CAPP Camp 2022: Getting Started with Linear Algebra

Steven Buschbach*

August 24, 2022

1 Introduction

This document serves as a guide to 2022's CAPP Camp lessons in linear algebra. The goal of CAPP Camp is to prepare incoming CAPP students for their linear algebra coursework by reviewing some foundational concepts, with a particular emphasis on the mechanics of working simple matrix and vector operations. Along with calculus and statistics, linear algebra is one of the most useful branches of applied mathematics, particularly for understanding many common machine learning techniques and for processing data efficiently. I will try to supplement the mathematics with CAPP-relevant examples, where possible.

These notes do not presuppose any particular math background other than a familiarity with high school algebra. For some students with a stronger math background, these notes may mostly cover material they are familiar with. I have included a few extra "challenge problems" for these students to test their understanding of the concepts. Many of the practice questions are adapted from Gilbert Strang's book, *Linear Algebra for Everyone*, and the material itself loosely follows the first chapter of that book. Another useful resource is Stephen Boyd and Lieven Vandenberghe's book *Introduction to Applied Linear Algebra*, which is freely available online. Lastly, for visual learners, I recommend Grant Sanderson's series of YouTube videos on linear algebra.

2 Introduction to Mathematical Proofs

Before diving into linear algebra, I want to familiarize you with the basics of mathematical proofs. In many high school and some undergrad math courses, we are simply asked to perform some mechanical calculations using formulas given to us by professors. In advanced math courses, we are more concerned with proving broad, generalized statements.

Often times, we can structure it as saying that conditions A imply result B. For example, we know the Pythagorean theorem gives us a formula for right triangle side lengths $a^2 + b^2 = c^2$. If we wanted to prove the Pythagorean theorem, we would need to show that for arbitrary side lengths a and b of a right triangle (this is our condition A), we always have hypotenuse of length $c = \sqrt{a^2 + b^2}$ (this is our result B). We can write this as $A \implies B$, which we describe as "A implies B."

How do we show $A \implies B$? There are often many ways to prove something, but I will describe a few common techniques:

1. **Direct proof**: This is the most intuitive way people attempt to approach proofs. We begin with condition(s) A and, step-by-step, use existing facts or definition to arrive at result B.

   For example, we can prove that the sum of any two even integers is even through a short direct proof. Consider any two even integers x, y. Since they are even, they can be written as $x = 2a$ and

*Please direct any questions or comments to me at sbuschbach@uchicago.edu

$y = 2b$ for some a, b. Then their sum can be written $x + y = 2a + 2b = 2(a + b)$. Since $x + y = 2(a + b)$, the sum of x and y is divisible by 2, and thus it is also an even number.

2. **Proof by contradiction**: In this sort of proof, we assume that A holds but that B *does not* hold, and then we show that it leads to some contradiction or nonsense. Since A holding and B not holding leads to a contradiction, it must therefore be that $A \implies B$. This is one of those most common sorts of proof you will see after direct proofs.

   As an example, consider the statement "There is no smallest rational number greater than 0." We can suppose that there was some rational number x that is the smallest rational number greater than 0. Since x is rational, it can be written $x = \frac{p}{q}$. But if we consider $y = \frac{p}{2q}$, this is a rational number that is smaller than x. So it cannot be that x is the smallest rational number greater than 0. Therefore, there is no smallest rational number greater than 0.

3. **Proof by induction**: You will use this style of proof a lot if you take Discrete Math. It commonly is used when we want to prove a statement $A \implies B$ for an infinite number of cases. If we can show that the statement holds for some "base case", and then use logical induction to establish a relationship between each case, we can set off a mathematical chain of dominoes that proves the statement for every case.

   As an example, consider the statement "for all positive integers n, the integer $2n - 1$ is odd." This is very easy to show for $n = 1$, since $2(1) - 1 = 1$ is trivially odd. So we use $n = 1$ as our base case. For our inductive step, we assume that the statement holds for some integer n and show that this implies it is also true for $n+1$. In other words, we want to show that $2n-1$ being odd implies that $2(n+1)-1$ is odd. First note that $2(n + 1) - 1 = 2n + 1 = (2n - 1) + 2$. Since we are assuming that $2n - 1$ is odd, and adding 2 to any odd number gives us another odd number, then it must be that $2(n+1)-1$ is odd. So with our base case of $n = 1$, the inductive step shows that this statement holds for $n = 2$, which then shows that it holds for $n = 3$, which then shows it holds for $n = 4$, and so on ad infinitum.

4. **Proof by contrapositive**: This style of proof is not as common as the others I mention, but it is another possible approach in case you get stuck. The idea here is that instead of arguing that $A \implies B$ directly, we argue that "not B" implies "not A."

3 Vectors

3.1 Understanding Vectors

Consider the set of all real numbers, commonly denoted as $\mathbb{R}$. Ignoring the technical definition, we can think of $\mathbb{R}$ as the full set of non-complex numbers one can imagine on a number line (e.g., $7, -12, \pi, \sqrt{2}, 1.1, \ldots$). We can write $x \in \mathbb{R}$ to denote that x is some arbitrary real number. More precisely, "$x \in \mathbb{R}$" means that x is some element in the set of real numbers.

Then in one sense, we can think of a vector as a generalization of numbers to multiple dimensions. A **vector** is an ordered list of numbers. The **dimension** (or **length**) of a vector is the number of elements it contains. So, for example,

$$\mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ 7 \end{bmatrix} \qquad \text{and} \qquad \mathbf{w} = \begin{bmatrix} 4 \\ 5 \\ 2 \\ 6 \\ 0 \end{bmatrix}$$

are 3-dimensional and 5-dimensional vectors, respectively. Since each element of a vector can be some arbitrary real number, we can write that in the examples above, $\mathbf{v} \in \mathbb{R}^3$ and $\mathbf{w} \in \mathbb{R}^5$. More generally, $\mathbf{v} \in \mathbb{R}^n$ is an arbitrary n-dimensional vector.[1] Note that a one-dimensional vector is just a number. There is a special word for this that will we use: a number (or 1-dimensional vector) is a **scalar**.

