Class 5 Notes: Introduction to Matrix Algebra

Myles D. Garvey, Ph.D

Summer, 2019

Contents

1 Introduction

Matrix Algebra allows us to express very complicated sets of equations in a simple and easy to use notation. When you are able to master the fundamentals of matrix algebra, all of the mathematics that you will encounter in your future courses, as well as material covered in multivariate calculus, as well as in statistics and machine learning, becomes easier to understand and to manipulate. In this lecture, we will review some of the basics of matrix algebra. I cannot promise that you will become an expert in a short time, this requires a lot of practice! But I can assure you that after this lecture, you should be able to identify a lot more of the notation that is used in textbooks and academic journal articles a like. So let's get going!

2 Fundamentals of Matrix Algebra

2.1 The Concept of a Matrix

Suppose we have gathered data on three variables, age, income, and spending levels for 5 individuals. Suppose this data is gathered and organized in the table below:

Person	Age	Income	Spending
1	23	34352	14083.02
2	62	74306	43939.13
3	29	74175	67823.23
4	29	24064	9071.93
5	53	57102	58731.75

Further suppose that we would like to find an equation to explain a person's spending. We may hypothesize that a person's age and income are *linearly related* with their spending levels. If this is the case, then we seek an equation like the following:
$Spending_i = \beta_0 + \beta_1 Age_i + \beta_2 Income_i + \epsilon_i$, where $i \in \{1, 2, 3, 4, 5\}$, $\beta_0, \beta_1, \beta_2$ are called *coefficients*, and they do not change with the values of age, income, or spending, and ϵ_i represents everything else about person i that impact's their spending levels. Our goal is to use the data to find the values of $\beta_0, \beta_1, \beta_2$ in such a way that $\sum_{i=1}^{5} \epsilon_i^2 = \epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \epsilon_4^2 + \epsilon_5^2$ is at it's lowest possible value. Since we have data on the variables $Spending_i, Income_i, Ages_i$, and since we have 5 data points, we can plug in our data into our general equation, to give us 5 equations (one for each data observation):

$$\begin{aligned}
14083.02 &= \beta_0 + \beta_1(23) + \beta_2(34352) + \epsilon_1 \\
43939.13 &= \beta_0 + \beta_1(62) + \beta_2(74306) + \epsilon_2 \\
67823.23 &= \beta_0 + \beta_1(29) + \beta_2(74175) + \epsilon_3 \\
9071.92 &= \beta_0 + \beta_1(29) + \beta_2(24064) + \epsilon_4 \\
58731.75 &= \beta_0 + \beta_1(53) + \beta_2(57102) + \epsilon_5
\end{aligned}$$

The above set of equations is referred to as a *system of linear equations*. We notice that each equation has the same variables in them ($\beta_0, \beta_1, \beta_2$), and soon in our course we will discuss how to solve these. However, our focus right now is not to solve these, but rather to represent the euqations in a more compact way. When we have systems of linear equations, things are, well, messy to say the least. We want to design a type of mathematics that will help us summarize in a more compact way our data and our operations on the data. We do so by leveraging *matrix theory*. Let us observe the right-hand sides equations. We notice that all equations share the same coefficients, but different values of the variables. Again, this is messy. But, we can leverage the idea of a matrix to summarize our data in more compact way. Let's see how. If we were to consider only the numbers that are not equal to the coefficients, then we obtain the following numbers:

$$\begin{gathered}1, (23), (34352)\\1, (62), (74306)\\1, (29), (74175)\\1, (29), (24064)\\1, (53), (57102)\end{gathered}$$

This organization of the information is certainly more compact, but not compact enough, and still seems arbitrary. So, we introduce the concept of a *matrix* to finish our organization process. A matrix is a table of rows and columns. Each entry in a matrix is identified by it's row and column. Typically, we write $(a_{i,j})$ or $[a_{i,j}]$ to indicate the element that is in the ith row and the jth column. Once the table is specified, we typically assign a capital (or bold faced) latter to the table, which we heretofore will refer to as a *matrix*. Let us see what our table looks like in matrix format:

$$A = \begin{bmatrix} 1 & 23 & 34352 \\ 1 & 62 & 74306 \\ 1 & 29 & 74175 \\ 1 & 29 & 24064 \\ 1 & 53 & 57102 \end{bmatrix}$$

It may not look like much, but believe it or not, this matrix hold a lot of information. Not to mention, we have defined a new type of mathematical object. As with all types of objects that we define in math, not only can we define objects, but conduct operations on them. Our specific matrix, A, is defined by 5 rows and 3 columns, and so we say that we have a 5x3 matrix. In general, an nxm matrix is a matrix that has n rows and m columns.

2.2 Types of Matrices

We also have special types of matrices that we often work with. The most fundamental of these is a special matrix that is known as a *vector*, and we have two types of these. A *row vector* is a 1xm matrix, and we say that the vector has length m. A *column vector* is an nx1 matrix, and we say that the vector has length n.

For example, we have a row vector of length 3:

$$\begin{bmatrix} 1 & 23 & 34352 \end{bmatrix}$$

and we have a column vector of length 5:

$$\begin{bmatrix} 23 \\ 62 \\ 29 \\ 29 \\ 53 \end{bmatrix}$$

If we are given a matrix A, then the jth column of A is commonly represented by $\mathbf{a}_j$, while the ith row of A is commonly represented by $\mathbf{a}'_i$. Here, the "prime" on the vector notation represents the matrix operation of a *transpose*. The transpose of a column vector with elements ordered from top to bottom is the same as a row vector with the elements ordered from left to right. Therefore, if we take the transpose of a column, we obtain a row, and if we take the transpose of a row, we obtain a column. We will return back to this idea in a moment.

