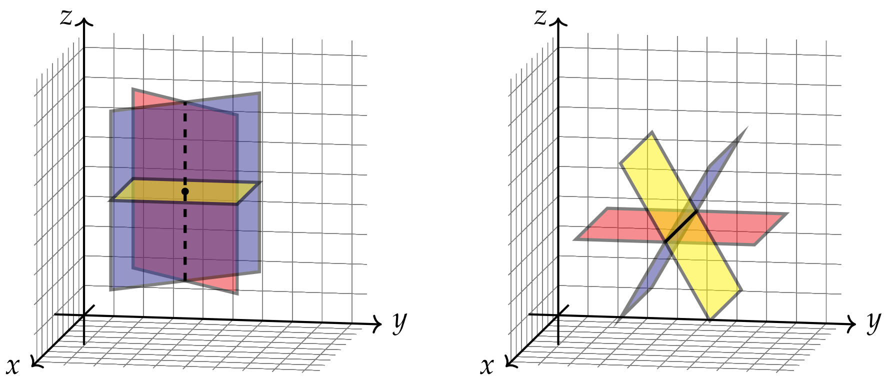
$$\begin{aligned} f(\alpha) &= \alpha x_1 + (1 - \alpha)x_2 = \alpha(x_1 - x_2) + x_2 \\ f(\alpha') &= \alpha' x_1 + (1 - \alpha')x_2 = \alpha'(x_1 - x_2) + x_2 \end{aligned}$$

If $f(\alpha) = f(\alpha')$, then $\alpha(x_1 - x_2) + x_2 = \alpha'(x_1 - x_2) + x_2$ and $\alpha = \alpha'$, proving the injectivity. Therefore, a system of linear equations can only have 0, 1, or infinitely many solutions.

In middle school, you probably saw a visual understanding of the corollary above. Let's do the same, except in 3-dimensional space. Below are two examples of when there exists no solution. Note that not all three planes intersect at a single point.




Continuing, the left diagram represents the case where there is a single solution while the right diagram shows the system with infinite solutions.




Wow, not going to lie, it took some time to draw the diagrams above. Anyhow, with that in mind, let's discuss other fundamental terms and concepts before we move on to the next section of the note.

**Definition 3.1.6**

A linear system is **consistent** if it retains at least one solution, whereas a system is **inconsistent** if there is no solution.

We could also define a system based on the column matrix $b$.

**Definition 3.1.7**

A linear system is **homogeneous** if the column matrix $b$ in the linear system $Ax = b$ is a zero matrix. If the matrix is not a zero matrix, the system is **nonhomogeneous**, or **inhomogeneous**.

Now from the definitions, we can see that all homogeneous systems are consistent. As you have guessed, we have a special name for a special solution in such cases.

**Definition 3.1.8**

**Trivial solution** is a solution where all the entries for $x$ in $Ax = b$ are zero.

With the fundamental terms in mind, let's take a look at how matrices can be used to solve complex linear system with Gaussian Elimination.

## §3.2 Gaussian Elimination

When you solve a linear system, you probably multiply constants and add or subtract equations to cancel variables. That is exactly what we will be doing in Gaussian Elimination, but with matrices. First, define how we solve a linear system in a more rigorous way.

**Definition 3.2.1**

For a system of linear equations, performing the following operations will result in a linear system with an identical solution.

1. Change the order of linear equations in which they are listed
2. Multiply a nonzero constant, or nonzero scalars, to an equation
3. Add equations from the system to form a new equation

Now before we connect this idea with matrices, let's define a matrix that we will use with the property of a linear system above.

**Definition 3.2.2**

An **augmented matrix** for a linear system $Ax = b$ is a partitioned matrix in the form $[A \mid b]$.

Consider the following system for an example.

$$\begin{aligned} x + 2y + 3z &= 4 \\ 2x + 3y + 4z &= 5 \\ 3x + 4y + 5z &= 6 \end{aligned}$$

In the linear system above, the augmented matrix will be the following matrix.

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{array}\right]$$

Now that we have an augmented matrix in mind, let's take a look at how the three properties that we discussed apply to matrices.

**Definition 3.2.3**

For an augmented matrix for a linear system, performing the following operations will result in a linear system with an identical solution as the initial system.

$R_1$. Rearrangement of the rows in the augmented matrix

$R_2$. Multiply a nonzero scalar to a row in the augmented matrix

$R_3$. Add a row multiplied by a scalar with another to replace a row with new row

Such operations are referred to as **elementary row operations**.

If you compare the operations one by one, you can notice that the two sets of properties refer to the same property. Now with the set of properties of augmented matrices in mind, here are the steps for the Gaussian Elimination algorithm.

**Theorem 3.2.4**

Execute the following steps in order for Gaussian Elimination.

1. Construct the augmented matrix.
2. Utilize the elementary row operations to transform the augmented matrix into REF.
3. Convert the augmented matrix in REF into equations of a transformed linear system, referred to as the **derived set**.
4. Back substitute to solve the system.

Let's take a look at a few examples.

**Exercise 3.2.5**

Find solution(s) to the following system of linear equations if any.

$$\begin{aligned} x + 2y + z &= 1 \\ 3x - y + z &= 6 \\ x + y - z &= -2 \end{aligned}$$

**Solution.**

First, the augmented matrix can be found from the system. Consider the matrix below.

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 3 & -1 & 1 & 6 \\ 1 & 1 & -1 & -2 \end{array}\right]$$

Using the elementary row operations, the augmented matrix can be transformed as the following.

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 3 & -1 & 1 & 6 \\ 1 & 1 & -1 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -4 & 4 & 12 \\ 1 & 1 & -1 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -4 & 4 & 12 \\ 0 & -1 & -2 & -3 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -1 & 1 & 3 \\ 0 & 1 & 2 & 3 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & -1 & 1 & 3 \\ 0 & 0 & 3 & 6 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & 1 & -1 & -3 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

Because the augmented matrix is now in REF, the transformed linear system can be rewritten.

$$\begin{aligned} x + 2y + z &= 1 \\ y - z &= -3 \\ z &= 2 \end{aligned}$$

Back substituting, it is evident that $z = 2$ and $y = -1$. Moreover, $x = 1 + 2 - 2 = 1$. Therefore, the triple $(1, -1, 2)$ is the solution to the linear system.

Try substituting to the original system! You can notice that the solution is valid. Below is the second example of using Gaussian Elimination.

**Exercise 3.2.6**

Find solution(s) to the following system of linear equations if any.

$$\begin{aligned} x + 2y + 3z &= 4 \\ y + 3z &= 5 \\ x + y &= 1 \end{aligned}$$

**Solution.**

First and foremost, the augmented matrix can be found.

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 3 & 5 \\ 1 & 1 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 3 & 5 \\ 0 & -1 & -3 & -3 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 1 & 3 & 5 \\ 0 & 0 & 0 & 2 \end{array}\right]$$

Writing the transformed augmented matrix into a system of linear equations, the following system is obtained.

$$
\begin{aligned}
x + 2y + 3z &= 4 \\
y + 3z &= 5 \\
0 &= 2
\end{aligned}
$$

Notice that $0 = 2$ can never be true. Therefore, no solution exists.

We could double check our answer without using matrices. Notice that $3z - x = 4$ and $x = 3z - 4$. Similarly, $y = 5 - 3z$. Substituting to the first equation, $3z - 4 + 10 - 6z + 3z = 4$, or $6 = 4$, which is not true. Therefore, our results align! Below is the third example of using the algorithm.

**Exercise 3.2.7**

Find solution(s) to the following system of linear equations if any.

$$
\begin{aligned}
x - y - 2z &= 3 \\
2x + y - 7z &= 3 \\
x + y - 4z &= 1
\end{aligned}
$$

**Solution.**

Consider the following augmented matrix.

$$
\left[\begin{array}{ccc|c}
1 & -1 & -2 & 3 \\
2 & 1 & -7 & 3 \\
1 & 1 & -4 & 1
\end{array}\right]
$$

Utilizing Gaussian Elimination, the augmented matrix can be transformed as following.

$$\left[\begin{array}{ccc|c} 1 & -1 & -2 & 3 \\ 2 & 1 & -7 & 3 \\ 1 & 1 & -4 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -1 & -2 & 3 \\ 2 & 1 & -7 & 3 \\ 0 & 2 & -2 & -2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -1 & -2 & 3 \\ 0 & 1 & -1 & -1 \\ 2 & 1 & -7 & 3 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|c} 1 & -1 & -2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 3 & -3 & -3 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & -1 & -2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Rewriting the system, the following equations are obtained.

$$\begin{aligned} x - y - 2z &= 3 \\ y - z &= -1 \end{aligned}$$

Notice that $y = z - 1$ and $x = y + 2z + 3 = z - 1 + 2z + 3 = 3z + 2$. Because there exists infinitely many $z$, the system has infinitely many solutions.

When writing such solutions, it could be written as the following.

$$x = \begin{bmatrix} 3z + 2 \\ z - 1 \\ z \end{bmatrix}$$

For the next section, let's discuss the inverse of a matrix and its application in linear systems.

## §3.3 The Multiplicative Inverse

As we discuss matrix multiplication, you may have wondered about matrix division. I mean it would be so nice if we could just divide the coefficient matrix to find our variables. Although there is no "division" operation per se, we could achieve similar operation with the multiplicative inverse of a matrix, or simply the inverse matrix. In other words, some solutions to linear systems can be written as $x = A^{-1}b$.

**Definition 3.3.1**

The **inverse of a matrix** is a matrix that produces an identity matrix when multiplied with the original matrix. The inverse of $A$ is denoted as $A^{-1}$.

We could see that matrices $A$ and $B$ are the inverse of each other if the following equation is satisfied.

$$AB = BA = I$$

Below is an example where $B = A^{-1}$.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} -5 & 2 \\ 3 & -1 \end{bmatrix}$$

One thing to note from the definition is that $A$ and $B$ must be square matrices. Say we have $A_{4\times 2}$ and $B_{2\times 4}$. Even if both $AB$ and $BA$ return an identity matrix, the order will be different. Therefore, for a matrix to have the inverse, it must be a square matrix.

**Definition 3.3.2**

A matrix is **invertible**, or **nonsingular**, if it retains an inverse matrix. A matrix is **singular** if it does not have an inverse matrix.

Now from the definition above, we could establish a theorem.

**Theorem 3.3.3**

If all diagonal elements of a matrix are nonzero constants, then the matrix is invertible.

*Proof.*

Consider the matrix $A_{m\times m}$ defined below.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2m} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mm} \end{bmatrix}$$

It suffices to show that there exists the inverse matrix for all values of the elements. Consider the inverse below.

$$A^{-1} = \begin{bmatrix} \frac{1}{a_{11}} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{a_{22}} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{a_{33}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{a_{mm}} \end{bmatrix}$$

Because $a_{ii}$ for integer $i \in [1, m]$ are nonzero numbers, all elements in the matrix $A^{-1}$ are defined, and the inverse matrix exists.

Now, it is important to note that not all square matrices have the inverse. For example, consider a matrix with zero row. The same row in the product matrices will become a zero row. By definition, it is not possible to make an identity matrix with a zero row.

Now that we discussed the existence, we could also investigate if the inverse matrix is unique just as a nonzero integer has a unique reciprocal.

**Theorem 3.3.4**

If there exists an inverse matrix, the inverse is unique.

*Proof.*

For the sake of contradiction, let $B$ and $C$ be different inverse of $A$.

Consider the following equation.

$$B = BI = B(AC) = (BA)C = IC = C$$

Because the fact that $B = C$ contradicts the initial assumption that $B \neq C$, all inverse matrices are unique.

Before we discuss how we actually find an inverse, let's discuss four important properties of inverse matrices.

**Theorem 3.3.5**

For invertible square matrices $A$ and $B$ and a nonzero scalar $\lambda$, the following equations hold.

1. $(A^{-1})^{-1} = A$
2. $(A^{\intercal})^{-1} = (A^{-1})^{\intercal}$
3. $(AB)^{-1} = B^{-1}A^{-1}$
4. $(\lambda A)^{-1} = (\frac{1}{\lambda})A^{-1}$

There's a lot to prove and let's do them one by one starting with the first one!

*Proof.*

Let matrix $B = A^{-1}$. Substituting, it suffices to show that $B^{-1} = A$. By definition of inverse matrix, the equation $AB = BA = 1$ holds. In other words, $A$ is the inverse of $B$ and the property holds.

Below is the proof for the second property.

*Proof.*

Notice that $(AB)^{\intercal} = B^{\intercal}A^{\intercal}$ from Theorem 2.4.4. Let matrix $B = A^{-1}$ and consider the following equation.

$$I = (AB)^{\intercal} = B^{\intercal}A^{\intercal} = \left(A^{-1}\right)^{\intercal} A^{\intercal}$$

In other words, $(A^{-1})^{\intercal}$ is the inverse of $A^{\intercal}$, or $(A^{\intercal})^{-1}$, proving the property.

Continuing, here is the proof for the third property.

*Proof.*

Consider the following equation that uses the associative property of matrix multiplication.

$$\begin{aligned}(B^{-1}A^{-1})(AB) &= B^{-1}(A^{-1}A)B = B^{-1}(I)B = B^{-1}(IB) \\ &= B^{-1}B = I\end{aligned}$$

Because $B^{-1}A^{-1}$ is the inverse of $AB$, the property is true.

Finally, here is the proof for the last property.

*Proof.*

Let $B = \lambda A$. Consider the following equation.

$$\frac{1}{\lambda}A^{-1}B = \frac{1}{\lambda}A^{-1}\lambda A = A^{-1}A = I$$

In other words, $\frac{1}{\lambda}A^{-1}$ is the inverse of $\lambda A$, proving the fourth property.

Now that we took a look at different properties of inverse matrices, let's discuss how to actually find one with tools to consider.

**Definition 3.3.6**

For a matrix $A$ that is not necessarily a square matrix, an **elementary matrix**, denoted as $E$, is a square matrix that performs an elementary row operation by multiplying with $A$.

Let's take a look at an example. Consider matrix $A$ defined below, and we want to perform elementary row operations of changing the order of rows 1 and 3 and add two times the fourth row to the second respectively.

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 2 & 0 \\ 2 & 4 & 6 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

The elementary row operations above can be complicated with words, but we can represent that with two elementary matrices respectively. Letting $E_1$ and $E_2$ be elementary matrices for each operation, we can represent the elementary row operations as the following.

$$E_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Let's multiply to see if we are right.

$$E_1 A = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 1 & 0 & 2 & 0 \\ 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad E_2 A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 2 \\ 2 & 4 & 6 & 8 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Now we can notice three observations and tips for constructing elementary matrices. For each elementary row operation, we could do the following.

**Definition 3.3.7**

**Methods to construct the elementary matrix for each elementary row operation:**

- To interchange row $a$ and $b$, write the corresponding identity matrix and interchange row $a$ and $b$ from the identity matrix.
- To multiply a nonzero scalar $k$ to $a^{\text{th}}$ row, start from the corresponding identity matrix and change the 1 from $a^{\text{th}}$ row of the identity matrix to $k$.
- To add row $a$ multiplied by the nonzero scalar $k$ to row $b$, start from the corresponding identity matrix and find the $b^{\text{th}}$ row in the matrix. Then, change the zero from the $a^{\text{th}}$ column of the row to $k$.

From this, we can notice the following theorem.

**Theorem 3.3.8**

All elementary matrices are invertible.

*Proof.*

The proof is rather simple. Because all elementary row operations can be inverted with the inverse operation, all elementary matrices are invertible.

Building onto the theorem, we could prove the uniqueness.