We can also define vectors geometrically. Specifically, we visualize a two-dimensional vector as an arrow with its base at the origin (0,0) and its head at the point defined by the vector elements. For example, Figure 1 shows the vectors $u = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $v = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ on a Euclidean plane.

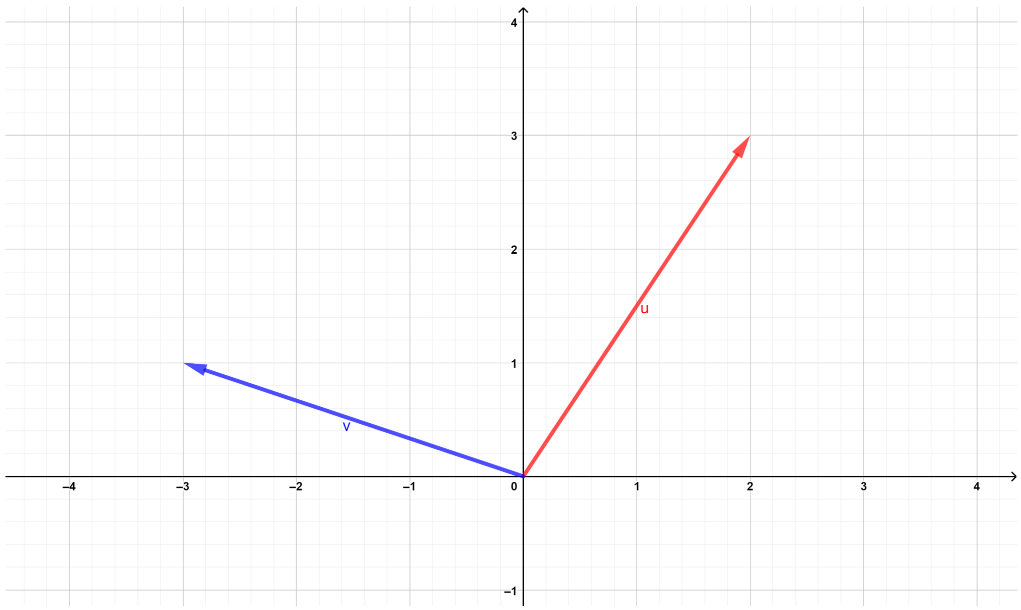

Figure 1: Two 2-dimensional vectors on a Euclidean plane

While intuitive, the geometric perspective of vectors falls short in higher dimensions. We can imagine arrows pointing in 2- or 3-dimensional space, but we are not equipped to think about geometry in 4-dimensional space intuitively. Still, the geometric perspective makes clear the two fundamental properties that define any given vector: (1) vector **direction** and (2) vector **length** (also called **magnitude**). We will come back to these properties.

Let's give some examples of where vectors could be used:

- <u>Color</u>: Computers can represent the RGB color of a pixel with a 3-dimensional vector. The first element defines the intensity of Red that color, the second element defines the intensity of Green, and the third element defines the intensity of Blue. A computer can represent a wide arrange array of colors this way. See Figure 2 for examples.

- <u>Word Counts</u>: Suppose there are n words in a dictionary. Then we can represent any text or document as an n-dimensional vector where the i-th entry is the number of times that the i-th word in the dictionary appears in that document. Note that this representation ignores the order of words, but it is can still be informative about a text. Also note that this vector is likely to be **sparse**, meaning that most entries will be zero. Any given document will only use a small fraction of the words that appear in the dictionary.

- <u>Investor Portfolios</u>: Suppose there are n available stocks for an investor to purchase. If each entry of a vector represents the number of shares that the investor owns of each stock, then an arbitrary vector $v \in \mathbb{R}^n$ represents a portfolio, and $\mathbb{R}^n$ is the space of all possible portfolios.

[1] Here, I only consider finite-dimensional real-valued vectors as the simplest case. But it is possible to consider infinite-dimensional vectors as well as vectors containing complex numbers. These complicate things, so for this introductory course you will not have to worry about them.

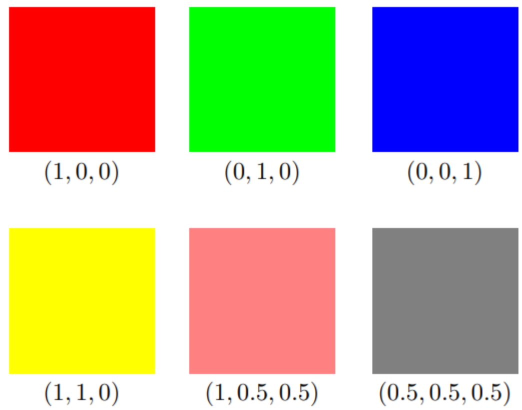

Figure 2: Six RGB color vectors, courtesy of Boyd and Vandenberghe, *Introduction to Applied Linear Algebra*

- Zero and Ones Vectors: Two common vector types that appear often enough to have special notation are zero vectors and ones vectors. We write $\mathbf{0}_n$ to describe an n-dimensional vector with all entries equal to zero, and we write $\mathbf{1}_n$ to describe an n-dimensional vector with all entries equal to one.

- Standard Basis Vectors: This is another common vector. A standard basis vector has a single 1 entry and the rest zero entries. In n-dimensions, there are n standard basis vectors. The i-th standard basis vector is written $\mathbf{e}_i$, and its i-th entry is a 1. Mathematically, we define the entries of a standard basis vector by

$$(\mathbf{e}_i)_j = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

  More concretely, consider the example of the 3 possible standard vectors in 3 dimensions. These are

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

  Sometimes, we use different notation for standard basis vectors in 2 or 3 dimensions. In two dimensions, $\mathbf{e}_1$ and $\mathbf{e}_2$ can be written as $\mathbf{i}$ and $\mathbf{j}$, respectively. In three dimensions, they become $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$.