For now, we would like to cover a few special types of matrices. A *square matrix* is a matrix whose number of rows equals it's number of columns. So an nxn matrix is a square matrix. We have a few special types of square matrices. A *symmetric matrix* is a matrix where $\forall i, j \in \{1, 2, \ldots, n\}, a_{ij} = a_{ji}$. An example would be

$$A = \begin{bmatrix} 4 & 8 & 2 \\ 8 & 10 & 98 \\ 2 & 98 & 2 \end{bmatrix}$$

A diagonal matrix is a matrix where $\forall i, j \in \{1, 2, \ldots, n\} s.t. i \neq j, a_{i,j} = 0$. An example would be:

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

A *scalar matrix* is a diagonal matrix where all the diagonal elements are the same, for example:

$$A = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

An identity matrix is a scalar matrix such that all the diagonal elements are equal to 1:

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

A triangular matrix is a matrix where either $\forall i, j \in \{1, 2, \ldots, n\} s.t. i < j, a_{i,j} = 0$ or $\forall i, j \in \{1, 2, \ldots, n\} s.t. j < i, a_{i,j} = 0$. Here are examples of lower triangular and upper triangular matrices, respectively: Lower Triangular:

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 31 & 6 & 0 & 0 \\ 13 & 32 & 5 & 0 \\ 4 & 6 & 4 & 2 \end{bmatrix}$$

$$A = \begin{bmatrix} 10 & 6 & 4 & 0 \\ 0 & 46 & 0 & 3 \\ 0 & 0 & 31 & 0 \\ 0 & 0 & 0 & 32 \end{bmatrix}$$

2.3 Inner Products

Suppose we have two vectors, $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix}$. The idea here is the we would like to combine the two vectors together to obtain something else. Multiplication of vectors works very differently than adding them (we will cover this below). There are two ways to "multiply" the vectors together. The first way is to combine the two vectors together to form a single number, and this type of product is referred to as the *inner product* of the two vectors. We will hold off on discussion on the second way.

First, whenever we multiply vectors, we can **never** multiply two column or two row vectors together. instead, we need to take the *transpose* of one of the vectors, and then multiply. For the inner product of the vectors **a** and **b**, we can denote this in one of two ways. Some academic journals and textbooks use a single "dot" in the middle, such as $\mathbf{a'} \cdot \mathbf{b}$, while others use this notation: $< \mathbf{a'}, \mathbf{b} >$. Some simply just place one vector next to the other **a'b** (notice in all of them, we use the " ′ " symbol, this is due to the fact that **a** is a column vector, and we need to convert it to a row vector by *transposing* it.). Regardless of the notation, they all mean the same thing: take the inner product of the two vectors, which is defined by matching up like elements in the two vectors, multiplying the like elements, and then adding all the results:

Let $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix}$. Then taking the transpose of **a** gives: $\mathbf{a'} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$ thus:

$$\begin{aligned} \mathbf{a'b} =< \mathbf{a'}, \mathbf{b} >= \mathbf{a'} \cdot \mathbf{b} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix} \\ = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \end{aligned}$$

For example, suppose we have $\mathbf{a} = \begin{bmatrix} 3 \\ 5 \\ 2 \\ 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 8 \\ 2 \\ 5 \\ 9 \end{bmatrix}$. In order to find $\mathbf{a'b}$, we first find $\mathbf{a'}$: $\mathbf{a'} = \begin{bmatrix} 3 & 5 & 2 & 1 \end{bmatrix}$, then match them up and use our equation to solve:

$$\begin{aligned} \mathbf{a'b} &= \begin{bmatrix} 3 & 5 & 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ 2 \\ 5 \\ 9 \end{bmatrix} \\ &= (3)(8) + (5)(2) + (2)(5) + (1)(9) \\ &= 24 + 10 + 10 + 9 \\ &= 53 \end{aligned}$$

Note that $\mathbf{a'b} \neq \mathbf{ba'}$. The latter term produces a matrix rather than a number, so we **must** be very careful with multiplying. However, it is true that $\mathbf{a'b} = \mathbf{b'a}$. Notice the careful placement of the transposes. Everything in Matrix Algebra is about careful placement of the terms (not everything is commutative like in regular algebra). We can show this is true very easily:

$$\begin{aligned} \mathbf{a'b} &= \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{bmatrix} \\ &= a_1b_1 + a_2b_2 + \cdots + a_nb_n \\ &= b_1a_1 + b_2a_2 + \cdots + b_na_n \\ &= \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix} \\ &= \mathbf{b'a} \end{aligned}$$

Note that we cannot find the inner product of all vectors. We need to always ensure that the length of our vectors is the same. For example, we cannot find the inner product of $\mathbf{a} = \begin{bmatrix} 3 \\ 100 \\ 10 \\ 52 \\ 313 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 100 \\ 13 \\ 3 \\ 7 \end{bmatrix}$, since the length of $\mathbf{a}$ is 5 and the length of $\mathbf{b}$ is 4 (in this situation, we **can** find something called an *outer product*, but we will avoid discussion on these for now).