**Theorem 3.3.9**

For each elementary row operation, the corresponding elementary matrix is unique.

*Proof.*

For the sake of contradiction, assume there exist elementary matrices $E$ and $F \neq E$. By definition, $EA = FA$ and $A = EE^{-1}A = E^{-1}FA$. By Theorem 2.3.7, $E^{-1}F = I$ and $E = F$ is obtained by Theorem 3.3.4.

By contradiction, it is shown that elementary matrices are unique.

There are more interesting lemmas and theorems, but let's discuss them in latter notes. However, let's discuss a few theorems for each elementary row operation before moving on to finding an inverse matrix.

**Theorem 3.3.10**

For each elementary matrix for corresponding elementary row operations, the following properties hold.

1. For an elementary matrix that changes the order of two rows, the inverse of such a matrix is itself.
2. For an elementary matrix that scales a row with nonzero constant $k$, the inverse can be obtained by replacing $k$ in the elementary matrix with $\frac{1}{k}$.
3. For an elementary matrix that adds the scaled row into the other, its inverse can be obtained by replacing the nonzero scalar $k$ with $-k$.

As always, let's start by proving the first property.

*Proof.*

The inverse of an elementary matrix can be considered as undoing the elementary row operation. To undo the interchanging of two rows, the two rows can be interchanged again. In other words, because the elementary row operation $EE = I$ by Theorem 2.3.7, $E = E^{-1}$.

Continuing, here is the proof for the second property.

*Proof.*

Let elementary matrices $E$ and $F$ be matrices that scale a row by a nonzero scalar $k$ and $\frac{1}{k}$ respectively. By definition, $F(EI) = I$ and $FE = I$. Therefore, $F = E^{-1}$ and the property is shown.

Lastly, here is the proof for the final property.

*Proof.*

Similar to the proof for the second property, let elementary matrices $E$ and $F$ represent the elementary row operations that add a row scaled by nonzero constant $k$ to another and the operation that adds a row scaled by $-k$ to another respectively. By definition, $F(EI) = I$ and $F = E^{-1}$ as demonstrated above.

Now before we actually find an inverse, we need to know which square matrices have the inverse matrix.

**Definition 3.3.11**

A square matrix $A$ is invertible if and only if it can be represented as an identity matrix after sequences of elementary row operations, i.e. there exist elementary matrices $E_i$ such that the following equation is true.

$$E_k E_{k-1} \cdots E_1 A = I$$

This is a very important result that we will prove later in a note, so don't worry. Finally, here are the steps to find an inverse if there exists one.

**Definition 3.3.12**

**Steps to Find the Inverse Matrix of $\mathbf{A_{m \times m}}$:**

1. Write an augmented matrix $[A \mid I]$ for $I_m$.
2. Utilize elementary row operations on the augmented matrix for the block $A$ to be in REF.
3. If the main diagonal of the left partition contains a zero, $A$ is singular. If not, the next step can be continued.
4. Use elementary row operations again to transform the block $A$ in REF to an identity matrix. The right partition block is the inverse of $A$.

Let's take a look at a few examples.

**Exercise 3.3.13**

Find the inverse of $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$.

**Solution.**

First, the augmented matrix can be constructed.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right]$$

Consider the following elementary row operations.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 8 & 10 & 12 & 0 & 2 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 2 & 3 & 0 & 2 & -1 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{array}\right]$$

It is evident that the main diagonal will contain a zero after the transformation. Therefore, the matrix is not invertible.

**Exercise 3.3.14**

Find the inverse of the following matrix $A$.

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

**Solution.**

First and foremost, the augmented matrix can be constructed.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 & 1 & 0 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & -1 & -2 & 1 & -1 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 & -1 & 1 \end{array}\right]$$

Continuing, the left partition can be transformed into an identity matrix with elementary row operations.

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 2 & -1 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -4 & 2 & -1 \\ 0 & 0 & 1 & 2 & -1 & 1 \end{array}\right]$$

Therefore, the inverse of $A$ is the following.

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

We could actually multiply to see if we get $I_3$.

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

Indeed it is the inverse! In the next section of the note, we will discuss LU decomposition before completing our note on linear systems.

## §3.4 LU Decomposition

The last popular method to solve a linear system that we will discuss is LU decomposition. As always, let's start by discussing fundamental definitions.

**Definition 3.4.1**

**LU decomposition** is the process of factorizing an invertible matrix into the product of lower and upper triangular matrices where the diagonal elements of the lower triangular matrix are 1.

Before we look at an important theorem and how it can be used to solve a linear system, here is an example of LU decomposition for matrix $A$ defined as the following.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 6 & 9 \\ 3 & 10 & 18 \end{bmatrix}$$

From here, notice that matrix $A$ can be transformed to upper triangular matrix with the following elementary row operations of $R_3$.

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 6 & 9 \\ 3 & 10 & 18 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 3 & 10 & 18 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 4 & 9 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix}$$

Notice that the operations above can be represented as the matrix multiplication with $R_3$ that are lower triangular matrices.

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

Therefore, the following equation is obtained.

$$
\begin{aligned}
A &= \left( \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \right) \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{bmatrix}
\end{aligned}
$$

$A$ is now represented as the product of lower and upper triangular matrices! Now that we know how LU decomposition works, let's take a look at an important lemma on triangular matrices to prove a theorem on LU decomposition.

**Lemma 3.4.2**

The inverse of upper and lower triangular matrices are upper and lower triangular matrices respectively.

*Proof.*

First and foremost, consider an invertible upper triangular matrix $A_{m\times m} = [a_{ij}]$ and its inverse $B = [b_{ij}]$. By definition, its product $C = [c_{ij}]$ can be represented as $C = [\sum_{k=1}^{m} a_{ij}b_{ji}]$. Because $c_{ij} = 0$ for $i \neq j$ and $c_{ij} = 1$ for $i = j$, the values of $b_{ji}$ can be found.

By definition, $a_{ij} = 0$ for $i > j$. Therefore, $c_{ij} = 0$ for $i > j$. Similarly, if $b_{ji} = 0$ for $j > i$, then $c_{ij} = 0$ for $i \neq j$. Continuing, if $b_{ij} = \frac{1}{a_{ij}}$ for $i = j$, $B$ can be constructed, which is an upper triangular matrix. Because the inverse matrix is unique, the inverse of an upper triangular matrix is an upper triangular matrix if there exists one.

Similarly, the fact that the inverse of an lower triangular matrix is a lower triangular matrix can be proven. Let $A'_{m\times m} = [a'_{ij}]$ be a lower triangular matrix and $B' = [b'_{ij}]$ be its inverse. By definition, their product $C' = [c'_{ij}] = [\sum_{k=1}^{m} a'_{ij} b'_{ji}]$ is the identity matrix $I_m$. Thus, it suffices to show that $c'_{ij} = 0$ for $i \neq j$ and $c'_{ij} = 1$ for $i = j$.

Note that by definition, $a'_{ij} = 0$ for $i < j$. Therefore, $c'_{ij} = 0$ for $i < j$. Similarly, if $b'_{ji} = 0$ for $i > j$, and $b'_{ij} = \frac{1}{a'_{ij}}$ for $i = j$, then the valid inverse matrix $B$ is constructed. Because $B$ is a lower triangular matrix and the inverse matrix is unique, the inverse of a lower triangular matrix is a lower triangular matrix.

With the lemma in mind, let's prove the following theorem.

**Theorem 3.4.3**

An invertible matrix has a LU decomposition if and only if multiplying a sequence of lower triangular elementary matrices for elementary row operations can transform the matrix into an upper triangular matrix.

*Proof.*

First, the statement that a nonsingular matrix has a LU decomposition if multiplying a sequence of lower triangular elementary matrices for elementary row operations can transform the matrix into an upper triangular matrix can be proven. Let $A$ be an invertible matrix that can be represented as the following.

$$E_{n,n-1} \cdots E_{31} E_{21} A = U$$

for lower matrices $E$ and an upper matrix $U$. Notice that $A$ can be written as the following.

$$A = \left( E_{21}^{-1} E_{31}^{-1} \cdots E_{n,n-1}^{-1} \right) U$$

From Lemma 3.4.2, the inverses of the elementary matrices are lower triangular matrices. Moreover, by Theorem 2.4.13, the product of the lower triangular matrices is also a lower triangular matrix.

Therefore, for a lower triangular matrix $L$, $A$ can be written as the following.

$$A = LU$$

Therefore, the first half of the statement holds.

For the second half of the statement, consider the following equation.

$$I = AA^{-1} = LUA^{-1} = L\left(UA^{-1}\right)$$

Therefore, $L$ is nonsingular. From the proof for Lemma 3.4.2, it is evident that the diagonal elements of an invertible lower triangular matrix are nonzero as the diagonal elements in the inverse cannot be defined if a zero is present in the diagonal elements of the lower triangular matrix.

Let $D$ be the diagonal matrix formed with the diagonal elements of $L$ and define $L'$ such that the equation $L'D = L$ is satisfied. Moreover, let $U' = DU$. Thus, the following equation is obtained.

$$A = LU = L'DU = L'U'$$

Recall that $L$ is invertible. Because $L^{-1}$ exists and $(L'D)^{-1} = D^{-1}L'^{-1}$ exists, $L'$ is also invertible. In other words, $L'^{-1}A = U'$. Therefore, it suffices to show that $L'^{-1}$ is a product of lower triangular elementary matrices. By definition of $D$, the diagonal elements of $L'$ are 1. Thus, it can be interpreted as a sequence of elementary row operations. In other words, it can be represented as a product of lower triangular elementary matrices, proving the latter half of the theorem.

Now that we discussed what LU decomposition is, let's see how we can use it to solve a linear system.

### 3.4.1 Application in Linear Systems

When we apply LU decomposition in solving linear systems, two relatively easy systems must be solved. Consider the system $Ax = b$ where $A = LU$. First, $y$ must be found from the following system with forward substitution.

$$Ly = b$$

Then, we can solve for $x$ with back substitution with the following system.

$$Ux = y$$

The good news is that both systems are relatively easy to solve. The reason is simple. Consider the following equation.

$$Ax = LUx = b$$

Let $y = Ux$, then $Ly = b$. If we find $Ly = b$, then we can find $Ux = y$ as we know $y$. This way, we can find $x$ while exploiting the triangular form. Let's take a look at an example.

**Exercise 3.4.4**

Solve the following system of linear equations.

$$x - 2y + z = 2$$
$$2x + y - z = 1$$
$$3x - y + 2z = 15$$

**Solution.**

First and foremost, the linear system can be represented as the matrix multiplication below.

$$\begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & -1 \\ 3 & -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 15 \end{bmatrix}$$

Continuing, the coefficient matrix $A$ can be transformed as the following.

$$\begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & -1 \\ 3 & -1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 1 \\ 0 & 5 & -3 \\ 3 & -1 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & 1 \\ 0 & 5 & -3 \\ 0 & 5 & -1 \end{bmatrix}$$
$$\to \begin{bmatrix} 1 & -2 & 1 \\ 0 & 5 & -3 \\ 0 & 0 & 2 \end{bmatrix}$$

Using elementary matrices, the transformation above can be written as the following.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} A = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 5 & -3 \\ 0 & 0 & 2 \end{bmatrix}$$

Completing the LU decomposition, $A$ can be represented as the product of the following lower and upper triangular matrices.

$$A = \left( \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \right) \begin{bmatrix} 1 & -2 & 1 \\ 0 & 5 & -3 \\ 0 & 0 & 2 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \\ 0 & 5 & -3 \\ 0 & 0 & 2 \end{bmatrix}$$

Continuing, the following system could be solved.

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 15 \end{bmatrix}$$

The matrix multiplication above can be written as the following system of linear equations.

$$\begin{aligned} x' &= 2 \\ 2x' + y' &= 1 \\ 3x' + y' + z' &= 15 \end{aligned}$$

Substituting, $y' = 1 - 4 = -3$ and $z' = 15 - 6 + 3 = 12$. Thus, the final system can be written as the following.

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 5 & -3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \\ 12 \end{bmatrix}$$

Writing in linear system, $x$, $y$, and $z$ can be found.

$$\begin{aligned} x - 2y + z &= 2 \\ 5y - 3z &= -3 \\ 2z &= 12 \end{aligned}$$

Therefore, $z = 6$, $y = 3$, and $x = 2$.

With LU decomposition, we completed our discussion on solving systems of linear equations! Below are some practice problems for the topics that we covered.

## §3.5 Practice Problems

1. Show that for an invertible square matrix $A$, the system $Ax = b$ is always consistent.

2. Show that the following system cannot have a unique solution for all $k \in \mathbb{R}$.

$$\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ k \end{bmatrix}$$

3. Use Gaussian Elimination to find the solution for the following system.

$$\begin{aligned} x + y + z &= 10 \\ 3x + 4y - z &= 2 \\ -x - 3y + z &= 4 \end{aligned}$$

4. Solve the same system above using LU decomposition.

5. Find the inverse of the following matrix.

$$\begin{bmatrix} 1 & 1 & 1 \\ 3 & 4 & -1 \\ -1 & -3 & 1 \end{bmatrix}$$

   Then, find the values of $x, y, z$ that satisfy the following system.

$$\begin{bmatrix} 1 & 1 & 1 \\ 3 & 4 & -1 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 4 \end{bmatrix}$$

Note 4

# Matrices and Vector Spaces

Now that we discussed matrices and their application in linear systems, let's return to the vector spaces to see what matrices represent in vector spaces. This note will introduce more concepts on vector spaces and matrices like row spaces and rank of a matrix and linear transformation.

## §4.1 Row Space and Rank of a Matrix

As always, let's get started with the definition.

**Definition 4.1.1**

The span of rows of a matrix $A_{m\times n}$ in $\mathbb{R}^n$ is the **row space** of the matrix $A$. The dimension of the row space is known as the **row rank**.

We can notice that by definition, the row space is a subspace of $\mathbb{R}^n$. Let's take a look at an example. Consider the following matrix $A$.

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

The row space $\mathrm{row}(A)$ is represented by $\mathrm{span}\{(1,2),(3,4),(5,6)\}$ and its rank $\mathrm{r}(A) = \dim(\mathrm{row}(A)) = \dim(\mathbb{R}^2) = 2$. Connecting this with our knowledge in matrices, we can establish the following theorem.

**Theorem 4.1.2**

A basis for the row space of a matrix $A$ in REF is the set of nonzero rows, and the number of such rows is the row rank.

*Proof.*

By definition, the nonzero rows span the row space since the zero rows do not contribute the vector addition. Therefore, it suffices to show that the set of nonzero vectors is linearly independent.

By definition of matrices in REF, there exist no two rows with the same initial nonzero entry, implying its uniqueness. Therefore, by induction, all coefficients in the linear combination of the set of nonzero vectors must be zero. Thus, the set of nonzero rows is a basis for the row space of $A$. Moreover, by definition, the rank of $A$ is the number of nonzero rows in $A$.

Continuing, we can also establish the following theorem.

**Theorem 4.1.3**

If $A$ and $B$ are matrices that can be obtained by applying elementary row operations to each other, then $\mathrm{row}(A) = \mathrm{row}(B)$.