Typically, we will be focusing on vertically-oriented vectors, known as **column vectors**. A **row vector** is horizontally-oriented. We can turn a column vector into a row vector with the **transpose** operation, denoted with a superscript T. For example,

$$\begin{bmatrix} 3 \\ 4 \\ 9 \end{bmatrix}^T = \begin{bmatrix} 3 & 4 & 9 \end{bmatrix}$$

3.2 Linear Combinations of Vectors

First we define **scalar multiplication of a vector**. Suppose we have some vector $\mathbf{v} \in \mathbb{R}^n$ and some scalar $c \in \mathbb{R}$. We can multiply the vector by the scalar $c\mathbf{v}$ simply by multiplying each entry of $\mathbf{v}$ by c. For example,

$$2 \begin{bmatrix} 3 \\ 1 \\ 7 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 14 \end{bmatrix}$$

Geometrically, scalar multiplication *changes the length of a vector but does not change the direction.* Consider the vector $\mathbf{u} = \mathbf{1}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. In Figure 3, we illustrate the vectors defined by $3\mathbf{u}$ and $(-1)\mathbf{u}$. More generally, we want to think about *all possible combinations of scalar multiplication*. In the above example, the space of all possible combinations of $c\mathbf{u}$ gives us a line of slope 1 through the origin. More generally, for an n-dimensional vector $v \in \mathbb{R}^n$, the space described by multiplying v by an arbitrary scalar $c \in \mathbb{R}$ is a line in n-dimensional space going through the origin.[2]

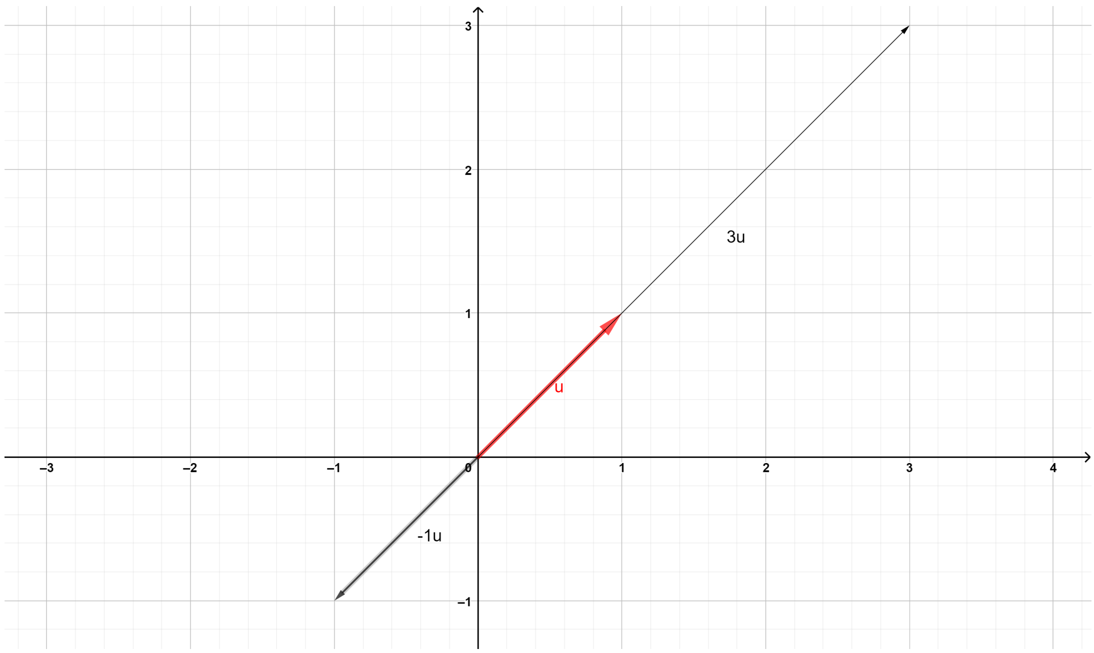

Figure 3: Scalar multiplication of the 2-dimensional ones vector

Now we define **vector addition**. Mechanically, vector addition is very intuitive - we can add two vectors v and w by adding each pair of elements. For example,

$$\begin{bmatrix} 4 \\ 9 \end{bmatrix} + \begin{bmatrix} 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 4+1 \\ 9+(-4) \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

Just as we have $a - b = a + (-1) \times b$ for scalars, **vector subtraction** is equivalent to vector addition after scalar multiplying by -1, i.e., $\mathbf{v} - \mathbf{w} = \mathbf{v} + (-1)\mathbf{w}$.

Note a few important properties about vector addition and subtraction. First, we can only add or subtract vectors that have the same dimension. Second, vector addition and subtraction are commutative. We have that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. This is important to note because matrix multiplication, as we will see, is not commutative.

The geometric intuition for vector addition is straightforward in 2 dimensions. If we add two vectors $\mathbf{u} + \mathbf{v}$, then the resulting vector is *the diagonal of the parallelogram formed by the two vectors*. In other words, we can find the head of the vector described by $\mathbf{u} + \mathbf{v}$ by taking $\mathbf{u}$ and stacking its base at the head of $\mathbf{v}$. See Figures 4 and 5 as a visual aids.

A **linear combination** just combines vector addition with scalar multiplication. So for some $c, d \in \mathbb{R}$, $c\mathbf{v} + d\mathbf{w}$ is a linear combination of $\mathbf{v}$ and $\mathbf{w}$. With explicit values for vectors and scalars, calculating a linear combination is simple arithmetic. For example, given

$$u = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \quad v = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad w = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$$

[2]This statement is technically not completely true. When is it not true?