Notice in our motivating example from earlier that we can shorten down our equations using inner products. Take the first equation in the system of equations. We had $14083.02 = \beta_0 + \beta_1(23) + \beta_2(34352) + \epsilon_1$. This can be shortened by defining two vectors: $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$ and $\boldsymbol{a_1} = \begin{bmatrix} 1 \\ 23 \\ 34352 \end{bmatrix}$. We can then express our equation from earlier in a much shorten notation: $14082.02 = \boldsymbol{\beta' a_1} + \epsilon_1$. We could have defined this differently as well. If we constructed a vector $\boldsymbol{\theta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \epsilon_1 \end{bmatrix}$ and $\boldsymbol{a_1} = \begin{bmatrix} 1 \\ 23 \\ 34352 \\ 1 \end{bmatrix}$, then in this instance, we could write our expression even more simplistically: $14082.02 = \boldsymbol{\theta' a_1}$. The power of this notation is the ability to take very long and complex equations and make the notation "cleaner" to work with.

3 Matrix Properties and Operations

One we have defined a few different matrices, we can conduct operations on them, just like we conduct operations on numbers. To most fundamental fact about matrices is their equality. We say that two matrices A and B, both of size nxm, are equal if and only if $\forall i \in \{1, \dots, n\}, j \in \{1, \dots, n\} a_{i,j} = b_{i,j}$. In such an instance, we write $A = B$. As for notation, we commonly use capital letters that are bold to represent a matrix. Therefore, if you see equations with bold font, it most likely is a matrix or a vector. So we really should say $\boldsymbol{A} = \boldsymbol{B}$.

3.1 Transposes

The transpose of an nxm matrix $\boldsymbol{A}$ is an mxn matrix where the rows are the columns of $\boldsymbol{A}$. We denote the transpose as A'. In some textbooks and journals, you may see the notation $\boldsymbol{A}^T$ as well. Both are equivalent notations. For example, if we have the matrix $\boldsymbol{A}$ defined below, it's transpose $\boldsymbol{A'}$ is also below:

$$\boldsymbol{A} = \begin{bmatrix} 6 & 3 & 7 & 3 \\ 13 & 36 & 6 & 3 \\ 26 & 2 & 62 & 7 \\ 2 & 7 & 0 & 2 \\ 74 & 23 & 43 & 6 \end{bmatrix}, \ \boldsymbol{A'} = \begin{bmatrix} 6 & 13 & 26 & 2 & 74 \\ 3 & 36 & 2 & 7 & 23 \\ 7 & 6 & 62 & 0 & 43 \\ 3 & 3 & 7 & 2 & 6 \end{bmatrix}$$

3.2 Matrix Addition

When given two matrices, we can add them together just like how we add two numbers together. If we have matrix $\boldsymbol{A} = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,n} \\ a_{1,0} & a_{1,1} & \dots & a_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,0} & a_{n,1} & \dots & a_{n,n} \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} b_{0,0} & b_{0,1} & \dots & b_{0,n} \\ b_{1,0} & b_{1,1} & \dots & b_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n,0} & b_{n,1} & \dots & b_{n,n} \end{bmatrix}$ Then the sum of the matrices would be:

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} a_{0,0} + b_{0,0} & a_{0,1} + b_{0,1} & \dots & a_{0,n} + b_{0,n} \\ a_{1,0} + b_{1,0} & a_{1,1} + b_{1,1} & \dots & a_{1,n} + b_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,0} + b_{n,0} & a_{n,1} + b_{n,1} & \dots & a_{n,n} + b_{n,n} \end{bmatrix}$$

For example, if $\boldsymbol{A} = \begin{bmatrix} 4 & 2 & 1 \\ 6 & 3 & 6 \\ 3 & 2 & 6 \end{bmatrix}$ and if $\boldsymbol{B} = \begin{bmatrix} 3 & 7 & 33 \\ 23 & 4 & 7 \\ 6 & 2 & 7 \end{bmatrix}$, then

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} 4 & 2 & 1 \\ 6 & 3 & 6 \\ 3 & 2 & 6 \end{bmatrix} + \begin{bmatrix} 3 & 7 & 33 \\ 23 & 4 & 7 \\ 6 & 2 & 7 \end{bmatrix} = \begin{bmatrix} (4+3) & (2+7) & (1+33) \\ (6+23) & (3+4) & (6+7) \\ (3+6) & (2+2) & (6+7) \end{bmatrix} = \begin{bmatrix} 7 & 9 & 34 \\ 29 & 7 & 13 \\ 9 & 4 & 13 \end{bmatrix}$$

Note that we can only add two matrices if and only if they are the same size. That is, both $\boldsymbol{A}$ and $\boldsymbol{B}$ both need to be an *nxm* matrix. If they are different sizes, then we **cannot** add them together.