*Proof.*

Let $A = \{v_1, \ldots, v_n\}$ and $B$ be a matrix that is obtained by a single row operation from $A$. For some $j, k \in [1, n]$ such that $j \neq k$, it is evident that $\mathrm{span}\{v_1, \ldots, v_j, \ldots, v_k, \ldots, v_n\} = \mathrm{span}\{v_1, \ldots, v_k, \ldots, v_j, \ldots, v_n\}$. Moreover, for some scalar $\lambda$, $\mathrm{span}\{v_1, \ldots, \lambda v_k, \ldots, v_n\} = \{v_1, \ldots, v_n\}$ and $\mathrm{span}\{v_1, \ldots, v_k + \lambda v_j, \ldots, v_n\}$ by statement four in Theorem 1.3.9. Continuing with induction, the same applies to $B$ that is obtained by a sequence of elementary operations from $A$.

Thus, the theorem holds.

From the two theorems above, we can notice that for any matrix $A$, we can find its basis and rank by transforming it into REF with elementary row operations and observing the nonzero rows. Below is a quick example.

**Exercise 4.1.4**

Find the basis, row space, and rank of the following matrix.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

**Solution.**

First and foremost, the matrix can be transformed into REF with elementary row operations.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 7 & 8 & 9 \end{bmatrix}$$
$$\to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -6 & -12 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore, the basis for the row space, row space, and rank of the matrix are $\{(1,2,3),(0,1,2)\}$, $\text{span}\{(1,2,3),(0,1,2)\}$, and 2 respectively.

We could also consider the following theorem.

**Theorem 4.1.5**

For a vector space $V$ and its basis $S = \{v_1, \ldots, v_n\}$ where $u, \{u_1, \ldots, u_m\} \in V$, $u \in \text{span}\{u_1, \ldots, u_m\}$ if and only if a linear combination of a coordinate representation of $\{u_1, \ldots, u_m\}$ with respect to $S$ is the coordinate representation of $u$ with respect to $S$.

*Proof.*

First and foremost, notice that $u \in \text{span}\{u_1, \ldots, u_m\}$ if and only if for some constants $c_k$, $u = \sum_{k=1}^{m} c_k u_k$. Let the coordinate representation of $u_i$ with respect to $S$ be $[a_{i1} \cdots a_{in}]$. Substituting, it could be inferred that $u \in \text{span}\{u_1, \ldots, u_m\}$ holds if and only if $u$ can be represented as the following.

$$u = \sum_{i=1}^{m} \sum_{j=1}^{n} c_i (a_{ij} v_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} (c_j a_{ji}) v_i$$

In other words, $u \in \text{span}\{u_1, \ldots, u_m\}$ if and only if the coordinate representation of $u$ with respect to $S$ is the following.

$$[u]_S = \begin{bmatrix} \sum_{j=1}^{m} (c_j a_{j1}) \\ \sum_{j=1}^{m} (c_j a_{j2}) \\ \vdots \\ \sum_{j=1}^{m} (c_j a_{jn}) \end{bmatrix} = \sum_{k=1}^{m} c_k \begin{bmatrix} a_{k1} \\ \vdots \\ a_{kn} \end{bmatrix}$$

Therefore, the theorem holds.

Continuing from this theorem, we can establish another theorem.

**Theorem 4.1.6**

For an $n$-dimensional vector space $V$ and its basis $B$, consider a subset $U = \{u_1, \ldots, u_m\} \subset V$. Construct a matrix $A = [a_{ij}]$ such that the coordinate representation of $u_i$ with respect to $B$ is the $i^{\text{th}}$ row of $A$. For some set of vectors $S \subset V$, if a basis for the row space of $A$ is the coordinate representation of vectors in $S$ with respect to $B$, then $S$ is a basis for $\text{span}(U)$.

*Proof.*

First and foremost, assume that the coordinate representation of vectors in $S$ with respect to $B$ is a basis for $\text{row}(A)$. By definition, the span of such representations is the row space of $A$. Notice that by Theorem 4.1.5, an arbitrary vector $v \in \text{span}(S)$ if and only if a linear combination of a coordinate representation of $S$ with respect to $B$ is the coordinate representation of $v$ with respect to $B$. In other words, $v \in \text{span}(S)$ if and only if the coordinate representation of $v$ is in $\text{row}(A)$. By Theorem 4.1.5 again, the statement implies that $v \in \text{span}(U)$. Thus, $v \in \text{span}(S)$ if and only if $v \in \text{span}(U)$ and $\text{span}(S) = \text{span}(U)$.

To show that $S$ is a basis for $\text{span}(U)$, or $\text{span}(S)$, it suffices to show that $S$ is linearly independent. Notice that because the set of coordinate representations of $S$ is a basis for $\text{row}(A)$, the set of coordinate representations of $S$ is linearly independent. Moreover, by the proof for the Theorem 4.1.5, $S$ is also linearly independent. Thus, $S$ is a basis for $\text{span}(U)$.

There are lots of theorems and results! But here is another one that connects vectors and matrices.

**Theorem 4.1.7**

For $S \subset \mathbb{R}^n$ such that $|S| = k$, construct matrix $A_{k\times n}$ such that the rows are each vectors in $S$. $S$ is linearly independent if and only if $\mathrm{r}(A) = k$.

*Proof.*

First and foremost, notice that by definition, $S$ is linearly independent if and only if it is a basis for its span. Moreover, by construction, $\mathrm{span}(S) = \mathrm{row}(A)$ and $S$ is linearly independent if and only if it is a basis for the row space of $A$. Continuing, $S$ is a basis for the row space of $A$ if and only if $\mathrm{r}(A) = |S|$ by definition. Therefore, $S$ is linearly independent if and only if $\mathrm{r}(A) = |S|$.

To be honest, I procrastinated a bit when I was learning to prove the theorems above. For some reason, the proofs were hard to understand and less intuitive than the proofs that I have done through olympiad style problems. I think what helped me to understand the proofs is to just continue looking at it over and over again until I fully knew the terms. Because the proofs above prioritize understanding of the basic concepts over creativity, I think procrastinating helped me fully understand the proofs.

So far, we discussed about row spaces and row ranks. How about columns? As you may have guessed, we could do the same for columns!

**Definition 4.1.8**

The span of the columns of a matrix $A_{m\times n}$ is the **column space** of $A$, denoted as $\mathrm{col}(A)$. The dimension of such space is **column rank** of $A$.

We could notice that because rows and columns of a matrix are not necessarily the same, the row space and column space of a matrix are generally different. One special example in which they are the same would be identity matrices.

Now when we discussed about row rank of a matrix $A$, we used the notation $\text{row}(A)$. That is because the row rank and column rank of a matrix are always equal.

**Theorem 4.1.9**

The row rank and column rank of a matrix $A$ are always equal.

*Proof.*

First and foremost, notice that by definition, $\text{row}(A^\intercal) = \text{col}(A)$ and $\text{row}(A) = \text{col}(A^\intercal)$. Therefore, the following equations are true.

$$\dim(\text{row}(A^\intercal)) = \dim(\text{col}(A))$$
$$\dim(\text{row}(A)) = \dim(\text{col}(A^\intercal))$$

Notice that if $\dim(\text{col}(A)) \leq \dim(\text{row}(A))$, then the following expression holds.

$$\dim(\text{row}(A)) = \dim(\text{col}(A^\intercal)) \leq \dim(\text{row}(A^\intercal)) = \dim(\text{col}(A))$$

Therefore, if $\dim(\text{col}(A)) \leq \dim(\text{row}(A))$, then $\dim(\text{row}(A)) \leq \dim(\text{col}(A))$ and $\dim(\text{col}(A)) = \dim(\text{row}(A))$. Thus, it suffices to show that $\dim(\text{col}(A)) \leq \dim(\text{row}(A))$.

Consider a matrix $A = [a_{ij}]$ that has the order $m \times n$. Notice that each row can be represented as an $n$-tuple. Let $U = \{u_1, \ldots, u_k\} \subset \mathbb{R}^n$ be a basis for $\text{row}(A)$. By definition, each row $A_i$ of $A$ can be written as the following for some constants $c_{ij}$.

$$A_i = \sum_{j=1}^{k} c_{ij} u_j$$

Moreover, notice that each element can be written as the following.

$$a_{ij} = \sum_{n=1}^{k} c_{in} u_{nj}$$

Continuing, let $V = \{v_1, \ldots, v_k\}$ be the set of such coefficients where $v_i$ is defined as the following.

$$v_i = \begin{bmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{mi} \end{bmatrix}$$

Therefore, each column of $A$ can be written as the following $m$-tuple.

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} = \sum_{n=1}^{k} \begin{bmatrix} c_{1n} \\ c_{2n} \\ \vdots \\ c_{mn} \end{bmatrix} u_{nj} = \sum_{n=1}^{k} u_{nj} v_n$$

Because each column is in the span of $V$, $\dim(\text{col}(A)) \leq k = \dim(\text{row}(A))$. Thus, the theorem holds.

Now that we have column space in mind, let's return to linear systems. Let's first start with an example before looking at the theorem.

$$\begin{aligned} x + 2y + 3z &= 4 \\ x - y + z &= -1 \\ 4x + 3y + 2z &= 1 \end{aligned}$$

Notice that we can write the system as the following.

$$x \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} + y \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} + z \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix}$$

This splits the coefficient matrix $A$ in $Ax = b$ to its columns. Now the left-hand side can be interpreted as the linear combination of the columns of $A$, or an element in $\text{col}(A)$. In other words, the solution exists if $b \in \text{col}(A)$ and the system is inconsistent if $b \notin \text{col}(A)$. If $b \in \text{col}(A)$, then the rank of the augmented matrix $[A \mid b]$ will equal $\mathrm{r}(A)$ since $\mathrm{col}(A) = \mathrm{col}([A \mid b])$. However, if $b \notin \mathrm{col}(A)$, then $\mathrm{r}([A \mid b]) = \mathrm{r}(A) + 1$ since we need an extra vector for the basis for the column space. We can now write this idea as a theorem.

**Theorem 4.1.10**

A linear system $Ax = b$ has at least one solution if and only if $\mathrm{r}(A) = \mathrm{r}([A \mid b])$.

*Proof.*

For a linear system $Ax = b$, let $A_{m\times n} = [a_{ij}]$, $A_i$ be the $i^{\text{th}}$ column of $A$, $x = [x_i]$, and $b = [b_i]$ where $x$ and $b$ are $m$-tuples. Notice that the system can be represented as the following.

$$\sum_{k=1}^{n} x_k A_k = b$$

The system will hold if and only if $b$ is in the span of the columns of $A$. Continuing, $\mathrm{r}(A) = \mathrm{r}([A \mid b])$ if and only if $b$ is in the span of the existing columns of $A$. Thus, a solution to the linear system exists if and only if $\mathrm{r}(A) = \mathrm{r}([A \mid b])$.

Though satisfying, the theorem may not be significantly useful since we anyways have to write the matrices in REF. However, we can notice another interesting result.

**Theorem 4.1.11**

For a consistent system with $n$ variables, if $\mathrm{r}(A) = k$, then the final solutions can be written with $n - k$ free variables.

*Proof.*

First and foremost, notice that because the system $Ax = b$ is consistent and $\mathrm{r}(A) = k$, $\mathrm{r}([A \mid b]) = k$. By Theorem 4.1.2, there exist $k$ nonzero rows, or $k$ variables with set values for $\mathrm{r}([A \mid b])$ in REF.

Therefore, there exist $n - k$ free variables.

Continuing with this theorem, we can set a corollary for homogeneous systems.

**Corollary 4.1.12**

For a homogeneous system $Ax = \mathbf{0}$ with $n$ variables, there exist nontrivial solutions if and only if $\mathrm{r}(A) \neq n$.

*Proof.*

By Theorem 4.1.11, there exist free variables if $\mathrm{r}(A) < n$, which are not necessarily trivial solutions. However, if $\mathrm{r}(A) = n$, then there exist zero free variables and a unique solution, which is the trivial solution.

Continuing from the system, we could also apply our knowledge to invertibility of a square matrix. Now before we prove an important theorem on the relationship between rank and invertibility, let's prove a property that we couldn't prove in our previous notes. In our previous notes, we proved the uniqueness of the inverse of a square matrix. Moreover, we also knew that if $AB = BA = I$, then $A$ and $B$ are inverse of each other. What if we truncate the equation to $AB = I$? The answer is that they imply the same thing! Indeed for two square matrices $A$ and $B$ with the same order, if $AB = I$, then $BA = I$. To prove this, let's first prove two lemmas.

**Lemma 4.1.13**

For two square matrices $A$ and $B$ with order $n \times n$ such that $AB = I, \mathrm{r}(A) = n$.

*Proof.*

First and foremost, notice that for all integers $i, j \in [1, n]$, the following equation is satisfied for $A_i$ the $i^{\text{th}}$ row of $A$ and $B_j$ the $j^{\text{th}}$ column of $B$.

$$A_i B_j = I_{ij}$$

Moreover, notice that if $\sum_{i=1}^{n} c_i A_i = \mathbf{0}$ for some constants $c_i$, then $\sum_{i=1}^{n} c_i A_i B_j = \mathbf{0}$ for any choice of $j$. By definition, $A_i B_j = 1$ if and only if $i = j$. In other words, if $\sum_{i=1}^{n} c_i A_i = \mathbf{0}$, then $c_i = 0$ for all integer $i \in [1, n]$. Therefore, the rows of $A$ are linearly independent. By definition, the rows compose the basis of $\text{row}(A)$ and $\text{r}(A) = n$.

Continuing, let's prove the second lemma.

**Lemma 4.1.14**

For a square matrix $A_{n \times n}$, if $\text{r}(A) = n$, then there exists a matrix $B$ where $BA = I$.

*Proof.*

Notice that because $\text{r}(A) = n$, none of the elements in the main diagonal for $A$ in REF is zero after applying series of elementary row operations. From $A$ in REF, more elementary row operations can be applied to convert $A$ in REF to an identity matrix. Thus, the lemma holds.

Now that we have the two lemmas in mind, let's prove the main result.

**Theorem 4.1.15**

For square matrices $A$ and $B$ with the order $n \times n$, if $AB = I$, then $BA = I$.

*Proof.*

First and foremost, by Lemma 4.1.13, $\text{r}(A) = n$. Moreover by Lemma 4.1.14, there exists matrix $C$ such that $CA = I$. Continuing, $C = CI = CAB = IB = B$. Thus, if $AB = I$, then $BA = I$.

This is the proof for the definition that we skipped in the earlier notes! Now with the two lemmas and the theorem in mind, we can derive the following result.

**Theorem 4.1.16**

A square matrix $A_{n\times n}$ is invertible if and only if $\mathrm{r}(A) = n$.

*Proof.*

First and foremost, the first half of the theorem can be proven. By definition, if $A$ is invertible, then there exists a square matrix of the same order $B$ such that $AB = I$. By Lemma 4.1.13, $\mathrm{r}(A) = n$.

Continuing, by Lemma 4.1.13, if $\mathrm{r}(A) = n$, then there exists a matrix $C$ such that $CA = I$. By Theorem 4.1.15 and definition of the inverse matrix, $A$ is invertible.

This is it for row space and rank of a matrix! In the next section, we will discuss about linear transformations, an important topic in Linear Algebra.

## §4.2 Linear Transformations

When we discussed about matrices, we saw that they kind of "change" the vector space. When I first learned about matrices in an academy as a middle school student, all I had in mind was "this is an interesting way of organizing items!" I knew what vectors and matrices were, but I didn't really know them. That changed until I learned about linear transformations. Of course, I wouldn't expect my middle school self to learn all these and truly understand what matrices represent in a vector space, but this moment, or section really changed my understanding. Like all the other sections, this is an interesting one, and let's get started!

### 4.2.1 Brief Review of Functions

Before we get into linear transformations, let's review about functions. I think this would be a good refresher on the terms.

**Definition 4.2.1**

A **function** is a special **relation** where each input from its domain has exactly one corresponding output.