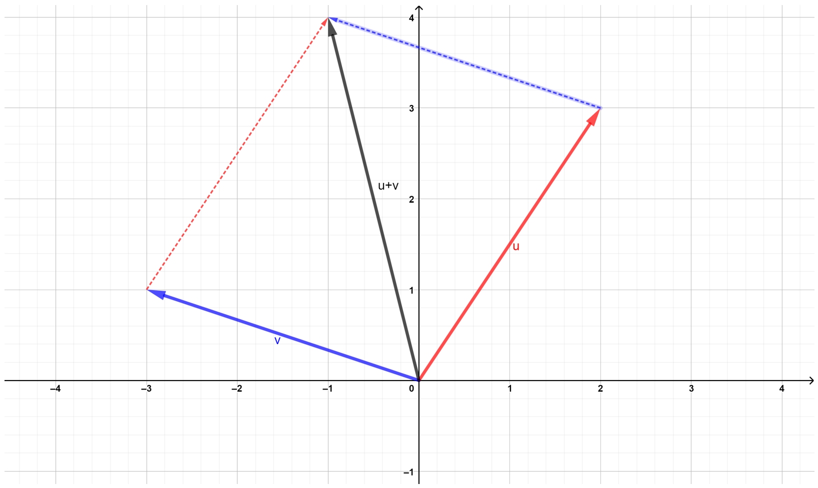

Figure 4: Adding [2,3] and [-3,1] gives the vector [-1,4]

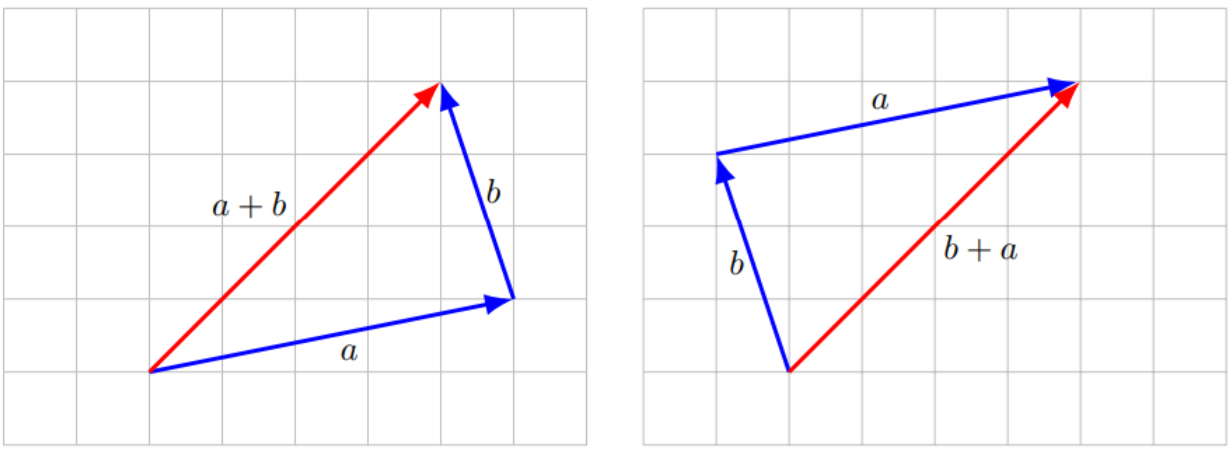

Figure 5: Vector addition visually, courtesy of Boyd and Vandenberghe, *Introduction to Applied Linear Algebra*

We can calculate one example of linear combination

$$\mathbf{u} + 4\mathbf{v} - 2\mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$$

Even more importantly, we want to be able to consider the **space of all linear combinations**. Consider now three arbitrary 3-dimensional vectors, $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$ and three arbitrary scalars, $c, d, e \in \mathbb{R}$:

1. The space of all combinations $c\mathbf{u}$ is a **1-dimensional line** that passes through the origin (0,0,0). We already showed this when discussing scalar multiplication.

2. Typically, the space of all combinations $c\mathbf{u} + d\mathbf{v}$ is a **2-dimensional plane** that passes through the origin (0,0,0).

3. Typically, the space of all combinations $c\mathbf{u} + d\mathbf{v} + e\mathbf{w}$ fills the whole **3-dimensional space**.

Importantly, note that I said *typically.* It is not always the case that the space of all linear combinations of n vectors is n-dimensional. Consider the two vectors

$$u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad v = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$$

Note that $\mathbf{v} = 2\mathbf{u}$, so $\mathbf{v}$ is already on the line that describes all combinations of $c\mathbf{u}$. Thus, the set of all linear combinations $c\mathbf{u} + d\mathbf{v}$ is still going to be a line! Similarly, if $\mathbf{w}$ is on a plane described by the combinations of $c\mathbf{u}+d\mathbf{v}$, then the space described by the combinations $c\mathbf{u}+d\mathbf{v}+e\mathbf{w}$ will still be a plane, not a 3-D space. If all the vectors under consideration are zero vectors, then the space of linear combinations is just the origin, a single point.

The space of all linear combinations of some set of vectors is called the **span** of that set of vectors.

3.3 Vector Length and Angles using Dot Products

We define the **dot product** (also called **inner product**) between two n-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$ to be the sum of elementwise products. That is,

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$$

Rather than the $\cdot$ notation, it is also possible to denote dot product using transposes. Instead of $\mathbf{u} \cdot \mathbf{v}$, we can write $\mathbf{u}^T\mathbf{v}$. Dot products are commutative and distributive over addition. An example of dot product computation is given below:

$$\mathbf{v} = \begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} \quad \implies \quad \mathbf{v} \cdot \mathbf{w} = (-1)(1) + (2)(0) + (3)(-3) = -1 + 0 - 9 = -10$$

Or consider the case of a linear regression with k variables. For an observation with outcome y and characteristics $x_1, x_2, \ldots, x_k$, we can model a linear relationship between the outcome and characteristics using coefficient vector β. That is,

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + \epsilon = \mathbf{x}^T\beta + \epsilon$$

where ϵ is some "error" term. Note how much simpler it is to express this model as the dot product between two vectors $\mathbf{x}$ and β compared to writing it as the sum of scalar multiplication.