3.3 Scalar Multiplication

In matrix algebra, there are two types of multiplication (techinically, three, if you count the *outer product* as a different type), scalar multiplication and matrix multiplication. If we are given a matrix $\boldsymbol{A} = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,n} \\ a_{1,0} & a_{1,1} & \dots & a_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,0} & a_{n,1} & \dots & a_{n,n} \end{bmatrix}$, and a number $k \in \mathbb{R}$, then we can multiply the number times the matrix by multiplying the number with every element in the matrix by k (for those of you with a background in Abstract Algebra, you should be able to identify this as a *group action*). That is:

$$k\boldsymbol{A} = k \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,n} \\ a_{1,0} & a_{1,1} & \dots & a_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,0} & a_{n,1} & \dots & a_{n,n} \end{bmatrix} = \begin{bmatrix} ka_{0,0} & ka_{0,1} & \dots & ka_{0,n} \\ ka_{1,0} & ka_{1,1} & \dots & ka_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ ka_{n,0} & ka_{n,1} & \dots & ka_{n,n} \end{bmatrix}$$

Notice that the k is not in bold. When working with vectors and matrices, when a symbol is not in bold, it typically is understood to represent a number, or a variable that a number will be assigned to. As an example of scalar multiplication, suppose we have $\boldsymbol{A} = \begin{bmatrix} 4 & 2 & 5 \\ 3 & 65 & 12 \\ 24 & 5 & 6 \end{bmatrix}$ and $k = 2 \in \mathbb{R}$. Then

$$(2)\begin{bmatrix}4 & 2 & 5\\ 3 & 65 & 12\\ 24 & 5 & 6\end{bmatrix} = \begin{bmatrix}(2)4 & (2)2 & (2)5\\ (2)3 & (2)65 & (2)12\\ (2)24 & (2)5 & (2)6\end{bmatrix} = \begin{bmatrix}8 & 4 & 10\\ 6 & 130 & 24\\ 48 & 10 & 12\end{bmatrix}$$

With the concept of addition and scalar multiplication, we can subtract two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. For example, suppose we had the same two matrices as before, $\boldsymbol{A} = \begin{bmatrix}4 & 2 & 1\\ 6 & 3 & 6\\ 3 & 2 & 6\end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix}3 & 7 & 33\\ 23 & 4 & 7\\ 6 & 2 & 7\end{bmatrix}$, then

$$\begin{aligned}
\boldsymbol{A} - \boldsymbol{B} &= \boldsymbol{A} + (-1)\boldsymbol{B}\\
&= \begin{bmatrix}4 & 2 & 1\\ 6 & 3 & 6\\ 3 & 2 & 6\end{bmatrix} + \begin{bmatrix}-3 & -7 & -33\\ -23 & -4 & -7\\ -6 & -2 & -7\end{bmatrix}\\
&= \begin{bmatrix}(4-3) & (2-7) & (1-33)\\ (6-23) & (3-4) & (6-7)\\ (3-6) & (2-2) & (6-7)\end{bmatrix}\\
&= \begin{bmatrix}1 & -5 & -32\\ -17 & -1 & -1\\ -3 & 0 & -1\end{bmatrix}
\end{aligned}$$

3.4 Matrix Multiplication

We first discussed the idea of *scalar multiplication*. But what if we wanted to multiply two matrices together? That is, instead of multiplying a single number times a matrix, can we multiply a matrix times a matrix. The answer to this question is: under certain circumstances, yes. However, matrix multiplication is not what many may think it is. Naturally, we may think that multiplying two matrices is the same as adding them. However, this is not correct. First, we need to make sure that the two matrices we multiply are *conformable*. If $\boldsymbol{A}$ is an n_1xm_1 matrix and $\boldsymbol{B}$ is an n_2xm_2 matrix, then we say that are *conformable* if $m_1 = n_2$. That is, the matrix $\boldsymbol{A}$ has the same number of columns as the number of rows as matrix $\boldsymbol{B}$. For example, if $\boldsymbol{A}$ is a 2x5 matrix and $\boldsymbol{B}$ is a 5x4 matrix, then we can multiply these two together and find $\boldsymbol{AB}$. However, we **cannot** multiply $\boldsymbol{BA}$, since $\boldsymbol{B}$ has 4 columns and $\boldsymbol{A}$ has 2 rows, which violates our definition of the matrices being *conformable*.

This raises a very important point when it comes to matrix multiplication: **order matters**. In general, matrix multiplication is not *commutative* like addition or scalar multiplication. So we must take the utmost care when multiplying matrices. A second point to raise is that matrix multiplication is found using a very specific procedure. If we multiply a matrix $\boldsymbol{A}$ that is n_1xm_1 by $\boldsymbol{B}$ which is m_1xm_2, then the resulting matrix is an n_1xm_2 matrix. In our example before, we multiplied a 2x5 with a 5 x 4, which will lead us to a 2x4 matrix. In general, we find the multiplication of two conformable matrices by conducting a procedure. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable, then the $c_{i,j}$ element in $\boldsymbol{AB}$ is defined as $c_{i,j} = \boldsymbol{a_i'b_j}$. That is, to find the element in the multiplied matrix in the ith row and the j th column, we take the inner product of the ith row in matrix $\boldsymbol{A}$ with the jth column in the matrix $\boldsymbol{B}$. Since we assumed the two matrices are conformable, the ith row of matrix $\boldsymbol{A}$ will be a vector of length m_1 and the jth column in matrix $\boldsymbol{B}$ will be a vector of length m_1.