For instance, a function $f : X \to Y$ is a special relation whose domain is $X$ and whose corresponding outputs lie in $Y$. As you probably know, there are special terms for $X$ and $Y$.

**Definition 4.2.2**

The set of all inputs of a function is called **domain** of the function. The set of all corresponding outputs is known as **image**, or range, of the function denoted as $\Im(f)$.

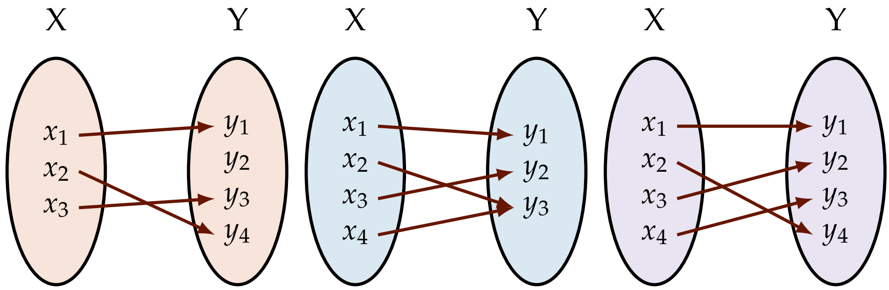
You might be familiar with the term **range**, but **image** and **range** refer to the same thing, and **image** is more commonly used in linear algebra and other courses. Now let's review about bijection, injection, and surjection.

**Definition 4.2.3**

A function is **injective**, or one-to-one, if each element in the domain maps to a unique element in the image. A function is **surjective**, or onto, if the image and codomain are identical. A **bijective** functions are functions that are both injective and surjective.

For more visual representation, below is a diagram that manifests the definition above.

| **Injective Function** | **Surjective Function** | **Bijective Function** |
|---|---|---|
| If $f(x_1) = f(x_2)$, then $x_1 = x_2$ | $\forall y_i \in Y, \exists x_j \in X$ s.t. $f(x_j) = y_i$ | $\forall y_i \in Y, \exists! x_j \in X$ s.t. $f(x_j) = y_i$ |




Below is a fun problem on bijectivity.

**Exercise 4.2.4**

Show that $f : \mathbb{R} \to \mathbb{R}$ is bijective if $f(f(x)) = x$.

**Solution.**

First, the injectivity can be proven. Let $f(a) = f(b)$ and consider the following equation.

$$a = f(f(a)) = f(f(b)) = b$$

Therefore, if $f(a) = f(b)$, then $a = b$ and the function is injective. Continuing, let $f(x) = y$. Substituting, $f(y) = x$ is true for all $x \in \mathbb{R}$. Because the image is $\mathbb{R}$, the function is bijective.

Now that we reviewed about functions, let's discuss linear transformation.

### 4.2.2 Basic Terms and Notations

If we have functions over different domains, can we have a function whose domain and range are vector spaces? The answer is yes, and we have a special name for such functions.

**Definition 4.2.5**

A function whose domain and range are defined over vector spaces is known as a **transformation**, denoted as $T$.

A transformation that maps a vector space $V$ to $W$ is denoted as $T : V \to W$. Just like how we have $f(x)$, we do the same like $T(v)$ for $v \in V$, but it is more common to write $Tv$. Continuing, the definitions for image, injection, surjection, and bijection all apply in similar fashion.

There are many different transformations, but some transformation moves the vector space in "parallel" in a way that preserves the fundamental definition of vector addition and scalar multiplication.

**Definition 4.2.6**

A **linear transformation** is a transformation that preserves the definition of vector addition and scalar multiplication of the input vector space. In other words, for scalars $\alpha$ and $\beta$ and input vectors $v_1$ and $v_2$, a transformation $T$ is linear if the following equation holds.

$$T(\alpha v_1 + \beta v_2) = \alpha T(v_1) + \beta T(v_2)$$

One way to think of the equation above is preservation of the **linear** combination. Let's take a look at an example.

**Exercise 4.2.7**

Show that a transformation defined as $T(v) = 2v$ is linear.

**Solution.**

Consider the following equation for some scalars $\alpha$ and $\beta$.

$$T(\alpha v_1 + \beta v_2) = 2(\alpha v_1 + \beta v_2) = 2\alpha v_1 + 2\beta v_2 = \alpha T(v_1) + \beta T(v_2)$$

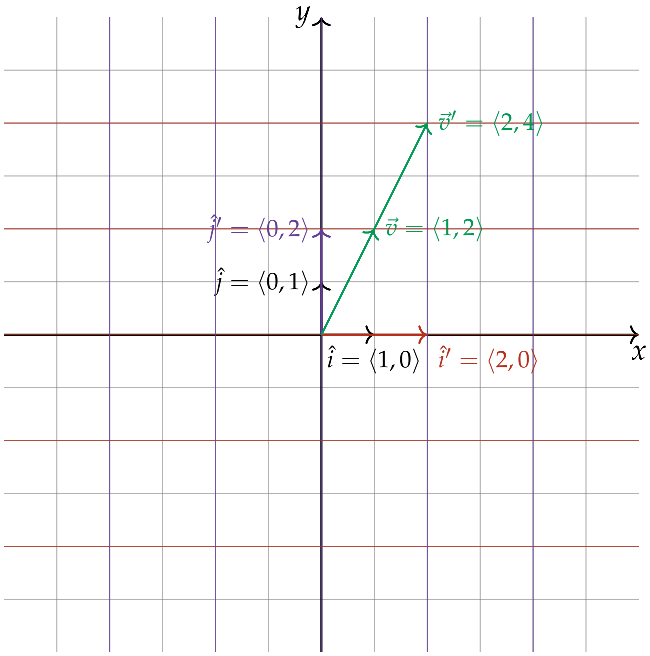
Thus, the transformation is linear.

As you can see above, if $T : \mathbb{R}^2 \to \mathbb{R}^2$, you can imagine the vectors dilating. Indeed, such linear transformations are called **dilation**

**Definition 4.2.8**

A **dilation** is a linear transformation that scales the vectors in the input vector space defined as $Tv = kv$ for some scalar $k$.

Let's take a look at the visual representation of the simple version of the example above. Consider the transformation defined as $T(\langle a, b \rangle) = \langle 2a, 2b \rangle$. We know from the previous example that the transformation is linear. Now let's try and graph it.




There are other special linear transformations that output the same vector space or the zero vector space.

**Definition 4.2.9**

A **zero transformation** is a linear transformation defined as $Tv = \mathbf{0}$. An **identity transformation** is a linear transformation that returns the same input vector as the output.

Let's take a look at a few more examples on linear transformations.

**Exercise 4.2.10**

Determine whether a transformation defined as $Tv = v^2$ is linear.

**Solution.**

Consider the following equation for arbitrary scalars $\alpha$ and $\beta$.

$$T(\alpha v_1 + \beta v_2) = (\alpha v_1 + \beta v_2)^2$$

Notice that the following equation holds.

$$= \alpha T(v_1) + \beta T(v_2) = \alpha v_1^2 + \beta v_2^2$$

Because $(\alpha v_1 + \beta v_2)^2 \neq \alpha v_1^2 + \beta v_2^2$ in general, the transformation is not linear.

Below is another example.

**Exercise 4.2.11**

Determine whether a transformation defined as $Ty = \frac{dy}{dx}$ for some differentiable function $y = f(x)$ is linear.

**Solution.**

Consider the following equation for some scalars $\alpha$ and $\beta$.

$$\begin{aligned} T(\alpha y_1 + \beta y_2) &= \frac{d}{dx}(\alpha y_1 + \beta y_2) = \frac{d}{dx}(\alpha y_1) + \frac{d}{dx}(\beta y_2) \\ &= \alpha \frac{d}{dx} y_1 + \beta \frac{d}{dx} y_2 = \alpha T(y_1) + \beta T(y_2) \end{aligned}$$

Thus, the transformation is linear.

One thing that we can notice from the problem above is that the transformation is not linear since changing the constants of an input will not change the output. The transformation is also not surjective as not all functions can be represented as the derivative of other functions. Continuing, we can consider the following transformation.

**Exercise 4.2.12**

Determine whether a transformation defined as $T(\langle a, b\rangle) = \langle -a, -b\rangle$ is linear.

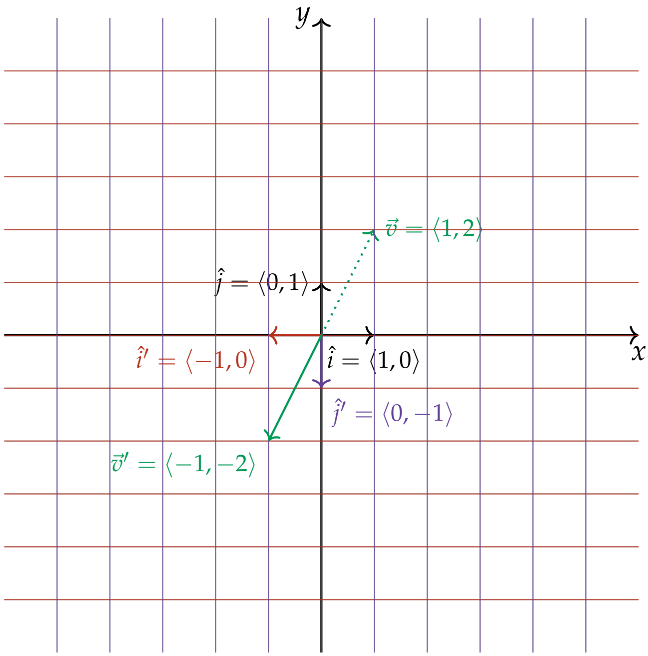
**Solution.**

Consider the following equation for arbitrary scalars $\alpha$ and $\beta$.

$$\begin{aligned} &T(\alpha\langle a, b\rangle + \beta\langle a', b'\rangle) \\ &= T(\langle \alpha a, \alpha b\rangle + \langle \beta a', \beta b'\rangle) = T(\langle \alpha a + \beta a', \alpha b + \beta b'\rangle) \\ &= \langle -\alpha a - \beta a', -\alpha b - \beta b'\rangle = \langle -\alpha a, -\alpha b\rangle + \langle -\beta a', -\beta b'\rangle \\ &= \alpha\langle -a, -b\rangle + \beta\langle -a', -b'\rangle = \alpha T(v_1) + \beta T(v_2) \end{aligned}$$

Thus, the transformation is linear.

Let's try and graph this to visually see how the vector space is transformed.




This transformation reflects the vectors across the origin! Now that we discussed what a linear transformation is, let's see how matrices relate to it.

### 4.2.3 Matrices and Linear Transformation

Recall that every vector in a vector space can be represented as a linear combination of the basis. We will apply our understanding of a linear transformation to vector representations in a vector space to find the relation.

Consider a basis $S = \{v_1, \ldots, v_n\}$ for a vector space $V$. Notice that by definition, an arbitrary vector $v \in V$ can be represented as a linear combination of a basis as shown below (where $c_k$ are constants).

$$v = \sum_{k=1}^{n} c_k v_k$$

Therefore, the following equation holds.

$$(v)_S = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Now returning to the definition of a linear transformation, we can consider the following equation.

$$Tv = T\left(\sum_{k=1}^{n} c_k v_k\right) = \sum_{k=1}^{n} c_k T v_k$$

Notice that the coordinate is preserved. In other words, the following equation holds for the new basis $S' = \{Tv_1, \ldots, Tv_n\}$

$$(Tv)_{S'} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

This could let us geometrically understand that a linear transformation "preserves" the general shape of a vector space since the new vectors have the same coordinates as the original ones, but with different basis. With this in mind, we can establish two important theorems.

**Theorem 4.2.13**

For a vector space $V$ and a basis $S = \{v_1, \ldots, v_n\}$, the transformation $T : V \to \mathbb{R}^n$ defined as $Tv = (v)_S$ is linear and bijective.

*Proof.*

First, the linearity of the transformation could be proven. Consider the following equations for $u_1, u_2 \in V$ and some constants $c_k$ and $d_k$.

$$\begin{aligned} T(\alpha u_1 + \beta u_2) &= T\left(\alpha \sum_{k=1}^{n} c_k v_k + \beta \sum_{k=1}^{n} d_k v_k\right) \\ &= T\left(\sum_{k=1}^{n} (\alpha c_k + \beta d_k) v_k\right) \\ &= \begin{bmatrix} \alpha c_1 + \beta d_1 \\ \alpha c_2 + \beta d_2 \\ \vdots \\ \alpha c_n + \beta d_n \end{bmatrix} = \begin{bmatrix} \alpha c_1 \\ \alpha c_2 \\ \vdots \\ \alpha c_n \end{bmatrix} + \begin{bmatrix} \beta d_1 \\ \beta d_2 \\ \vdots \\ \beta d_n \end{bmatrix} = \alpha \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} + \beta \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \\ &= \alpha T(u_1) + \beta T(u_2) \end{aligned}$$

Thus, $T$ is linear. Continuing, notice that if $T(u_1) = T(u_2)$, then $(u_1)_S = (u_2)_S$ and $u_1 = u_2$. Therefore, the transformation is injective. The transformation is also surjective as the entries $c_k$ in the coordinates are arbitrary real numbers. Thus, $T$ is linear and bijective.

Continuing, below is an important theorem and definition.

**Theorem 4.2.14**

For vector spaces $V$ and $W$ with basis $B = \{v_1, \ldots, v_n\}$ and $C = \{w_1, \ldots, w_m\}$ respectively, there exists a matrix $A$ such that the following equation holds for a linear transformation $T : V \to W$ and arbitrary $v \in V$.

$$(Tv)_C = A(v)_B$$

*Proof.*

First and foremost, consider the following equation for any $Tv_j \in W$ and constant $a_{ij}$.

$$Tv_j = \sum_{i=1}^{m} a_{ij} w_i$$

Let $A = [a_{ij}]$ and consider the following equation.

$$A(v_2)_B = A \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}$$

Generalizing, $(Tv_j)_C = A(v_j)_B$ and the following equations hold for constants $c_j$.

$$\begin{aligned}
(Tv)_C &= \left( \sum_{j=1}^{n} c_j T v_j \right)_C = \sum_{j=1}^{n} (c_j T v_j)_C = \sum_{j=1}^{n} c_j (Tv_j)_C \\
&= \sum_{j=1}^{n} c_j A(v_j)_B = A \left( \sum_{j=1}^{n} (c_j v_j)_B \right) = A \left( \sum_{j=1}^{n} c_j v_j \right)_B \\
&= A(v)_B
\end{aligned}$$

Thus, there always exists such a matrix $A$.

The matrix $A$ above has a special name. The matrix is known as the matrix representation of $T$ with respect to $B$ and $C$, denoted as $A_B^C$. In short, we also say $A$ from $B$ to $C$.

Continuing from the theorem, we can also notice that for a vector space $V$ and its basis $A$ and $B$, there exists a matrix $P_A^B$ such that $(v)_B = P_A^B(v)_A$ holds for all $v \in V$. To prove this, consider the identity transformation $T : V \to V$. By theorem 4.2.14, there exists a matrix $P$ such that $(v)_B = P_A^B(v)_A$. This matrix too has a special name.

**Definition 4.2.15**

A **transition matrix** is a matrix that converts the coordinate representation of a vector in a vector space $V$ from a basis in $V$ to another basis in $V$.

Now from the definitions, we could notice that linear transformations are highly related to the basis of a vector space. If we know how the basis is mapped, we can understand the linear transformation. Consider the following theorem.