Lastly, consider an example in economics. Suppose we have a firm that produces m goods to sell, and in doing so it uses n input goods. We can stack these goods into an $m+n$ dimensional quantity vector $\mathbf{q}$, where input good amounts have a negative sign. With price vector $\mathbf{p}$, we can calculate the firm's total profit as the dot product between $\mathbf{q}$ and $\mathbf{p}$. That is, total profit is $q \cdot p = p_1 q_1 + \ldots + p_{m+n} q_{m+n}$.

We can use a vector's dot product with itself to define its length. The **length** (or **magnitude**) of a vector $\mathbf{v} \in \mathbb{R}^n$ is defined as

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \ldots v_n^2}$$

In 2 or 3 dimensions, the vector magnitude fits our intuitive understanding of Euclidean distance.[3] In 2 dimensions, this fits perfectly with our understanding of the Pythagorean theorem. That is, given a right triangle with sides a and b, the length of a hypotenuse c is defined by the formula $c^2 = a^2 + b^2$.

We can scalar divide any vector by its length to create a vector of length 1. We say that any vector $\mathbf{u}$ with length $||\mathbf{u}|| = 1$ is a **unit vector**. Given an arbitrary vector $\mathbf{v}$, if $\mathbf{v} \neq 0$, then

$$u = \frac{v}{||v||}$$

[3] Another word for the function calculating vector length is a vector norm. Here, we only focus on the standard Euclidean norm, but there are other vector norms you may encounter.

is the corresponding unit vector. Our standard basis vectors **i** and **j** in 2-dimensions are unit vectors. More generally, in two dimensions, we can define the unit vector **u** that makes the angle θ with the x-axis as

$$\mathbf{u} = \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix}$$

This comes from the fact that $\cos^2(\theta) + \sin^2(\theta) = 1$ for any angle θ.

Dot products are also related to the angle between any two vectors. We can calculate the angle θ between any two vectors **u** and **v** using the **cosine formula** below:

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

Note a few things about this formula:

- If two vectors are **perpendicular** (also called **orthogonal**), there is a 90° right angle between them. Since $\cos(90°) = 0$, then this means that the dot product between them is 0. This is an important fact: **perpendicular vectors have $\mathbf{v} \cdot \mathbf{w} = \mathbf{0}$**. We can also show that this means that $||v + w||^2 = ||v||^2 + ||w||^2$. To see this, note that

  $$||v + w||^2 = (v + w) \cdot (v + w) = v \cdot v + v \cdot w + w \cdot v + w \cdot w = ||v||^2 + ||w||^2$$

  This is a useful fact that we can use to take a vector and find vector(s) perpendicular to it. For example, consider the vector $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Arbitrary vector $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \in \mathbb{R}^2$ is perpendicular if we have

  $$4a + 2b = 0$$

  There are many pairs of a and b we can use to make this equation true. For example, $a = 1$ and $b = -2$. Our system is underdetermined, and so there are many possible vectors perpendicular to **u**.

- If two vectors are both unit vectors (that is, both of them have length 1), then the denominator of the left-hand side of the cosine formula goes away, and we have that $\mathbf{u} \cdot \mathbf{v} = \cos(\theta)$.

- The sign of the dot product gives us an idea about whether the angle between two vectors is **acute** or **obtuse**. Since $\cos(\theta)$ is positive when $\theta < 90°$, then the dot product must be positive when the angle between two vectors is acute. Similarly, the dot product must be negative when the angle between two vectors is obtuse. For example, the following dot product

  $$\mathbf{u} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \quad \mathbf{u} \cdot \mathbf{v} = (4)(-1) + (1)(3) = -1$$

  tells us that the angle between **u** and **v** is acute. If we wanted, we could use the cosine formula to calculate the exact angle.

The last two items in this section to cover are two useful inequalities. The is called the **Cauchy-Schwarz inequality**. It states that the absolute value of the dot product between two vectors is less than or equal to the product of the their two lengths. More precisely, for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$

$$|\mathbf{u} \cdot \mathbf{v}| \leq ||\mathbf{u}||\ ||\mathbf{v}||$$

Using the fact that $|\cos(\theta)| \leq 1$, this is a direct result of the cosine formula.

Lastly, we have the **triangle inequality**. This is a direct result of Cauchy-Schwarz. It states that

$$||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$$

The triangle inequality also has a nice geometric interpretation from which it gets its name. That is, for any triangle, the sum of length of any two of its sides must be greater than or equal to the length of its third side.

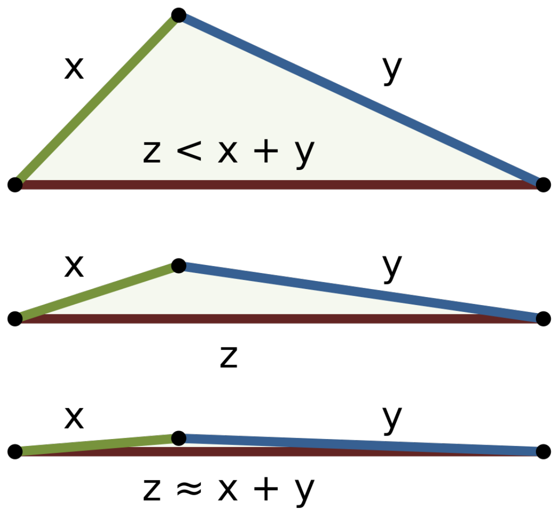

Figure 6: The triangle inequality, visualized. Courtesy of Wikipedia

3.4 Vector Review Questions

3.4.1 Vector Addition and Linear Combinations

1. Draw $\mathbf{v} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$ and $\mathbf{v} + \mathbf{w}$ and $\mathbf{v} - \mathbf{w}$ in a single xy-plane.

2. If $\mathbf{v} + \mathbf{w} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ and $\mathbf{v} - \mathbf{w} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$, compute and draw the vectors $\mathbf{v}$ and $\mathbf{w}$.