Let us illustrate this process with a matrix $\boldsymbol{A} = \begin{bmatrix} 4 & 6 & 2 \\ 3 & 1 & 4 \\ 9 & 4 & 1 \\ 4 & 6 & 2 \end{bmatrix}$ and $\boldsymbol{B} = \begin{bmatrix} 1 & 5 & 3 & 2 & 2 \\ 1 & 2 & 4 & 2 & 1 \\ 10 & 2 & 2 & 4 & 4 \end{bmatrix}$ Notice that matrix $\boldsymbol{A}$ has 4 rows and 3 columns, while matrix $\boldsymbol{B}$ has 3 rows and 5 columns. These matrices are conformable for $\boldsymbol{AB}$, since $\boldsymbol{A}$ has 3 columns and $\boldsymbol{B}$ has 3 rows. Notice how $\boldsymbol{BA}$ is **not** conformable, since $\boldsymbol{B}$ has 5 columns and $\boldsymbol{A}$ has 4 rows. An easy way to remember if two matrices are conformable is to write the size of the first matrix followed by the size of the second matrix. If the two middle numbers are the same, they are conformable, if they are different, then we cannot multiply them. For example, for $\boldsymbol{AB}$, we have (4x3)(3x5), the two middle numbers are the same, and so they are conformable. For $\boldsymbol{BA}$, we have (3x5)(4x3), and we see that the two middle numbers are **not** conformable, and hence cannot be multiplied!

Lets see how the multiplication works. We know that $\boldsymbol{C} = \boldsymbol{AB}$ exists since this multiplication is conformable. We also know that since we are doing (4x3)(3x5), the resulting matrix will be (4x5) (this is always true, look at the sizes, the resulting size will be the first number for the number of rows and the last number for the number of columns in the statement (4x3)(3x5)). So we have 20 different inner products to compute. Lets go element by element to compute each one. First, lets compute the element $c_{0,0}$, which is found by taking the inner product of the first row of matrix $\boldsymbol{A}$ and the first column of matrix $\boldsymbol{B}$, that is:

$$\begin{aligned} c_{0,0} &= \boldsymbol{a}_0' \boldsymbol{b}_0 \\ &= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 10 \end{bmatrix} \\ &= (4)(1) + (6)(1) + (2)(10) \\ &= 4 + 6 + 20 \\ &= 30 \end{aligned}$$

So we would have

$$\boldsymbol{AB} = \begin{bmatrix} 30 & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \end{bmatrix}$$

Next, we move to the next column in the matrix $\boldsymbol{B}$ (alternatively, you could move row to row instead of column to column, the choice is yours!). We have:

$$\begin{aligned}
c_{0,1} &= \boldsymbol{a}_0'\boldsymbol{b}_1 \\
&= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 2 \\ 2 \end{bmatrix} \\
&= (4)(5) + (6)(2) + (2)(2) \\
&= 20 + 12 + 4 \\
&= 36
\end{aligned}$$

So we would have

$$\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}$$

The rest of the calculations are below (you can skip ahead if you get the point):

$$\begin{aligned}
c_{0,2} &= \boldsymbol{a}_0'\boldsymbol{b}_2 \\
&= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix} \\
&= (4)(3) + (6)(4) + (2)(12) \\
&= 12 + 24 + 24 \\
&= 60
\end{aligned}$$

$$\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}$$

$$\begin{aligned}
c_{0,3} &= \boldsymbol{a}_0'\boldsymbol{b}_3 \\
&= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \\
&= (4)(2) + (6)(2) + (2)(4) \\
&= 8 + 12 + 8 \\
&= 28
\end{aligned}$$

$$\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}$$

$$
\begin{aligned}
c_{0,4} &= \boldsymbol{a}_0' \boldsymbol{b}_4 \\
&= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \\
&= (4)(2) + (6)(1) + (2)(4) \\
&= 8 + 6 + 8 \\
&= 32
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{1,0} &= \boldsymbol{a}_1' \boldsymbol{b}_0 \\
&= \begin{bmatrix} 3 & 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 10 \end{bmatrix} \\
&= (3)(1) + (1)(1) + (4)(10) \\
&= 3 + 1 + 40 \\
&= 44
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{1,1} &= \boldsymbol{a}_1' \boldsymbol{b}_1 \\
&= \begin{bmatrix} 3 & 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 2 \\ 2 \end{bmatrix} \\
&= (3)(5) + (1)(2) + (4)(2) \\
&= 10 + 2 + 8 \\
&= 20
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{1,2} &= \boldsymbol{a}_1' \boldsymbol{b}_2 \\
&= \begin{bmatrix} 3 & 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix} \\
&= (3)(3) + (1)(4) + (4)(2) \\
&= 9 + 4 + 8 \\
&= 21
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{1,3} &= \boldsymbol{a}_1' \boldsymbol{b}_3 \\
&= \begin{bmatrix} 3 & 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \\
&= (3)(2) + (1)(2) + (4)(4) \\
&= 6 + 2 + 16 \\
&= 24
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{1,4} &= \boldsymbol{a}_1' \boldsymbol{b}_4 \\
&= \begin{bmatrix} 3 & 1 & 4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \\
&= (3)(2) + (1)(1) + (4)(4) \\
&= 6 + 1 + 16 \\
&= 23
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix}
$$

$$\begin{aligned}
c_{2,0} &= \boldsymbol{a}_2'\boldsymbol{b}_0 \\
&= \begin{bmatrix} 9 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 10 \end{bmatrix} \\
&= (9)(1) + (4)(1) + (1)(10) \\
&= 9 + 4 + 10 \\
&= 23
\end{aligned}$$

$$\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & - & - & - & - \\ - & - & - & - & - \end{bmatrix}$$

$$\begin{aligned}
c_{2,1} &= \boldsymbol{a}_2'\boldsymbol{b}_1 \\
&= \begin{bmatrix} 9 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 2 \\ 2 \end{bmatrix} \\
&= (9)(5) + (4)(2) + (1)(2) \\
&= 45 + 8 + 2 \\
&= 55
\end{aligned}$$