**Theorem 4.2.16**

For a vector space and its bases $A$ and $B$, the coordinate representation of the $j$ vector of $A$ with respect to $B$ is the $j^{\text{th}}$ column of the transition matrix $P_A^B$.

*Proof.*

Let $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$. For some scalars $p_{ij}$, $a_i$ can be written as $a_j = \sum_{i=1}^n p_{ij} b_i$. Let a vector $v$ in the vector space be written as $v = \sum_{k=1}^n a'_k + a_k = \sum_{k=1}^n b'_k + b_k$. Continuing, the following equation holds.

$$v = \sum_{j=1}^n a'_j a_j = \sum_{j=1}^n a'_j \left( \sum_{i=1}^n p_{ij} b_i \right) = \sum_{i=1}^n \left( \sum_{j=1}^n p_{ij} a'_j \right) b_i$$

In other words, $b'_i = \sum_{j=1}^n p_{ij} a'_j$ and $(v)_B = [p_{ij}](v)_A = P_A^B(v)_A$.

Continuing with the theorem above, we can derive the following.

**Theorem 4.2.17**

A transition matrix $P_A^B$ is always invertible and $(P_A^B)^{-1} = P_B^A$.

*Proof.*

Consider the following equations for a vector $v$.

$$\begin{aligned}(v)_B &= P_A^B (v)_A \\ (v)_A &= P_B^A (v)_B\end{aligned}$$

Therefore, $(v)_B = P_A^B(v)_A = P_A^B P_B^A (v)_B$ for all $(v)_B$. Thus, $P_A^B P_B^A = I$ and the theorem holds.

Now how about matrix representations like $A_S^S$? We have an interesting theorem that links such matrices with transition matrices.

**Theorem 4.2.18**

Let $A_B^B$ and $A_C^C$ be the matrix representations of linear transformations for bases $B$ and $C$. If $P_B^C$ represents the transition matrix, then $A_B^B = P_C^B A_C^C P_B^C$.

*Proof.*

For a vector $v$ in the same vector space, notice that $(Tv)_B = A_B^B(v)_B$ and $(v)_C = P_B^C(v)_B$ hold by definition. Consider the following equation.

$$\begin{aligned}(Tv)_B &= P_C^B (Tv)_C = P_C^B A_C^C (v)_C = P_C^B A_C^C P_B^C (v)_B \\ &= A_B^B (v)_B\end{aligned}$$

Thus, $P_C^B A_C^C P_B^C = A_B^B$.

Let's use $A$ for $A_B^B$, $\tilde{A}$ for $A_C^C$, and $P$ for $P_B^C$ where $P^{-1} = P_C^B$. Notice that we can write $A = P^{-1}\tilde{A}P$ and $PA = \tilde{A}P$. Here, we say that $A$ and $\tilde{A}$ are **similar**.

**Definition 4.2.19**

Two matrices $A$ and $B$ are **similar** if the linear transformation that they represent with different bases is identical. In other words, for transition matrix $P = P_A^B$, $PA = BP$.

Continuing with the definition, we can establish the following theorem.

**Theorem 4.2.20**

For a vector space $V$ with basis $B$, let $A$ be the matrix representation of the linear transformation $T : V \to V$ from $B$ to $B$. If $A$ is similar to another matrix representation $\tilde{A}$, then there exists a basis $C$ such that $\tilde{A}$ represents the same linear transformation from $C$ to $C$.

*Proof.*

First and foremost, let a transition matrix from a basis $C$ to $B$ be $P$ and let $v$ be a vector in the vector space. Consider the following equations.

$$\begin{aligned}(Tv)_B = A(v)_B &= AP(v)_C \\ &= P(Tv)_C\end{aligned}$$

Therefore, $AP(v)_C = P(Tv)_C$ and $(Tv)_C = A'(v)_C = P^{-1}AP(v)_C$. In other words, for each basis $C$, there exists a transition matrix similar to $A$. Moreover, because the whole process is invertible, the theorem holds.

Now that we discussed that matrices can be used to represent linear transformations, let's discuss about topics that are never left out in linear algebra.

### 4.2.4 Kernels and Images

As you could have noticed, sometimes it is easier to discuss a linear transformation with intuitions and sometimes matrices are much more straightforward. Recall that linear transformations are functions for vector spaces. Sometimes it is easier to look at the functions directly and sometimes it is more convenient to look at the matrix representation of such functions. That is what we are going to do with kernels and images. As always, let's get started with definitions.

**Definition 4.2.21**

The **kernel** of a linear transformation $T : V \to W$, also known as the **nullspace** of $T$, is the set of vectors in $V$ that are mapped to zero vectors when transformed with $T$.

$$\ker(T) = \{v \in V \mid Tv = \mathbf{0}\}$$

Continuing with the definition, we can establish two fundamental properties of the kernel of a transformation.

**Theorem 4.2.22**

Consider the following statements for all linear transformations $T : V \to W$.

1. $\mathbf{0} \in \ker(T)$
2. $\ker(T)$ is a subspace of $V$.

Let's start by proving the first statement. The proof is rather straightforward with the properties of linear transformation.

*Proof.*

Consider the following equation.

$$T(\mathbf{0}) = T(0\mathbf{0}) = 0T(\mathbf{0}) = \mathbf{0}$$

Because $\mathbf{0}$ is mapped to $\mathbf{0}$ after any linear transformation $T$, $\mathbf{0} \in \ker(T)$ for all linear transformations $T$.

Continuing, below is the proof for the second statement.

*Proof.*

By definition, $\ker(T) \in V$. Therefore, it suffices to show that the kernel of $T$ is closed under vector addition and scalar multiplication. Because $\ker(T)$ is not empty by the first statement, consider the following equations for $u, v \in \ker(T)$ and some scalars $\alpha$ and $\beta$.

$$T(\alpha u + \beta v) = \alpha Tu + \beta Tv = \alpha \mathbf{0} + \beta \mathbf{0} = \mathbf{0}$$

In other words $\alpha u + \beta v \in \ker(T)$ and $\ker(T)$ is a subspace of $V$.

Recall that a linear transformation can be represented as some matrix $A$. Therefore, kernel of $A$ can be interpreted as the set of solutions to $Ax = \mathbf{0}$. Notice that some solutions retain finite and infinite number of vectors. Therefore, we can establish the following definition.

**Definition 4.2.23**

The **nullity** of $T$, denoted as $\text{null}(T)$, is the dimension of $\ker(T)$ for some linear transformation $T$.

Continuing, we can define the image of a transformation, which should be unsurprising.

**Definition 4.2.24**

The subset of the codomain $W$ for a transformation $T : V \to W$ that contains all output vectors is known as the **image** of $T$. Below defines $\Im(T)$ with $v \in V$.

$$\Im(T) = \{w \in W \mid w = Tv\}$$

Although the kernel and the image of a transformation are very different, we can establish very similar statements for the image just as how we derived them for the kernel.

**Theorem 4.2.25**

Consider the following statements for all linear transformations $T : V \to W$.

1. $\mathbf{0} \in \Im(T)$
2. $\Im(T)$ is a subspace of $W$.

Let's start by proving the first statement.

*Proof.*

By theorem 4.2.22, there exists a vector in $V$ such that it is mapped to $\mathbf{0}$ in $W$. Thus, the statement holds.

Continuing, below is the proof for the second statement.

*Proof.*

First and foremost, $\Im(T) \in W$ by definition. Therefore, it suffices to show that the image is closed under vector addition and scalar multiplication. Consider the following equation for $v_1, v_2 \in V$, $Tv_1 = w_1$, $Tv_2 = w_2$, and some scalars $\alpha$ and $\beta$.

$$\alpha w_1 + \beta w_2 = \alpha T v_1 + \beta T v_2 = T(\alpha v_1 + \beta v_2)$$

Because $T(\alpha v_1 + \beta v_2) \in \Im(T)$, $\alpha w_1 + \beta w_2 \in \Im(T)$ and the statement holds.

Now recall the kernel of a matrix $A$ is the set of all solutions to the equation $Ax = \mathbf{0}$. Similarly, the image is the set of all possible outcomes of the product $Ax$ for some $x \in V$. Consider the equations below.

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n a_{1k} x_k \\ \sum_{k=1}^n a_{2k} x_k \\ \vdots \\ \sum_{k=1}^n a_{mk} x_k \end{bmatrix} = \sum_{k=1}^n x_k \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix}$$

In other words, the image of $A$ is the column space of $A$. Therefore, we can establish the following definition.

**Definition 4.2.26**

The **rank** of a transformation $T$, denoted as $\mathrm{r}(T)$, is $\dim(\Im(T))$.

Continuing from the definition and ideas of column space and the image, we can derive the following theorem.

**Theorem 4.2.27**

Let $B$ and $C$ be $n$-dimensional and $m$-dimensional bases for $V$ and $W$ in a linear transformation $T : V \to W$ respectively. If the matrix representation of $T$ is $A = A_B^C$, then $\mathrm{r}(T) = \mathrm{r}(A)$.

*Proof.*

First and foremost, let a basis for $\Im(T)$ be $\{w_1, \dots, w_k\}$. Notice that if $\{(w_1)_C, \dots, (w_k)_C\}$ is a basis for $\mathrm{col}(A)$, then $\dim(\Im(T)) = \dim(\mathrm{col}(A))$ and $\mathrm{r}(T) = \mathrm{r}(A)$. Thus, it suffices to show that $\{(w_1)_C, \dots, (w_k)_C\}$ is a basis for $\mathrm{col}(A)$.

Let $c_k$ be some scalars that $x \in \mathrm{col}(A)$ is represented as the linear combination of the columns $A_k$ of $A$ where $x = \sum_{k=1}^{n} c_k A_k$. Note that $c_k$ can be written as an $n$-tuple. Thus, let $v \in V$ be a vector such that the following equation holds.

$$(v)_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

Continuing, $x = A(v)_B = (Tv)_C$. Because $Tv \in \Im(T)$, it is evident that $x \in \mathrm{span}\{(w_1)_C, \dots, (w_k)_C\}$. Moreover, because $x$ is an arbitrary vector from $\mathrm{col}(A)$, it is evident that $\mathrm{col}(A) \subset \mathrm{span}\{(w_1)_C, \dots, (w_k)_C\}$ and $\{(w_1)_C, \dots, (w_k)_C\}$ spans $\mathrm{col}(A)$.

By definition, $\{w_1, \ldots, w_k\}$ is linearly independent. For some scalars $d_k$, for $\sum_{k=1}^{n} d_k (w_k)_C$ to be a zero vector, $(\sum_{k=1}^{n} d_k w_k)_C = \mathbf{0}$ and $\sum_{k=1}^{n} d_k w_k = \mathbf{0}$, which only happens when $d_k = 0$ for all integer $k \in [0, n]$. Thus, $\{(w_1)_C, \ldots, (w_k)_C\}$ is linearly independent and is a basis for $\text{col}(A)$.

As you can see above by the theorem, although the kernel and the image of a linear transformation appear to be totally unrelated, they are somehow related. Below is another theorem that connects the bases of the kernel and the image of a linear transformation.

**Theorem 4.2.28**

Let $\{v_1, \ldots, v_k\}$ be the basis of the kernel of a linear transformation $T : V \to W$. If $\{v_1, \ldots, v_k, \ldots v_n\}$ is a basis for $V$, then a basis of $\Im(T)$ is $\{Tv_{k+1}, \ldots, Tv_n\}$.

*Proof.*

To prove that $\{Tv_{k+1}, \ldots, Tv_n\}$ is a basis of $\Im(T)$, it suffices to show that the set of vectors is linearly independent and spans $\Im(T)$. First, to show that the set is linearly independent, consider the following equations for some scalars $c_i$.

$$\sum_{i=k+1}^{n} c_i T v_i = T\left(\sum_{i=k+1}^{n} c_i v_i\right) = \mathbf{0}$$

In other words, $\sum_{i=k+1}^{n} c_i v_i \in \ker(T)$ and it can be written as the linear combination of the basis, or $\sum_{i=k+1}^{n} c_i v_i = \sum_{i=1}^{k} c_i v_i$. By definition, because $\{v_1, \ldots, v_k, \ldots v_n\}$ is linearly independent, $c_i$ must be all zero for integer $i \in [1, n]$ for $\sum_{i=k+1}^{n} c_i v_i - \sum_{i=1}^{k} c_i v_i = \mathbf{0}$ to satisfy. Thus, $\{Tv_{k+1}, \ldots, Tv_n\}$ is linearly independent.

Continuing, the fact that the set $\{Tv_{k+1}, \ldots, Tv_n\}$ spans $\Im(T)$ can be proven. Let $w = Tv$ for any $w \in \Im(T)$. Consider the following equations below.

$$\begin{aligned} w = Tv = T\left(\sum_{i=1}^{n} c_i v_i\right) &= \sum_{i=1}^{n} c_i T v_i \\ = \sum_{i=1}^{k} c_i T v_i + \sum_{i=k+1}^{n} c_i T v_i &= \sum_{i=k+1}^{n} c_i T v_i \in \operatorname{span}\{Tv_{k+1}, \ldots, Tv_n\} \end{aligned}$$

In other words, $\{Tv_{k+1}, \ldots, Tv_n\}$ spans $\Im(T)$. Thus, the theorem holds.

Continuing from this theorem, we can establish another interesting corollary as a theorem.

**Theorem 4.2.29**

For a finite dimensional vector space $V$, the following equation is true for a linear transformation $T : V \to W$.

$$\mathrm{r}(T) + \mathrm{null}(T) = \dim(V)$$

*Proof.*

Let $k = \dim(\ker(T))$ and $n = \dim(V)$. By Theorem 4.2.28, there exists a basis for $\Im(T)$ with $n-k$ vectors, or $\mathrm{r}(T) = \dim(\Im(T)) = n-k$. Thus, $\mathrm{r}(T) + \mathrm{null}(T) = n - k + k = n = \dim(V)$.

Moving on from the relationship with the kernel and the image of a linear transformation, let's discuss about injection, surjection, and bijection. Consider the following theorem.

**Theorem 4.2.30**

A linear transformation $T : V \to W$ is injective if and only if $\ker(T) = \{\mathbf{0}\}$.

*Proof.*

By definition, $T$ is injective if $v_1, v_2 \in V$ such that $v_1 \neq v_2$ implies $Tv_1 \neq Tv_2$. Therefore, if $v \in \ker(T)$, then $Tv = \mathbf{0} = T\mathbf{0}$ and $\ker(T)$

must contain no nonzero vectors for $T$ to be injective. Thus, $\ker(T)$ must only contain the zero vector.

Continuing, below is another theorem on injection and surjection.

**Theorem 4.2.31**

For a linear transformation $T : V \to W$ with $\dim(V) = \dim(W) = n$, $T$ is injective if and only if $T$ is surjective.

*Proof.*

By Theorem 4.2.30, $T$ is injective if and only if $\ker(T) = \{\mathbf{0}\}$. Moreover by definition, $\ker(T) = \{\mathbf{0}\}$ if and only if $\mathrm{null}(T) = 0$ and $\mathrm{r}(T) = \dim(V) - \mathrm{null}(T) = n$ by Theorem 4.2.29. Continuing by definition, $\mathrm{r}(T) = n$ if and only if $\dim(\Im(T)) = n = \dim(W)$. Therefore, $T$ is injective if and only if $T$ is surjective.

With the interesting properties, we can assign a special name for bijective linear transformation.

**Definition 4.2.32**

A bijective linear transformation is called an **isomorphism**. Moreover, a vector space $V$ is said to be **isomorphic** to another vector space $W$ if there exists an isomorphism $T : V \to W$, denoted as $V \cong W$.

Continuing from the definition, we can derive the following properties.