3. Describe geometrically (line, plane, or all of $\mathbb{R}^3$) all linear combinations of
$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 \\ 6 \\ 9 \end{bmatrix}$$

4. Describe geometrically (line, plane, or all of $\mathbb{R}^3$) all linear combinations of
$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$$

5. Describe geometrically (line, plane, or all of $\mathbb{R}^3$) all linear combinations of
$$\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$$

6. What values of c and d give
$$c\begin{bmatrix} 1 \\ 2 \end{bmatrix} + d\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 14 \\ 8 \end{bmatrix}$$
Express this question as two equations for the coefficients c and d in the linear combination.

7. **Challenge**: Write down three equations for c, d, e so that $c\mathbf{u} + d\mathbf{v} + e\mathbf{w} = \mathbf{b}$. Find c, d, e.
$$\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

8. **Challenge**: How many corners does a cube have in 4 dimensions? How many 3D faces? How many edges? A typical corner is $(0, 0, 1, 0)$. A typical edge goes to $(0, 1, 0, 0)$.

3.4.2 Vector Lengths and Angles Using Dot Products

9. Calculate the dot products $\mathbf{u} \cdot \mathbf{v}$ and $\mathbf{u} \cdot \mathbf{w}$ and $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w})$ for

$$\mathbf{u} = \begin{bmatrix} -0.6 \\ 0.8 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

10. For the vectors in the previous problem, compute the lengths $||\mathbf{u}||$ and $||\mathbf{v}||$ and $||\mathbf{w}||$. Check the Cauchy-Schwarz inequality for $|\mathbf{u} \cdot \mathbf{v}| \leq ||\mathbf{u}||\ ||\mathbf{v}||$ and $|\mathbf{v} \cdot \mathbf{w}| \leq ||\mathbf{v}||\ ||\mathbf{w}||$.

11. Find unit vectors in the directions of $\mathbf{v}$ and $\mathbf{w}$ in Problem 1.

12. For any *unit vectors* $\mathbf{v}$ and $\mathbf{w}$, find the dot products of (a) v and $-\mathbf{v}$, and (b) $\mathbf{v} + \mathbf{w}$ and **v**-**w**.

13. Describe every vector $\mathbf{w} = (w_1, w_2)$ that is perpendicular to $\mathbf{v} = (2, -1)$.

14. All vectors perpendicular to $\mathbf{v} = (1, 1, 1)$ lie on a __________ in 3 dimensions. The vectors perpendicular to both (1,1,1) and (1,2,3) lie on a __________.

15. Find the angle θ between $\mathbf{v} = (3, 1)$ and $\mathbf{w} = (-1, -2)$.

16. Find the angle θ between $\mathbf{v} = (2, 2, -1)$ and $\mathbf{w} = (2, -1, 2)$.

17. Find the angle θ between $\mathbf{v} = (1, \sqrt{3})$ and $\mathbf{w} = (-1, \sqrt{3})$.

18. With $\mathbf{v} = (1, 1)$ and $\mathbf{w} = (1, 5)$ choose a number c so that $\mathbf{w} - c\mathbf{v}$ is perpendicular to v. Then find the formula for c starting from *any* nonzero $\mathbf{v}$ and $\mathbf{w}$.

19. How long is the vector $\mathbf{v} = (1, 1, \ldots, 1)$ in 9 dimensions? Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$ and a unit vector $\mathbf{w}$ that is perpendicular to $\mathbf{v}$.

20. **Challenge**: In the xy plane, when could four vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{v}_4$ not be the four sides of a quadrilateral?

4 Matrices

4.1 Overview

We can form a **matrix** by horizontally concatenating n vectors of dimension m. This gives us a matrix $A \in \mathbb{R}^m \times \mathbb{R}^n$. Another way of saying it is that a matrix with m rows and n columns is an m by n matrix. Typically, we define matrices with capital letters. Matrix

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

is a 3 by 2 matrix. The (i,j)-th entry of a matrix is the entry in row i and column j. For example, $A_{(1,2)}$ in the example above is 2, whereas $A_{(2,1)}$ is 3.

Let's go over some simple categorizations of matrices:

- A **square matrix** has an equal number of columns and rows, i.e., $m = n$.
- A **diagonal matrix** has all entries off the main diagonal all equal to zero. The main diagonal consists of all (i, j)-th entries where $i = j$. The following is an example of a 3 by 3 diagonal matrix:

  $$A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

  for some $a, b, c \in \mathbb{R}$.
- The **identity matrix** is an n by n diagonal matrix with every entry on the main diagonal equal to 1. We denote this as $\mathbb{I}_n$ or just $\mathbb{I}$.
- A **triangular matrix** has either all entries below the main diagonal or above the main diagonal equal to zero. If all entries below the main diagonal are zero, then it is an upper triangular matrix. One example of an upper triangular matrix is

  $$A = \begin{bmatrix} 2 & 1 & -3 \\ 0 & 4 & 7 \\ 0 & 0 & 5 \end{bmatrix}$$

  A lower triangular matrix has entries above the main diagonal equal to zero.
- A **symmetric matrix** has every (i, j)-th entry equal to every (j, i)-th entry. Below is one example of a symmetric matrix:

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

4.2 Multiplying a Matrix and a Vector

Now, we can define matrix-vector multiplication. An $m \times n$ matrix can be multiplied by an n-dimensional vector to get an m-dimensional vector output. It's important to note that the vector dimension *must* be equal to the number of columns in the matrix. Otherwise, matrix-vector multiplication is not well-defined. The number of rows gives us the dimension of the resulting vector. There are two ways to conceptualize the mechanics of matrix-vector multiplication:

1. **Row / Dot Product Approach**: In this picture, multiplying $m \times n$ matrix A by n-dimensional vector $\mathbf{x}$ gives us

   $$A\mathbf{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{1\cdot}^T \\ \mathbf{a}_{2\cdot}^T \\ \cdots \\ \mathbf{a}_{m\cdot}^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{1\cdot} \cdot \mathbf{x} \\ \mathbf{a}_{2\cdot} \cdot \mathbf{x} \\ \cdots \\ \mathbf{a}_{m\cdot} \cdot \mathbf{x} \end{bmatrix}$$

where $\mathbf{a}_{1\cdot}^T$ is the row vector of the entries in the first row of matrix A. In words, this approach forms the result of $A\mathbf{x}$ by taking the dot product of each row of A with $\mathbf{x}$.