$$\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & - & - & - \\ - & - & - & - & - \end{bmatrix}$$

$$\begin{aligned}
c_{2,2} &= \boldsymbol{a}_2'\boldsymbol{b}_2 \\
&= \begin{bmatrix} 9 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix} \\
&= (9)(3) + (4)(4) + (1)(2) \\
&= 27 + 16 + 2 \\
&= 45
\end{aligned}$$

$$\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & - & - \\ - & - & - & - & - \end{bmatrix}$$

$$
\begin{aligned}
c_{2,3} &= \boldsymbol{a}_2'\boldsymbol{b}_3 \\
&= \begin{bmatrix} 9 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \\
&= (9)(2) + (4)(2) + (1)(4) \\
&= 18 + 8 + 4 \\
&= 30
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & 30 & - \\ - & - & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{2,4} &= \boldsymbol{a}_2'\boldsymbol{b}_4 \\
&= \begin{bmatrix} 9 & 4 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \\
&= (9)(2) + (4)(1) + (1)(4) \\
&= 18 + 4 + 4 \\
&= 26
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & 30 & 26 \\ - & - & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{3,0} &= \boldsymbol{a}_3'\boldsymbol{b}_0 \\
&= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 10 \end{bmatrix} \\
&= (4)(1) + (6)(1) + (2)(10) \\
&= 4 + 6 + 20 \\
&= 30
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & 30 & 26 \\ 30 & - & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{3,1} &= \boldsymbol{a}_3'\boldsymbol{b}_1 \\
&= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 2 \\ 2 \end{bmatrix} \\
&= (4)(5) + (6)(2) + (2)(2) \\
&= 20 + 12 + 4 \\
&= 36
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & 30 & 26 \\ 30 & 36 & - & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{3,2} &= \boldsymbol{a}_3'\boldsymbol{b}_2 \\
&= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix} \\
&= (4)(3) + (6)(4) + (2)(2) \\
&= 12 + 24 + 4 \\
&= 40
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & 30 & 26 \\ 30 & 36 & 40 & - & - \end{bmatrix}
$$

$$
\begin{aligned}
c_{3,3} &= \boldsymbol{a}_3'\boldsymbol{b}_3 \\
&= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix} \\
&= (4)(2) + (6)(2) + (2)(4) \\
&= 8 + 12 + 8 \\
&= 28
\end{aligned}
$$

$$
\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & 30 & 26 \\ 30 & 36 & 40 & 28 & - \end{bmatrix}
$$

$$\begin{aligned} c_{3,4} &= \boldsymbol{a}_3'\boldsymbol{b}_4 \\ &= \begin{bmatrix} 4 & 6 & 2 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} \\ &= (4)(2) + (6)(1) + (2)(4) \\ &= 8 + 6 + 8 \\ &= 22 \end{aligned}$$

$$\boldsymbol{AB} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & 30 & 26 \\ 30 & 36 & 40 & 28 & 22 \end{bmatrix}$$

Therefore, $\boldsymbol{AB} = \begin{bmatrix} 4 & 6 & 2 \\ 3 & 1 & 4 \\ 9 & 4 & 1 \\ 4 & 6 & 2 \end{bmatrix} \begin{bmatrix} 1 & 5 & 3 & 2 & 2 \\ 1 & 2 & 4 & 2 & 1 \\ 10 & 2 & 2 & 4 & 4 \end{bmatrix} = \begin{bmatrix} 30 & 36 & 60 & 28 & 32 \\ 44 & 20 & 21 & 24 & 23 \\ 23 & 55 & 45 & 30 & 26 \\ 30 & 36 & 40 & 28 & 22 \end{bmatrix}$

3.5 Identities of Matrix Operations

Now that we have defined the basic operations for matices, we can set forth some general rules that help us work with the alegbra of matrices. If $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are conformable, then the following is true:

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}) \tag{1}$$

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC} \tag{2}$$

$$(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}' \tag{3}$$

$$(\boldsymbol{ABC})' = \boldsymbol{C}'\boldsymbol{B}'\boldsymbol{A}' \tag{4}$$

$$(\boldsymbol{A} + \boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}' \tag{5}$$

I leave it as an exercise for you to practice with your matrix algebra, and to show that these identities are true. These will come in handy when we get to working with the mathematics of our regression model.

3.6 Geometric Interpretation of Vectors and Matrices

Many of the very important results that come from Matrix Algebra that helps make the math of some topics, like Machine Learning and Econometrics for example "cleaner" is actually motivated from geometric interpretations of vectors and matrices. All we need to do is interpret the elements of our vectors in a special way, and we can identify interesting geometric properties, which of course trickle down to algebraic properties. Typically, we interpret a vector $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