**Theorem 4.2.33**

For vector spaces $U$, $V$, and $W$, isomorphism is reflexive, symmetric, and transitive.

1. $U \cong U$
2. If $U \cong V$, then $V \cong U$.
3. If $U \cong V$ and $V \cong W$, then $U \cong W$.

Let's start by proving the first statement.

*Proof.*

Notice that it suffices to show that the identity transformation is bijective. By definition, because each vector is mapped to itself, identity transformation is bijective and $U \cong U$.

Below is the proof for the second property.

*Proof.*

First and foremost, notice that if there exists $T : U \to V$ that is bijective, then there exists its inverse $T^{-1} : V \to U$ that is also bijective. Continuing, consider the equation below for some vectors $v_1, v_2 \in V$ and scalars $\alpha$ and $\beta$.

$$\begin{aligned}
\alpha T^{-1}v_1 + \beta T^{-1}v_2 &= T^{-1}\left(T\left(\alpha T^{-1}v_1 + \beta T^{-1}v_2\right)\right) \\
&= T^{-1}\left(\alpha T\left(T^{-1}v_1\right) + \beta T\left(T^{-1}v_2\right)\right) \\
&= T^{-1}\left(\alpha v_1 + \beta v_2\right)
\end{aligned}$$

Therefore, $T^{-1}$ is linear and $T^{-1}$ is an isomorphism.

Continuing, below is the proof for the last statement.

*Proof.*

First and foremost, let $T_1 : U \to V$ and $T_2 : V \to W$ be isomorphisms. Therefore, it suffices to show that $T_2 \circ T_1$ is an isomorphism. Notice that because $T_1$ and $T_2$ are bijective, $T_2 \circ T_1$ must also be bijective. Moreover, consider the following equation for some vectors $v_1, v_2 \in V$ and scalars $\alpha$ and $\beta$.

$$\begin{aligned}
(T_2 \circ T_1)(\alpha v_1 + \beta v_2) &= T_2(\alpha T_1 v_1 + \beta T_1 v_2) \\
&= \alpha T_2(T_1 v_1) + \beta T_2(T_1 v_2) \\
&= \alpha (T_2 \circ T_1)(v_1) + \beta (T_2 \circ T_1)(v_2)
\end{aligned}$$

Thus, $T_2 \circ T_1$ is linear and the statement is true.

As you can see above, the three statements in Theorem 4.2.33 each shows a relation called reflexive, symmetric, and transitive. Continuing from this terms, we can assign a special name for a relation.

**Definition 4.2.34**

An **equivalence relation** is a relation that is reflexive, symmetric, and transitive.

Below is another interesting theorem on isomorphism.

**Theorem 4.2.35**

For finite dimensional vector spaces $V$ and $W$, $V \cong W$ if and only if $\dim(V) = \dim(W)$.

*Proof.*

First, the statement if $\dim(V) = \dim(W)$, then $V \cong W$ can be proven. Notice that the coordinate representations of infinitely many vectors in $V$ and $W$ are unique. Therefore, $V \cong \mathbb{R}^n$ and $W \cong \mathbb{R}^n$. By Theorem 4.2.33, $\mathbb{R}^n \cong W$ and $V \cong W$.

Continuing, if $V \cong W$, then there exists a linear transformation $T : V \to W$ that is bijective. By Theorem 4.2.30, $\ker(T) = \{\mathbf{0}\}$ and $\text{null}(T) = 0$. Moreover by definition, because $T$ is surjective, $\text{r}(T) = \dim(\Im(T)) = \dim(W)$. Finally by Theorem 4.2.29, $\dim(W) + 0 = \dim(V)$ and the theorem holds.

From the theorem above, we can notice that if we consider a bijective linear transformation that maps to coordinate representation, then all $n$-dimensional vector spaces are isomorphic to $\mathbb{R}^n$. In other words, every $n$-dimensional vector space over a given field for some finite $n$ can be considered as the same vector space, but just with different representations.

This is it for linear transformations, and let's discuss about eigenvalues and eigenvectors!

## §4.3 Eigenvalues and Eigenvectors

As always, let's start with definitions.

**Definition 4.3.1**

For a linear transformation $T : V \to V$ with a nonzero vector $v \in V$ and some scalar $\lambda$,if $Tv = \lambda v$, then $v$ is an **eigenvector** of $T$ and $\lambda$ is an **eigenvalue** of $T$.

From here, it is important to note that while $v$ must be a nonzero vector, $\lambda$ could equal zero. Let's take a look at an example.

**Exercise 4.3.2**

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation defined as $T(a, b) = (a, -b)$. Find an eigenvector and eigenvalue of $T$.

**Solution.**

Consider the following equation.

$$T(0, 1) = (0, -1) = -1 \cdot (0, 1)$$

Therefore, $(0, 1)$ is an eigenvector of $T$ and $-1$ is the corresponding eigenvalue of $T$.

Below is another example for a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$.

**Exercise 4.3.3**

Find, if any, eigenvectors and eigenvalues of $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $T(a, b) = (b, -a)$.

**Solution.**

For some scalar $\lambda$, consider the following equation.

$$T(a, b) = (b, -a) = \lambda(a, b) = (\lambda a, \lambda b)$$

In other words $b = \lambda a$ and $-a = \lambda b$. Substituting, $-a = \lambda^2 a$ and $(\lambda^2 + 1)a = 0$. Because $\lambda \in \mathbb{R}$, $a = 0$ and $b = 0$. However, because eigenvectors are nonzero vectors by definition, $T$ has not real eigenvalues and eigenvectors.

Now eigenvectors and eigenvalues are important because they make our lives easier for investigating matrix representation of a linear transformation. Consider the following theorem.

**Theorem 4.3.4**

For a linear transformation $T : V \to V$ for a finite dimensional vector space $V$, there exists a basis whose matrix representation is a diagonal matrix if and only if there exists a basis composed of only eigenvectors of $T$.

*Proof.*

First, let $B = \{v_1, \ldots, v_n\}$ be a basis for $V$ and let the matrix representation of $T$ be a diagonal matrix $A$ whose diagonal entries are $\lambda_1, \ldots, \lambda_n$. Thus

$$(Tv_j)_B = \begin{bmatrix} 0 \\ \vdots \\ \lambda_j \\ \vdots \\ 0 \end{bmatrix}$$

hold by construction and $Tv_j = \lambda_j v_j$. In other words, $v_j$ and $\lambda_j$ are an eigenvector and an eigenvalue of $T$ and there exists a basis composed of only eigenvectors of $T$.

Continuing, let $B = \{v_1, \ldots, v_n\}$ be a basis for $V$ whose elements are all eigenvectors of $T$. Therefore, $Tv_j = \lambda v_j$ and the following equation holds.

$$(Tv_j)_B = \begin{bmatrix} 0 \\ \vdots \\ \lambda_j \\ \vdots \\ 0 \end{bmatrix}$$

Therefore, there exists a basis whose matrix representation is a diagonal matrix.

With this theorem and connecting vector spaces with matrix again, we can redefine eigenvectors and eigenvalues with different interpretations.

**Definition 4.3.5**

For a square matrix $A_{n\times n}$ and a nonzero $n$-tuple $x \in \mathbb{R}^n$, if there exists a scalar $\lambda$ such that $Ax = \lambda x$, then $x$ is an **eigenvector** and $\lambda$ is an **eigenvalue** of $A$.

When you think about this definition, we can see that both definitions imply the same thing as the $A$ is the matrix representation of a linear transformation and $x$ is a vector in the corresponding vector space. Let's take a look at a quick example.

**Exercise 4.3.6**

Find an eigenvector and eigenvalue of matrix $A$ defined as the following.

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

**Solution.**

Consider the following equation.

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} = -1 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Thus, $-1$ is an eigenvalue and $[01]$ is an eigenvector.

If this example looks familiar, this is actually the same question as the first example from this section. Continuing, below are two interesting theorems on eigenvectors and eigenvalues.

**Theorem 4.3.7**

For a linear transformation $T : V \to V$, $v \in V$ is an eigenvector with the corresponding eigenvalue $\lambda$ if and only if $(v)_B$ is an eigenvector with the corresponding eigenvalue $\lambda$ for $A_B^B$, the matrix representation of $T$ for any basis $B$.

*Proof.*

Notice that $Tv = \lambda v$ if and only if $A_B^B(v)_B = (Tv)_B = (\lambda v)_B = \lambda(v)_B$. Thus, the theorem holds.

This theorem will come in handy as we can either find all eigenvectors and corresponding eigenvalues of the transformation or its matrix representation. Continuing, below is another theorem.

**Theorem 4.3.8**

For a linear transformation $T : V \to V$ and eigenvalue $\lambda$, $S_\lambda = \{v \in V | Tv = \lambda v\}$ is a subspace of $V$.

*Proof.*

First, notice that because $S_\lambda \subset V$, it suffices to show that the set is closed under vector addition and scalar multiplication. Consider the following equations for $u, v \in S_\lambda$ and some scalars $\alpha$ and $\beta$.

$$\begin{aligned} T(\alpha u + \beta v) &= \alpha Tu + \beta Tv = \alpha\lambda u + \beta\lambda v \\ &= \lambda(\alpha u + \beta v) \end{aligned}$$

In other words, $\alpha u + \beta v \in S_\lambda$ and the theorem holds.

As you may have guessed, we can establish a new definition from the theorem above.

**Definition 4.3.9**

For a transformation $T : V \to V$ and its eigenvalue $\lambda$, $S_\lambda = \{v \in V | Tv = \lambda v\}$ defined as the **eigenspace** of $T$ for the eigenvalue $\lambda$.

Just as we redefined eigenvectors and eigenvalues with matrices, we could do the same for eigenspace. Now that we discussed what eigenvectors and eigenvalues are, let's discuss how to find one.

**Theorem 4.3.10**

For a square matrix $A_{n\times n}$, $\lambda$ is its eigenvalue if and only if $\det(A - \lambda I_n) = 0$.

*Proof.*

First notice that by definition, $\lambda$ is an eigenvalue of $A$ if and only if $Ax = \lambda x$ holds for some nonzero vector $x$. Continuing, $Ax = \lambda x$ holds if and only if $Ax = \lambda I_n x$ and $(A - \lambda I_n)x = 0$. In other words, the columns of the matrix $A - \lambda I_n$ are linearly dependent and invertible. Therefore by Theorem 2.5.13, the theorem holds.

Being one of the most important theorems in linear algebra, we can establish another definition from the theorem.

**Definition 4.3.11**

The **characteristic equation** of a square matrix $A$ is defined as $\det(A - \lambda I_n) = 0$ for scalar variable $\lambda$.

An important thing to note from this definition is that $\det(A - \lambda I_n)$ is a polynomial in $\lambda$ with degree $n$. Moreover, the real roots of the polynomial can be interpreted as eigenvalues of $A$. Let's take a look at an example of how the characteristic equation can be used to find all eigenvalues of a matrix.

**Exercise 4.3.12**

Find all eigenvalues and the corresponding eigenspaces of matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

**Solution.**

To find all eigenvalues of $A$, it suffices to find all solutions for the characteristic equation of $A$ by Theorem 4.3.10.

$$A - \lambda I_2 = \begin{bmatrix} 1-\lambda & 0 \\ 0 & -1-\lambda \end{bmatrix}$$

Therefore, $\det(A - \lambda I_2) = (1-\lambda)(-1-\lambda)$ and $\lambda = 1$ or $\lambda = -1$ are the solutions to the characteristic equation.

Continuing, the corresponding eigenspaces could be found. For $\lambda = 1$, the eigenspace is defined as $S_1 = \{x | Ax = x\}$. Solving for the equation $(A - I)x = 0$, the following equations hold.

$$\begin{bmatrix} 0 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$
$$-2x_2 = 0$$

Therefore, the eigenspace for the eigenvalue 1 can be represented as the following

$$S_1 = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

Similarly, $S_{-1} = \{x | Ax = -x\}$ and $(A + I)x = 0$ could be solved.

$$\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$
$$2x_1 = 0$$

Therefore,

$$S_{-1} = \text{span}\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

and the corresponding eigenspaces for all eigenvalues of $A$ are found.

We got a nice matrix here, but it's worth noting that $\lambda$ need not be real. As demonstrated from the simple example above, solving the characteristic equation will lead to all eigenvalues for a matrix. I will leave more complex and interesting examples in the practice problems below. In the next part of the section, we will discuss the properties of eigenvalues and eigenvectors.

### 4.3.1 Properties of Eigenvalues

Eigenvalues and eigenvectors have interesting properties and relationship with vector spaces. Consider the following theorem.

**Theorem 4.3.13**

For two square matrices $A$ and $B$ such that $A \cong B$, the characteristic equation and eigenvalues for each matrix are equal.

*Proof.*

By definition of similar matrices, there exists some invertible matrix $P$ such that $B = P^{-1}AP$. Therefore, the following equations hold.

$$P^{-1}(A - \lambda I)P = P^{-1}AP - P^{-1}\lambda IP = B - \lambda I$$

Therefore, $\det(B - \lambda I) = \det(P^{-1}(A - \lambda I)P) = \det(A - \lambda I)$ and the characteristic equations are the same. In other words, the eigenvalues for $A$ and $B$ are equal and the theorem holds.

Below is another theorem on the eigenvalues of triangular matrices.

**Theorem 4.3.14**

The eigenvalues of triangular (upper or lower) matrices are the diagonal elements.

*Proof.*

By definition, notice that for a triangular matrix $A$, $A - \lambda I$ is also triangular. Therefore by Theorem 2.5.4,

$$\det(A - \lambda I) = \prod_{k=1}^{n} (\lambda_k - \lambda)$$

and $\det(A - \lambda I_n) = 0$ if and only if $\lambda = \lambda_k$. Because $\lambda_k$ are the diagonal elements of $A$, the theorem holds.

Continuing, below is another theorem on the relationship between eigenvalues and invertibility.

**Theorem 4.3.15**

A square matrix $A$ contains a zero eigenvalue if and only if it is not invertible.

*Proof.*

A square matrix $A$ contains a zero eigenvalue if and only if $\det(A - 0I) = 0$ and $\det(A) = 0$. Because $A$ is invertible if and only if $\det(A) \neq 0$ by Theorem 2.5.13, the theorem holds.

If you think this theorem is nice, below is another theorem on invertibility and eigenvalues.

**Theorem 4.3.16**

For an invertible matrix $A$ with eigenvector $x$ and its corresponding eigenvalue $\lambda$, $x$ is also an eigenvalue of $A^{-1}$ with corresponding eigenvalue $\frac{1}{\lambda}$.

*Proof.*

By definition, $Ax = \lambda x$ and $x = A^{-1}\lambda x$. Continuing, $A^{-1}x = \frac{1}{\lambda} \cdot x$. Therefore, $x$ is also an eigenvector of $A^{-1}$ with the corresponding eigenvalue $\frac{1}{\lambda}$.

Continuing, below is a theorem on the relationship between eigenvalues and linear independence.

**Theorem 4.3.17**

If $x_1, \ldots, x_k$ are eigenvectors of a square matrix $A$ and the corresponding eigenvalues $\lambda_1, \ldots, \lambda_k$ are all distinct, then $\{x_1, \ldots, x_k\}$ is linearly independent.