2. **Column / Linear Combination Approach**: In this picture, multiplying $m \times n$ matrix A by n-dimensional vector $\mathbf{x}$ gives us a linear combination of the columns in A, where the entries of $\mathbf{x}$ are the multiplying scalars of each column. That is,

$$A\mathbf{x} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \cdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \cdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \cdots \\ a_{mn} \end{bmatrix}$$

These two approaches are mechanically equivalent. I personally find it easier to do computation with the dot product approach, but it is often easier to get geometric intuition with the linear combination approach.

Another way to think about a matrix is as a **linear transformation**. A matrix can take any arbitrary vector $\mathbf{x}$ and, through matrix-vector multiplication, output a new vector. Let's call this new vector $\mathbf{b}$. So the matrix A in the equation $A\mathbf{x} = \mathbf{b}$ can be viewed as a linear transformation that takes vector $\mathbf{x}$ and outputs $\mathbf{b}$. The columns of a matrix define the result of applying the linear transformation to each standard basis vector. Grant Sanderson at the YouTube channel 3Blue1Brown has an incredibly helpful video on this concept, with dynamic visualizations of matrices as linear transformations.[4] For example, we can call the following 2-D matrices *rotation matrices*:

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

The matrix A multiplied by $\mathbf{x}$ gives a 90° counter-clockwise rotation of $\mathbf{x}$, and the matrix B multiplied by $\mathbf{x}$ gives a 180° counter-clockwise rotation of $\mathbf{x}$.

Lastly, we can think of matrices as describing a **system of linear equations**. Suppose we want to solve the following simple system of 2 linear equations with 2 unknown variables:

$$\begin{aligned} 3x_1 - 4x_2 &= 2 \\ 9x_1 + x_2 &= 1 \end{aligned}$$

Hopefully, this should not be too difficult for you to solve algebraically. But what if we had a very large number of equations and unknown variables? This complicates the standard algebraic approach, and a matrix-representation approach becomes preferable. In the example above, we can equivalently write

$$\underset{A}{\begin{bmatrix} 3 & -4 \\ 9 & 1 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}} \underset{=}{} \underset{\mathbf{b}}{\begin{bmatrix} 2 \\ 1 \end{bmatrix}}$$

So solving for x_1 and x_2 is equivalent to finding vector $\mathbf{x}$ that solves $A\mathbf{x} = \mathbf{b}$. But how do we know there even exists a solution to $A\mathbf{x} = \mathbf{b}$? This can be answered by thinking about the properties of matrix A.

4.3 Column Space and Rank of Matrices

We begin this section by defining new terms related to our earlier discussion on linear combinations of vectors. We say that a set of vectors are **dependent** if one of the vectors in that set is a linear combination of some of the other vectors. For example, consider

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 6 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 0 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 5 \\ 9 \\ 6 \end{bmatrix}$$

[4] I highly recommend his whole series of videos on The Essence of Linear Algebra as a useful aid for ideas covered in this Math Camp and your CAPP Linear Algebra course.

Note that we have $\mathbf{x} + 2\mathbf{y} = \mathbf{z}$, so this set of vectors are dependent. A set of vectors is **independent** if none of the columns are linear combinations of the others. For example, consider

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 4 \\ 5 \end{bmatrix}, \quad \mathbf{z} = \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}$$

There is no way to linearly combine two of these vectors to get the third.

The notion of dependence and independence is closely related to our earlier idea of the dimensionality of the space of all linear combinations of vectors. In the first example above, because $\mathbf{z}$ lies on the plane spanned by $\mathbf{x}$ and $\mathbf{z}$, the set of all linear combinations is just a 2-D plane. Whereas in the second example, because all three vectors are independent (in a sense, they point in different "directions"), then the space of all linear combinations is all of $\mathbb{R}^3$. See Figure 7 for geometric intuition about vector independence.

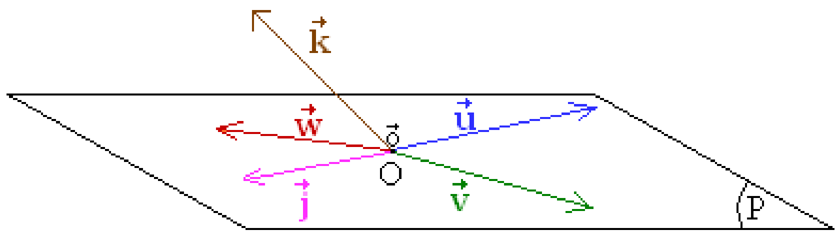

Figure 7: Vector dependence and independence, courtesy of Wikipedia. Vectors $\mathbf{u}$ and $\mathbf{v}$ are independent, defining plane P. Vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are dependent because they are all in the same plane. Vectors $\mathbf{u}$ and $\mathbf{j}$ are on the same line, so they are dependent.

Since matrix columns are vectors, we can apply the idea of independence or dependence to them. The columns of a matrix are independent if no column can be described as a linear combination of the other columns. For matrices, the space of all linear combinations of columns is known as its **column space**. Using the column / linear combination approach to understanding matrix-vector multiplication, we can equivalently say that the column space of A is the set of all vectors $A\mathbf{x}$. A vector $\mathbf{b}$ is in the column space of A if there exists some $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$. The column space of matrix A is sometimes denoted $\mathbf{C}(A)$.