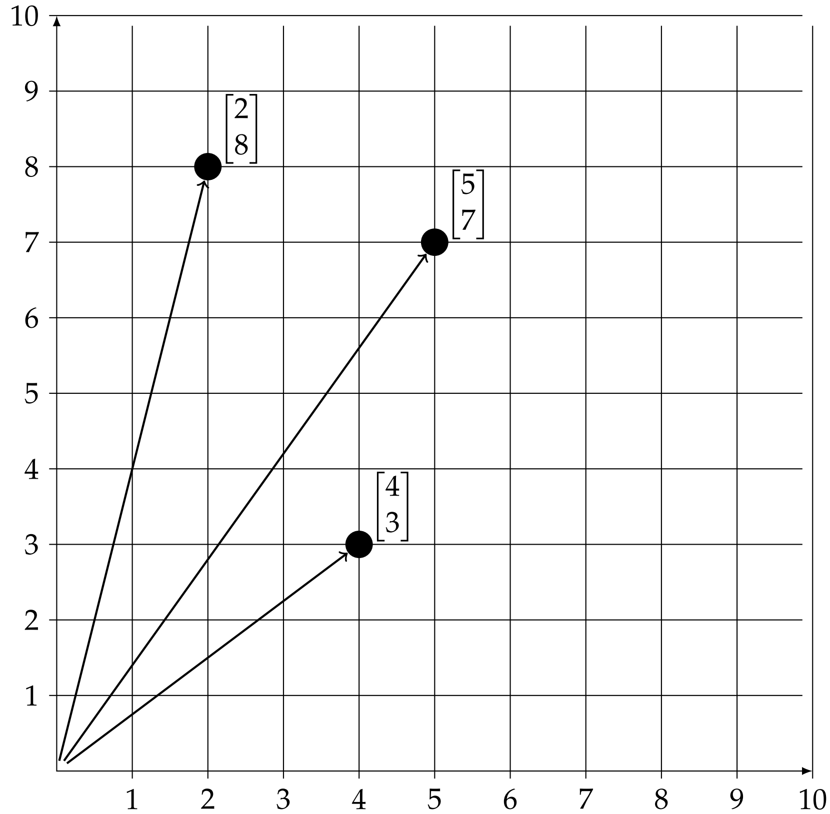

Figure 1

as a point in n-dimensional space. For example, if $n = 2$, then given the vector $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, this represents a point in 2-dimensional space, which we can visualize with the standard xy-plot. x_1 would tell us how much to the left (if it is negative) or to the right (if it is positive) we should move on the horizontal axis, which x_2 would tell us how much up (if it is positive) or down (if it is negative) we should move on the vertical axis. At this point in space, we then can think of the vector as a line that starts at the point $(0, 0)$ and "points" to the point (x_1, x_2). Generalizing this idea, we can think of each position n in the vector as the value that lies in the i-th axis (it is very difficult, but not impossible, to visualize vectors after $n = 3$ dimensions/axis, so we typically do much of our demonstrations in $n = 2$ or $n = 3$ dimensional space, despite that fact that many applications occur within very large dimensions).

As we can see from Figure 1, we have three vectors in 2-dimensional space. Namely, we have the vectors $\boldsymbol{x_1} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \boldsymbol{x_2} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}, \boldsymbol{x_3} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$. First, notice the indexing on the vectors themselves. You will see this from time to time in the literature. **DO NOT** confuse the indexing on the vector and the indexing on the elements of the vector. The notation $\boldsymbol{x_1}$ and x_1 are very different. $\boldsymbol{x_1}$ refers to the vector $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$, whereas x_1 refers to the value of the first dimension in some vector $\boldsymbol{x}$ (which we do not have defined in our current discussion). So, to refer to the first element in vector $\boldsymbol{x_1}$, we would write x_{11}, the second element in vector $\boldsymbol{x_1}$ would be written as x_{12}, etc.

There are two important geometric interpretations when it comes to operations on vectors.