*Proof.*

First, notice that $\{x_1\}$ is linearly independent. Therefore, it suffices to show that if the theorem holds for $k - 1 > 0$, then the statement holds for $k$. Assume $\{x_1, \ldots, x_{k-1}\}$ is a linearly independent set of eigenvectors with distinct corresponding eigenvalues $\lambda_1, \ldots, \lambda_{k-1}$. Consider the following equations for some scalars $c_i$.

$$\sum_{i=1}^{k} c_i x_i = 0$$
$$\sum_{i=1}^{k} c_i A x_i = 0$$
$$\sum_{i=1}^{k} c_i \lambda_i x_i = 0$$
$$\sum_{i=1}^{k-1} c_i \lambda_k x_i = 0$$

Subtracting the last two equations, $\sum_{i=1}^{k-1} c_i(\lambda_i - \lambda_k)x_i = 0$ is obtained. Because $\{x_1, \ldots, x_{k-1}\}$ is linearly independent, $c_i(\lambda_i - \lambda_k) = 0$ for all integer $i \in [1, k)$. Continuing, because $\lambda_i - \lambda_k \neq 0$ by construction, $c_i = 0$ for all integer $i \in [1, k)$ and $c_k = 0$ since $x_k \neq 0$. Thus $\{x_1, \ldots, x_k\}$ is linearly independent and the theorem holds by induction.

Naturally, we can consider the dimension of an eigenspace.

**Theorem 4.3.18**

For a matrix $A_{n\times n}$ with an eigenvalue $\lambda$, the following equation holds.

$$\dim(S_\lambda) = n - \mathrm{r}(A - \lambda I)$$

*Proof.*

By definition, the eigenspace for $\lambda$ is the set of all vectors $v$ that satisfy $Av = \lambda v$. Because $(A - \lambda)v = 0$, $S_\lambda$ can be interpreted as $\ker(A - \lambda I)$. Moreover by Theorem 4.2.29, the following equation holds.

$$\mathrm{r}(A - \lambda I) + \mathrm{null}(A - \lambda I) = n$$

Because $\dim(\ker(A - \lambda I)) = \mathrm{null}(A - \lambda I)$ holds by definition, $\dim(S_\lambda) = n - \mathrm{r}(A - \lambda I)$.

We could really go on forever with interesting theorems, but before we conclude this part of the note, I would like to discuss two more theorems.

**Theorem 4.3.19**

Consider the following statements for a matrix $A$ with an eigenvector $x$ with its corresponding eigenvalue $\lambda$.

1. For an arbitrary scalar $k$, $x$ is also an eigenvector of $kA$ with corresponding eigenvalue $k\lambda$.
2. For some $n \in \mathbb{N}$, $x$ is also an eigenvector of $A^n$ with corresponding eigenvalue $\lambda^n$.

Let's start by proving the first statement.

*Proof.*

Consider the following equation.

$$(kA)x = k(Ax) = k\lambda x = (k\lambda)x$$

Therefore, $x$ is an eigenvector of $kA$ and its corresponding eigenvalue is $k\lambda$.

Below is the proof for the second statement.

*Proof.*

Assume that if the statement holds for $n-1$, then the statement holds for $n > 1$. Consider the following equations.

$$\begin{aligned} A^n x &= A\left(A^{n-1}x\right) = A\left(\lambda^{n-1}x\right) = \lambda^{n-1}Ax \\ &= \lambda^{n-1}\lambda x = \lambda^n x \end{aligned}$$

Because $Ax = \lambda x$, the statement holds by induction.

Before we move on to the final content with the last theorem for this part, below is a definition for the theorem.

**Definition 4.3.20**

The **trace** of a matrix $A = [a_{ij}]$, denoted as $\mathrm{tr}(A)$, is the sum of the diagonal elements of $A$.

With the definition, we can continue with the theorem.

**Theorem 4.3.21**

For a square matrix $A_{n\times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$, the following equations hold.

1. $\mathrm{tr}(A) = \sum_{k=1}^{n} \lambda_k$
2. $\det(A) = \prod_{k=1}^{n} \lambda_k$

Let's start by proving the first equation.

*Proof.*

By the Fundamental Theorem of Algebra, the following equation holds.

$$\det(A - \lambda I) = \prod_{k=1}^{n} (\lambda_k - \lambda) = 0$$

Therefore, all eigenvalues are the diagonal elements and by the definition of trace of $A$, the sum of all eigenvalues are the trace of $A$.

Continuing, below is the proof for the second equation.

*Proof.*

Substituting $\lambda = 0$ into the equation

$$\det(A - \lambda I) = \prod_{k=1}^{n} (\lambda_k - \lambda),$$

$\det(A) = \prod_{k=1}^{n} \lambda_k$ and the equation holds.

It is important to note that the second equation counts multiplicity. In other words, even if $\lambda_p = \lambda_q$, you still count both. Wow! I can't believe that we are already almost done with the notes! For the final concept of my notes on linear algebra, we will discuss diagonalization.

### 4.3.2 Diagonalization

Diagonalization in linear algebra is a technique that I think really makes our lives easier in calculations. To get into it, let's get started with definition and theorems.

**Definition 4.3.22**

A square matrix that is similar to a diagonal matrix is known to be **diagonalizable**.

Now we cannot always use a brute force method to see if two matrices are similar. Fortunately, the following theorem will help in checking whether a square matrix is diagonalizable.

**Theorem 4.3.23**

A square matrix $A_{n\times n}$ is diagonalizable if and only if it retains a set of $n$ eigenvectors that is linearly independent.

*Proof.*

First and foremost, notice that $A$ and a diagonal matrix $D$ of the same order with entries $\lambda_i$ are similar if and only if there exists an invertible matrix $P$ such that $P^{-1}AP = D$ or $AP = PD$. Let $x_j$ and $e_j$ be the $j^{\text{th}}$ column of $P$ and $I$ respectively. Then, the following equations hold.

$$\begin{aligned} Ax_j = APe_j &= PDe_j = P\lambda_j e_j \\ &= \lambda_j Pe_j = \lambda_j x_j \end{aligned}$$

Thus, $x_j$ and $\lambda_j$ are an eigenvector and the corresponding eigenvalue of $A$. By definition, because $M$ is invertible, $x_j$ are all linearly independent and $A$ has $n$ linearly independent eigenvectors. In other words, if $A$ is diagonalizable, then $A$ has $n$ linearly independent eigenvectors.

Continuing, the converse can be proven. Let $A$ have $n$ linearly independent eigenvectors denoted as $x_j$ with corresponding eigenvalues $\lambda_j$. Construct matrices $P$ and $D$ such that the $j^{\text{th}}$ column of $P$ is $x_j$ and the diagonal element of $D$ is $\lambda_j$. If $e_j$ is the $j^{\text{th}}$ column of $I$, then the $j^{\text{th}}$ column of $AP$ can be written as the following.

$$APe_j = Ax_j = \lambda_j x_j$$

Similarly, the $j^{\text{th}}$ column of $PD$ can be written as $PDe_j = M\lambda_j e_j = \lambda_j Me_j = \lambda_j x_j$ and $AP = PD$ holds. Moreover, because $P$ is invertible by construction, $P^{-1}AP = D$ and $A$ is diagonalizable.

Continuing from the theorem, we can establish the following corollary.

**Corollary 4.3.24**

A matrix $A_{n\times n}$ with $n$ distinct real eigenvalues is diagonalizable.

*Proof.*

First and foremost, notice that by Theorem 4.3.17, the set of corresponding eigenvectors $\{x_1, \ldots, x_n\}$ of $A$ is linearly independent. Thus by Theorem 4.3.23, the corollary holds.

This theorem also naturally leads to the question if a square matrix can be similar to more than one diagonal matrix. The answer to that question can be found with the theorem below.

**Theorem 4.3.25**

If a square matrix is similar to both diagonal matrices $D_1$ and $D_2$, then the diagonal entries of $D_1$ and $D_2$ are the same with possibly different orderings.

*Proof.*

First and foremost, notice that by Theorem 4.3.13, because $A$ and $D_1$ are similar, they have the same characteristic equation. Similarly, because $A$ and $D_2$ are similar, they also have the same characteristic equation and the characteristic equations of $D_1$ and $D_2$ are equal.

$$\det(D_1 - \lambda I) = \det(D_2 - \lambda I) = 0$$

By construction, notice that $D_1 - \lambda I$ and $D_2 - \lambda I$ are diagonal matrices. Thus by Corollary 2.5.5, the determinant is the product of all $(\lambda - \lambda_k$ for integer $k \in [1, n]$ where $\lambda_k$ are the diagonal entries for $D_1$. Because multiplication is commutative, the diagonal entries for $D_1$ and $D_2$ are the same.

Naturally, we can see that if $A \cong D_1$ and $A \cong D_2$, then $D_1 \cong A$ and $D_1 \cong D_2$. In other words, if $A$ is similar to multiple diagonal matrices, then the diagonal matrices are similar to each other.

Now that we discussed when a matrix is diagonalizable, let's discuss how to diagonalize and its applications.

**Definition 4.3.26**

**Diagonalization** is the process of writing a diagonalizable matrix $A$ in the form of $A = PDP^{-1}$.

Diagonalization is particularly useful as it makes calculations easier. Before concluding our notes, below is the final theorem!

**Theorem 4.3.27**

Let $A_{n\times n}$ be a matrix with real and distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. The matrix $A$ is diagonalizable if and only if the following equation holds.

$$\sum_{i=1}^{k} \dim(S_{\lambda_i}) = n$$

*Proof.*

First, the statement if $\sum_{i=1}^{k} \dim(S_{\lambda_i}) = n$, then $A$ is diagonalizable can be proven. Let $B_i$ be a basis for eigenspace $S_{\lambda_i}$. By definition, vectors $v$ in eigenspace $S_{\lambda_i}$ satisfy $Av = \lambda_i v$. Because $\lambda_i$ are all distinct, no eigenspace shares the same eigenvector and thus no basis of two different eigenspaces shares a vector. In other words, $B$ contains $n$ distinct eigenvectors of $A$. Let $v_i \in S_{\lambda_i}$ be some linear combination of vectors in $B_i$. Notice that if $\sum_{i=1}^{k} v_i = 0$, then the coefficients must be zero since $\{v_1, \ldots, v_k\}$ is linearly independent by Theorem 4.3.17. In other words, all coefficients for the linear combination of $B_i$ that forms $v_i$ must be zero for $\sum_{i=1}^{k} v_i = 0$ to hold. Thus, $B$ is linearly independent and $A$ is diagonalizable by Theorem 4.3.23.

Continuing, the converse can be proven. Assume that $\sum_{i=1}^{k} \dim(S_{\lambda_i})$ $> n$. Continuing from the previous definition, $B \subset \mathbb{R}^n$ and the inequality $\sum_{i=1}^{k} \dim(S_{\lambda_i}) > n$ leads to contradiction. Moreover, if

> $\sum_{i=1}^{k} \dim(S_{\lambda_i}) < n$ then $A$ has no set that contains $n$ linearly independent eigenvectors. Thus, $A$ is not diagonalizable if $\sum_{i=1}^{k} \dim(S_{\lambda_i}) \neq n$.

This is it for eigenvalues and eigenvectors! This note was a particularly long one, but I think it also means that the relationships between matrices and vector spaces are one of the most essential understandings for linear algebra. Below are a few practice problems on the topics that we discussed.

## §4.4 Practice Problems

1. Determine whether a transformation defined as $T(\langle a, b\rangle) = a + b$ is linear. How about $T(\langle a, b\rangle) = ab$?

2. For square matrices $A$ and $B$ of order $n \times n$, show that if $\lambda$ is an eigenvalue of $AB$, then the eigenvalues of $BA$ also include $\lambda$ [Erd20].

3. Find bases for the kernel and the image of $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$.

4. Show that $A \cong B$ if and only if $A^\intercal \cong B^\intercal$.

5. Find the eigenvalues and corresponding eigenspace for $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{bmatrix}$.

Note 5

# Appendix A

Here are the solutions to the practice problems at the end of each note.

## §5.1 Solutions for Note 1

1. It suffices to show that $\alpha(x, y, z) + \beta(x', y', z') \in S$ for any scalars $\alpha$ and $\beta$. With standard vector addition and scalar multiplication, $\alpha(x, y, z) + \beta(x', y', z') = (0, \alpha y + \beta y', \alpha z + \beta z') \in S$. Thus, the set is a vector space.

2. Let $f$ and $g$ be functions in $\mathbb{R}$ such that $f'(0) = 0$ and $g'(0) = 0$. Notice that $(f+g)'(0) = f'(0) + g'(0) = 0$ and $f + g \in S$. Moreover, for any scalars $\alpha$, $(\alpha f)'(0) = \alpha f'(0) = 0$. Thus $\alpha f \in S$ and $S$ is a vector space.

3. First, for any scalars $\alpha$, notice that $\alpha(x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n)$. If each coefficient is divided by $\alpha \neq 0$, the scalar multiplication holds. The case where $\alpha = 0$ naturally holds. Continuing, notice that $(x_1, \ldots, x_n) + (x'_1, \ldots, x'_n) = (x_1 + x'_1, \ldots, x_n + x'_n) \in S$ due to the property of addition. Thus, $S$ is a vector space.

4. Let $V = W_1 \cap \cdots \cap W_n$ where $W_i$ for integer $i \in [1, n]$ is a subspace

of $V$. For $u, v \in V$, notice that since $u, v \in W_i$ for all $i$ and $\alpha u + \beta v \in W_i$ for any scalars $\alpha$ and $\beta$ by construction, $\alpha u + \beta v \in V$ and the intersection of subspaces of a vector space is also a subspace of the vector space.

5. First and foremost, it is evident that $e^{a_1 x}$ is linearly independent since $e^{a_1 x} > 0$ for all $a_1$ and $x$. Therefore, it suffices to show that if the set is linearly independent for $n-1$, then it is linearly independent for $n$. Consider the following equations for some $c_k$.

$$\sum_{k=1}^{n} c_k e^{a_k x} = 0$$
$$\sum_{k=1}^{n} c_k a_k e^{a_k x} = 0$$
$$\sum_{k=1}^{n} c_k a_n e^{a_k x} = 0$$

Subtracting the third equation, which is obtained by multiplying $a_n$ in both sides, to the second equation obtained by differentiating both sides,

$$\sum_{k=1}^{n} c_k (a_k - a_n) e^{a_k x} = \sum_{k=1}^{n-1} c_k (a_k - a_n) e^{a_k x} = 0$$

is obtained. Because $a_k \neq a_n$ for integer $k \in [1, n-1]$ and $\{e^{a_1 x}, \ldots, e^{a_{n-1} x}\}$ is linearly independent $c_1 = \cdots = c_{n-1} = 0$. In other words, $c_n e^{a_n x} = 0$ and $c_n = 0$. Therefore, $\{e^{a_1 x}, \ldots, e^{a_n x}\}$ is linearly independent by induction.

6. By definition, note that $C = A + B$. Therefore, it suffices to show that $A \cap B = \{0\}$. Notice that $A$ and $B$ can be written as $A = \text{span}\{(1,1,1)\}$ and $B = \text{span}\{(20,15,12)\}$. Because $\lambda(1,1,1) = (20,15,12)$ if and only if $\lambda = 0$, $A \cap B = \{0\}$ and $C = A \oplus B$.

7. Consider the following equations for some scalars $k_1, k_2, k_3$.

$$\begin{aligned} k_1(a+b)+k_2(b+c)+k_3(c+a) &= 0 \\ a(k_3+k_1)+b(k_1+k_2)+c(k_2+k_3) &= 0 \end{aligned}$$

The linear combinations of $\{a+b, b+c, c+a\}$ equal zero if and only if the linear combinations of $a$, $b$, and $c$ equal zero. Moreover, from the equation above, $k_1 = -k_2 = k_3 = -k_1$. In other words, $k_1 = k_2 = k_3 = 0$ and $\{a+b, b+c, c+a\}$ is linearly independent.