How do we characterize the dimension of a matrix's column space? This can be done by counting the number of independent columns! Consider the case of a matrix A with m-dimensional columns. There are 4 possibilities for its column space:

1. With **3 independent columns**, the column space $\mathbf{C}(A)$ is all of $\mathbb{R}^3$.
2. With **2 independent columns**, the column space is a plane in $\mathbb{R}^3$ going through the origin.
3. With **1 independent column**, the column space is a line in $\mathbb{R}^3$ going through the origin.
4. If A is a matrix of all zeros, then the column space is just the origin (0,0,0). Every vector, when multiplied by A, gets mapped to the origin.

The dimension of the column space (equivalent to the number of independent columns) is r, we say that this matrix's **rank** is r. We denote this $\text{rank}(A) = r$. [5] If a matrix has rank r, then the first r columns are a **basis** for the column space. Loosely speaking, a basis for a given space is a set of vectors whose linear combinations can reach any point on that space.

[5] A matrix's **row space** can be defined analogously to a column space. Remarkably, an important result you will see later is that the number of independent columns is equal to the number of independent rows. So we can identify a matrix's rank by looking at either its row space or column space.

4.4 Multiplying a Matrix and a Matrix

Our last topic will involve multiplying two matrices, A and B. If the row-length of A is equal to the column-length of B, then we can multiply them together to get a new matrix AB. If not, then matrix multiplication is not defined. Specifically, suppose that A is a $m \times n$ matrix and B is a $n \times p$ matrix. Then AB will be a $m \times p$ matrix.

$$\underset{(m\times n)}{A} \text{ times } \underset{(n\times p)}{B} = \underset{(m\times p)}{AB}$$

We call two matrices **conformable** if they can be multiplied, that is, if the row-length of A is equal to the column-length of B.

The mechanics for multiplying two matrices is really just an extension of matrix-vector multiplication. We can think of a vector as a matrix with just one row. So we can extend our two perspectives on matrix-vector multiplication accordingly:

1. **Dot product view**: The (i, j)-th entry of product AB is equal to the dot product of the i-th row of A and the j-th column of B.
2. **Linear combination view**: The i-th column of product AB is a linear combination of all columns of A, with scalar weights described by the i-th column of B.

Let's do a simple matrix multiplication as an example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} (1)(5)+(2)(7) & (1)(6)+(2)(8) \\ (3)(5)+(4)(7) & (3)(6)+(4)(8) \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Some standard facts about matrix multiplication:

- Recall that we defined the **identity matrix** $\mathbb{I}$ as a matrix with 1s along the main diagonal and 0s elsewhere. For any matrix A, we have $A\mathbb{I} = A$ and $\mathbb{I}A = A$, if conformable. In words, the identity matrix times any other matrix returns the same matrix.
- Matrix multiplication is **not commutative**. This means that, generally speaking,

  $$AB \neq BA$$

  So unlike scalar multiplication or matrix and vector addition, order *very much matters*. Depending on the shapes of A and B, it may be the case that AB is well-defined, but BA is not.
- Even though it is not commutative, matrix multiplication is **associative**. That is, for matrices A, B, C

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

  So we can change parentheses in the context of matrix multiplication without changing the final result.

Recall that we conceptualized a matrix as a linear transformation. Matrix multiplication of A and B to get new matrix AB can be thought of as the **composition** of two linear transformations. So, for example, if A is the matrix that rotates all vectors counter-clockwise by 90°, then A times A (i.e., A^2) should be equal to the matrix that rotates all vectors by 180°. We can confirm this:

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

which we know to be the 180° rotation matrix.

I will add one more way to consider matrix multiplication, through **outer products**. The outer product of two vectors $\mathbf{u}$ and $\mathbf{v}$ is a matrix defined by $\mathbf{u}\mathbf{v}^T$, where $\mathbf{v}^T$ is a row vector. We can write matrix multiplication as the sum of outer products:

$$AB = \underset{\text{columns of A}}{\begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}} \underset{\text{rows of B}}{\begin{bmatrix} \mathbf{b}_1^T \\ \dots \\ \mathbf{b}_n^T \end{bmatrix}} = \mathbf{a}_1\mathbf{b}_1^T + \mathbf{a}_2\mathbf{b}_2^T + \ldots + \mathbf{a}_n\mathbf{b}_n^T$$

4.5 Matrix Practice Questions

1. Multiply $A\mathbf{x}$, $B\mathbf{y}$, and $C\mathbf{z}$:

$$A\mathbf{x} = \begin{bmatrix} 2 & 1 & 2 \\ 4 & 2 & 4 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$
$$B\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \\ 10 \end{bmatrix}$$
$$C\mathbf{z} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

2. In the previous question, how many independent columns does A have? How many independent columns does B have? How many independent columns are in $A + B$?

3. Describe the column space of the following matrices: a point, a line, a plane, or all of $\mathbb{R}^3$:

$$A_1 = \begin{bmatrix} 2 & 2 \\ 1 & 1 \\ 5 & 6 \end{bmatrix}$$
$$A_2 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$
$$A_3 = \begin{bmatrix} 1 & 5 \\ 2 & 10 \\ 1 & 5 \end{bmatrix}$$
$$A_4 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

4. Complete A and B so that they are rank one matrices. What are the column spaces of A and B? What are the row spaces of A and B?

$$A = \begin{bmatrix} 3 & \\ 5 & 15 \end{bmatrix}$$
$$B = \begin{bmatrix} 1 & 2 & -5 \\ 4 & & \end{bmatrix}$$

5. Set up the system of equations

$$x_1 + 3x_2 = 4$$
$$2x_1 + 4x_2 = 6$$

in the form $A\mathbf{x} = \mathbf{b}$. Solve for x_1 and x_2, and relate this solution to the column space of A.

6. Which numbers q would leave A with two independent columns?

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

7. Suppose A times $\mathbf{x}$ equals $\mathbf{b}$. If you add $\mathbf{b}$ as an extra column of A, explain why the rank r (number of independent columns) stays the same.

8. Multiply the following matrices:

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

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

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

9. Test the truth of the associativity of matrix multiplication, i.e., $(AB)C = A(BC)$, for

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