8. Let $f$ and $g$ be polynomials such that $f(0) = g(0) = 0$. Therefore, $(f+g)(0) = 0$ and $f+g$ is in the set. Moreover, $(\alpha f)(0) = 0$ for any scalars $\alpha$. Therefore, the set of all polynomials in $\mathbb{P}^3$ that includes the origin is a subspace of $\mathbb{P}^3$.

## §5.2 Solutions for Note 2

1. (a) Consider the following computation.

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

(b) Consider the following equation.

$$AB = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} -1 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} -1 & 8 \\ -2 & 7 \end{bmatrix}$$

(c) Because $AB$ is found above, $BA$ could be found.

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

Therefore, the following equation is true.

$$
\begin{aligned}
(AB - BA)^{\mathsf{T}} &= \left( \begin{bmatrix} -1 & 8 \\ -2 & 7 \end{bmatrix} - \begin{bmatrix} 3 & -2 \\ 6 & 0 \end{bmatrix} \right)^{\mathsf{T}} = \left( \begin{bmatrix} -4 & 10 \\ -8 & 7 \end{bmatrix} \right)^{\mathsf{T}} \\
&= \begin{bmatrix} -4 & -8 \\ 10 & 7 \end{bmatrix}
\end{aligned}
$$

2. By the definition of matrix multiplication, $a_2 B = [4, 5, 6]$.

3. Notice that $A$ is a $2 \times 2$ square matrix. Therefore, define $A$ as the following.

$$
A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}
$$

Therefore, the following equation is obtained.

$$
\begin{aligned}
2A - 3A^{\mathsf{T}} &= \begin{bmatrix} 2a_{11} & 2a_{12} \\ 2a_{21} & 2a_{22} \end{bmatrix} - \begin{bmatrix} 3a_{11} & 3a_{21} \\ 3a_{12} & 3a_{22} \end{bmatrix} \\
&= \begin{bmatrix} 2a_{11} - 3a_{11} & 2a_{12} - 3a_{21} \\ 2a_{21} - 3a_{12} & 2a_{22} - 3a_{22} \end{bmatrix} = \begin{bmatrix} -1 & -5 \\ 0 & -4 \end{bmatrix}
\end{aligned}
$$

Continuing, the elements of $A$ could be found.

$$
\begin{aligned}
2a_{11} - 3a_{11} &= -1 \\
2a_{12} - 3a_{21} &= -5 \\
2a_{21} - 3a_{12} &= 0 \\
2a_{22} - 3a_{22} &= -4
\end{aligned}
$$

Therefore, $a_{11} = 1$, $a_{12} = 2$, $a_{21} = 3$, and $a_{22} = 4$. Thus, $A$ can be defined as the following.

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

4. For any square matrix $S$, define $A$ and $B$ as the following.

$$A = \frac{S + S^\intercal}{2}, \quad B = \frac{S - S^\intercal}{2}$$

Notice that $S = A + B$. Therefore, it suffices to show that $A$ is symmetric and $B$ is skew-symmetric. By Theorem 2.4.4, the following equations hold.

$$A^\intercal = \frac{S + S^\intercal}{2}, \quad B^\intercal = \frac{S^\intercal - S}{2} = -B$$

Thus, all square matrices can be represented as the sum of a symmetric and skew-symmetric matrices.

5. By definition, there exist $n, m \in \mathbb{Z}$ such that $A^n = 0$ and $B^m = 0$. Because $AB = BA$, $A^{nm}B^{nm} = 0$ can be written as $(AB)^{nm} = 0$ and $AB$ is nilpotent.

6. By Theorem 2.4.4, $(AB + (AB)^\intercal)^\intercal = (AB)^\intercal + AB = AB + (AB)^\intercal$. Therefore, $AB + (AB)^\intercal$ is symmetric.

7. An easy trap for this problem is that you might be tempted to do $\det(A^2 - B^2) = \det(A + B)\det(A - B)$. However, notice that this does not hold in general as $AB \neq BA$ in general.

Consider the following matrix $P$ and its inverse $P^{-1}$.

$$P = \begin{bmatrix} I & I \\ I & -I \end{bmatrix}, \quad P^{-1} = \frac{1}{2}\begin{bmatrix} I & I \\ I & -I \end{bmatrix}$$

Continuing,

$$\begin{aligned} P^{-1}\begin{bmatrix} A & B \\ B & A \end{bmatrix}P &= \frac{1}{2}\begin{bmatrix} A+B & B+A \\ A-B & B-A \end{bmatrix}\begin{bmatrix} I & I \\ I & -I \end{bmatrix} \\ &= \begin{bmatrix} A+B & 0 \\ 0 & A-B \end{bmatrix} \end{aligned}$$

holds and $\begin{bmatrix} A & B \\ B & A \end{bmatrix} \cong \begin{bmatrix} A+B & 0 \\ 0 & A-B \end{bmatrix}$. Therefore, the determinant of the original matrix is $\det(A+B)\det(A-B)$.

8. By definition, $A$ is skew-symmetric if $A^\intercal = -A$. Notice that $(A^3)^\intercal = (A^\intercal)^3$ by Theorem 2.4.4. Therefore, $(A^3)^\intercal = (A^\intercal)^3 = -A^3$. Thus, if $A$ is skew-symmetric, then $A^3$ is also skew-symmetric.

## §5.3 Solutions for Note 3

1. If $A$ is invertible, then $A^{-1}Ax = A^{-1}b$ and $x = A^{-1}b$. Therefore, the system $Ax = b$ is always consistent.

2. Consider the following representation of the system.

$$\begin{aligned} x + 2y &= 1 \\ 2x + 4y &= k \end{aligned}$$

Notice that if $k = 2$, then there exist infinitely many solutions while $k \neq 2$ gives no solutions. Thus the system never has a unique solution.

3. First and foremost, consider the following augmented matrix.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 10 \\ 3 & 4 & -1 & 2 \\ -1 & -3 & 1 & 4 \end{array}\right]$$

Using elementary row operations, the augmented matrix can be transformed as following.

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 10 \\ 3 & 4 & -1 & 2 \\ -1 & -3 & 1 & 4 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 10 \\ 0 & 1 & -4 & -28 \\ -1 & -3 & 1 & 4 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 1 & 1 & 10 \\ 0 & 1 & -4 & -28 \\ 0 & -2 & 2 & 14 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|c} 1 & 1 & 1 & 10 \\ 0 & 1 & -4 & -28 \\ 0 & 0 & -6 & -42 \end{array}\right]$$

Therefore, $-6z = -42$ and $z = 7$. Moreover, $y - 4z = -28$ and $y = 0$. Finally, $x + y + z = 10$ and $x = 3$. Therefore, the solution to the system is $(3, 0, 7)$.

4. First and foremost, let $A = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 4 & -1 \\ -1 & -3 & 1 \end{bmatrix}$. Using elementary row operations, $A$ can be transformed as following.

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

$$\rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -4 \\ 0 & 0 & -6 \end{bmatrix}$$

In other words, $A$ can be represented as the matrix multiplication below.

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

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

Consider the following system for equation for $Ly = b$.

$$\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 4 \end{bmatrix}$$

It is evident that $x' = 10$, $y' = 2 - 3x' = -28$, and $z' = 4 + x' + 2y' = -42$. Continuing, the system for $Ux = y$ can be solved.

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -4 \\ 0 & 0 & -6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 10 \\ -28 \\ -42 \end{bmatrix}$$

Therefore, $z = 7$, $y = 0$, and $x = 3$ holds and the solution for the system is $(3, 0, 7)$ as shown in the previous problem.

5. To find the inverse of $A = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 4 & -1 \\ -1 & -3 & 1 \end{bmatrix}$, consider the following augmentation and transformations.

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 3 & 4 & -1 & 0 & 1 & 0 \\ -1 & -3 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -4 & -3 & 1 & 0 \\ -1 & -3 & 1 & 0 & 0 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -4 & -3 & 1 & 0 \\ 0 & -2 & 2 & 1 & 0 & 1 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -4 & -3 & 1 & 0 \\ 0 & 0 & -6 & -5 & 2 & 1 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -4 & -3 & 1 & 0 \\ 0 & 0 & 1 & \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{6} & \frac{2}{3} & \frac{5}{6} \\ 0 & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \end{array}\right]$$

Therefore, $A^{-1} = \begin{bmatrix} -\frac{1}{6} & \frac{2}{3} & \frac{5}{6} \\ \frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \end{bmatrix}$ and the system could be solved by multiplying $A^{-1}$ and $b$.

$$\begin{bmatrix} -\frac{1}{6} & \frac{2}{3} & \frac{5}{6} \\ \frac{1}{3} & -\frac{1}{3} & -\frac{2}{3} \\ \frac{5}{6} & -\frac{1}{3} & -\frac{1}{6} \end{bmatrix} \begin{bmatrix} 10 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 7 \end{bmatrix}$$

Thus, the solution is $(3, 0, 7)$, which is consistent with the previous problems.

## §5.4 Solutions for Note 4

1. For transformation $T(\langle a, b\rangle) = a + b$, consider the following equations for some scalars $\alpha$ and $\beta$.

$$\begin{aligned} T(\alpha\langle a_1, b_1\rangle + \beta\langle a_2, b_2\rangle) &= T(\langle \alpha a_1, \alpha b_1\rangle + \langle \beta a_2, \beta b_2\rangle) \\ &= \alpha a_1 + \alpha b_1 + \beta a_2 + \beta b_2 \\ &= \alpha(a_1 + b_1) + \beta(a_2 + b_2) \\ &= \alpha T(\langle a_1, b_1\rangle) + \beta T(\langle a_2, b_2\rangle) \end{aligned}$$

Therefore by the definition of linear transformation, $T$ is linear.

For $T(\langle a,b\rangle) = ab$, consider the following for arbitrary scalars $\alpha$ and $\beta$.

$$\begin{aligned} T(\alpha\langle a_1,b_1\rangle + \beta\langle a_2,b_2\rangle) &= T(\langle \alpha a_1, \alpha b_1\rangle + \langle \beta a_2, \beta b_2\rangle) \\ &= T(\langle \alpha a_1 + \beta b_2, \alpha b_1 + \beta b_2\rangle) \\ &= (\alpha a_1 + \beta b_2)(\alpha b_1 + \beta b_2) \end{aligned}$$

Moreover,

$$\alpha T(\langle a_1,b_1\rangle) + \beta T(\langle a_2,b_2\rangle) = \alpha a_1 b_1 + \beta a_2 b_2$$

holds. Because $(\alpha a_1 + \beta b_2)(\alpha b_1 + \beta b_2) \neq \alpha a_1 b_1 + \beta a_2 b_2$ in general, the transformation is not linear.

2. By definition, if $\lambda$ is an eigenvalue of $AB$, then there exists some nonzero vector $v$ such that $ABv = \lambda v$. First, let $\lambda \neq 0$. Notice that $u = Bv \neq \mathbf{0}$. Therefore, $BAu = BA(Bv) = B(ABv) = B(\lambda v) = \lambda(Bv) = \lambda u$ and $\lambda$ is also an eigenvalue of $BA$. Continuing if $\lambda = 0$, then $\det(AB) = \det(A)\det(B) = 0 = \det(B)\det(A) = \det(BA)$. In other words, if $AB$ is singular then $BA$ is also singular and contains the eigenvalue $\lambda$. Therefore, $\lambda$ is an eigenvalue of $AB$ if and only if it is an eigenvalue of $BA$.

3. First, to find bases for the kernel of the matrix, the following equation must be solved.

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solving the system, $y = 0$ and $x = -z$. Therefore the basis for the kernel is $\left\{ \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$. Continuing, the basis for the image can be found. Consider the following equation.

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x+y+z \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 1 \\ 1 \end{bmatrix} + z \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Therefore,

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

is a basis of the image of the given matrix.

4. By definition, $A \cong B$ if and only if there exists an invertible matrix $P$ such that $A = P^{-1}BP$. Continuing, $A = P^{-1}BP$ if and only if $A^\intercal = (P^{-1}BP)^\intercal = P^\intercal B^\intercal (P^{-1})^\intercal = P^\intercal B^\intercal (P^\intercal)^{-1}$. In other words, $A \cong B$ if and only if $A^\intercal \cong B^\intercal$.

5. To find the eigenvalues for the given matrix say $A$, it suffices to solve $\det(A - \lambda I) = 0$ where $\lambda$ is an eigenvalue.

$$\det \left( \begin{bmatrix} 1-\lambda & 2 & 3 \\ 0 & 1-\lambda & 0 \\ 0 & 1 & 2-\lambda \end{bmatrix} \right) = 0$$

Continuing with the expansion along the first column,

$$\begin{aligned} &\det \left( \begin{bmatrix} 1-\lambda & 2 & 3 \\ 0 & 1-\lambda & 0 \\ 0 & 1 & 2-\lambda \end{bmatrix} \right) \\ &= (1-\lambda) \det \left( \begin{bmatrix} 1-\lambda & 0 \\ 1 & 2-\lambda \end{bmatrix} \right) \\ &= (1-\lambda)(1-\lambda)(2-\lambda) \end{aligned}$$

Therefore, the eigenvalues are 1 and 2. For eigenvalue 1, it suffices to solve the equation $Av = v$ or $(A - I)v = 0$. Consider the following system.

$$\begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Therefore, $2y + 3z = 0$ and $y + z = 0$. In other words, $y = z = 0$. Thus, $S_1 = \text{span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ holds for $\lambda = 1$. Continuing with $\lambda = 2$, it suffices to solve $Av = 2v$ or $(A - 2I)v = 0$.

$$\begin{bmatrix} -1 & 2 & 3 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Solving the system, $y = 0$ and $x = 3z$. Thus, $S_2 = \text{span} \left\{ \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\}$ holds for $\lambda = 2$.

Note 6

# Appendix B

I wanted to include my final thoughts on these notes and notice on some topics that were not covered here, and I guess I can use the help of appendix.

First, as mentioned in the prefix, these are the notes that I took while learning the essence of linear algebra in 8 weeks. That being said, while most of the fundamental concepts were covered, other topics like dual space, cross product, euclidean inner product, and more advanced concepts including Markov chains and Jordan canonical forms are not included here. After the initial release, I will continue to update these notes for the second edition as I will likely take linear algebra course again in college as a student who will major in pure math. In the meantime, I highly recommend watching 3Blue1Brown series on linear algebra as it really helped me visually understand different concepts and formulas.

As I also study different concepts in machine learning, I realized how important linear algebra is outside "mathematics." Literally, vectors are used everywhere including, but not limited to encoding/decoding and backpropagation. One of the reasons why I find math so fascinating is due to duality and applications that just seems so perfect. Matrices can be interpreted with heavy computation while it can also be interpreted geometrically with change in area. The two seemingly unrelated topics actually imply the same exact concept. Similarly, what else can we use instead of matrix multiplication and addition for forward pass in machine learning? The way it seems so intuitive and "perfect" is I think one of the most beautiful side of mathematics. Really, if you completed these notes, I highly recommend learning math behind machine learning if you like finding different applications of mathematics. More information can be found in the machine learning section of LibreNotebook.

Note 7

# References


[Ric07] Gabriel B. Costa Richard Bronson. *Linear Algebra*. Elsevier, 2007. ISBN: 9780080510262.

[Erd20] J.M. Erdman. *Exercises And Problems In Linear Algebra*. World Scientific Publishing Company, 2020. ISBN: 9789811220425.

[Kan26] Margarita Kanarsky. *Linear Algebra (XM511)*. 2026.