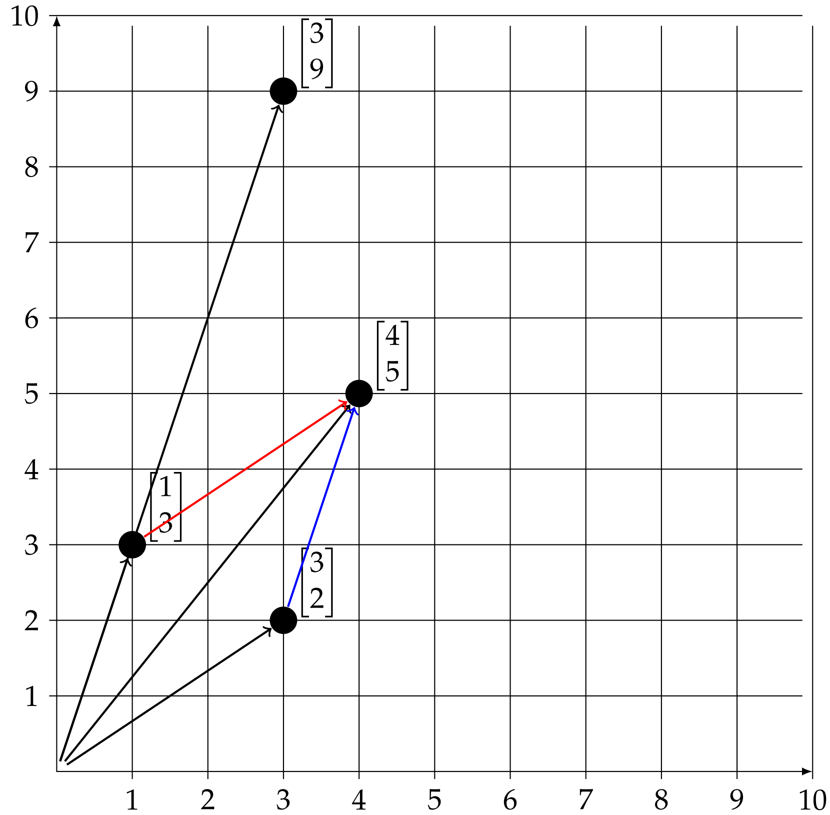

Figure 2

When we multiply a vector by a number (scalar multiplication), we obtain another vector that is in the same *direction* but is either shorter, longer, or 180 degrees opposite in length. Observe Figure 2, where $\boldsymbol{x_1} = \begin{bmatrix}1\\3\end{bmatrix}$ and $\boldsymbol{x_2} = \begin{bmatrix}3\\9\end{bmatrix}$. Notice that $\boldsymbol{x_2} = \begin{bmatrix}3\\9\end{bmatrix} = 3\begin{bmatrix}1\\3\end{bmatrix} = 3\boldsymbol{x_1}$. Notice how the vector $\boldsymbol{x_2}$ is on the same "line" as $\boldsymbol{x_1}$. This is always true of scalar multiplication of vectors. That is, the scalar multiple of a vector will be a new vector that is "longer" or "shorter", but on the same line as the original vector.

Also observe vector $\boldsymbol{x_3} = \begin{bmatrix}3\\2\end{bmatrix}$ and $\boldsymbol{x_4} = \begin{bmatrix}4\\5\end{bmatrix}$. Notice that $\boldsymbol{x_4} = \boldsymbol{x_1} + \boldsymbol{x_3} = \begin{bmatrix}1\\3\end{bmatrix} + \begin{bmatrix}3\\2\end{bmatrix}$. The geometric interpretation of vector addition is a "movement" of the point in the same direction and length of another vector. Think of it like this. We start at point $(0,0)$. We first "move" along vector $\boldsymbol{x_1}$ and land on the point $(1,3)$. When we "add" to this point the vector $\boldsymbol{x_3}$, the vector tells us to move 3 units to the right and 2 units up from the point $(1,3)$ (illustrated by the red line). We finally land on the point $(4,5)$, which is the same as vector $\boldsymbol{x_4}$. That is, vector addition tells us how to "move" in one or more "steps" to get to a point. Since addition is commutative, we should obtain the same result if we started to move from the origin using vector $\boldsymbol{x_3}$ first. Doing so would first move us from $(0,0)$ to the point $(3,2)$. We then "add" the vector $\boldsymbol{x_1}$ by starting at $(3,2)$ and moving 1 unit to the right and 3 units up, that is, we move according to the values in the vector $\boldsymbol{x_1}$ (illustrated by the blue line). The result is of course the same. We have shown geometrically that $\boldsymbol{x_4} = \boldsymbol{x_1} + \boldsymbol{x_3} = \boldsymbol{x_3} + \boldsymbol{x_1}$.

In general, we say that an operation $f(x, y)$ on a set A is *closed* if and only if $\forall x, y \in$

$A, f(x,y) \in A$. That is, an operation on a set is closed if and only if the operation on any two elements in the set gives us another element that is also in the set. Consider for example the set $\{0,1,2,3\}$. This is an example of a set that is **not** closed under the operation of addition. While some operations give us results that are in the set, such as 0+1 = 1, 1+2 = 3, other operations do not, like 3+1 = 4, and 4 is not in our set. However, if we were to take the same set and define a different operation, say the remainder of dividing the sum of any two elements by 3, then this is an operation that is closed under this set. If we consider the two elements from before, 3 and 1, we would have (3+1) mod 3 = 4 mod 3 = 1 (that is, the remainder after dividing 4 by 3 is 1), and this element is in the set.

The closure property is a very important one for our purposes, and we use this definition to define a very important concept known as a *vector space*. First, let us consider the set $\mathbb{R}^2 = \{(x,y)|x \in \mathbb{R} \wedge y \in \mathbb{R}\} = \{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} |x_1, x_2 \in \mathbb{R}\}$. We have two operations on vectors that we care about, namely scalar multiplication and vector addition. It is easy to show that these operations are closed under the set $\mathbb{R}^2$ (and, even generally in $\mathbb{R}^n$). Let's first consider scalar multiplication. Let $k \in \mathbb{R}$ and let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}$. Then $k\boldsymbol{x} = k \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} kx_1 \\ kx_2 \end{bmatrix}$. Since kx_1 is also in $\mathbb{R}$, and since kx_2 is also in $\mathbb{R}$, then by definition of the set $\mathbb{R}^2$, we must have $k\boldsymbol{x} \in \mathbb{R}^2$. A similar proof can be used to show that vector addition is closed in $\mathbb{R}^2$

When a subset $A \subset \mathbb{R}^n$ is shown to be closed under both addition and scalar multiplication, we say that A is a *vector space*. Let us consider the set of all vectors $A = \left\{ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \middle| x_1 \in \mathbb{R} \right\}$. We know that vectors in A have any number in the first element, but the second element must be 0.

For example $\begin{bmatrix} 4 \\ 0 \end{bmatrix} \in A, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in A, \begin{bmatrix} 50.3 \\ 0 \end{bmatrix} \in A$, but the vectors $\begin{bmatrix} 4 \\ 1 \end{bmatrix} \notin A, \begin{bmatrix} 6 \\ 56 \end{bmatrix} \notin A$, are not in A, because the second element is not 0, which means by definition of the set A, it would not belong in A. It would be easy to show that our set $A = \left\{ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \middle| x_1 \in \mathbb{R} \right\}$. Let $\boldsymbol{x_1}, \boldsymbol{x_2} \in A$. Then by definition, we know that $\exists k_1, k_2 \in \mathbb{R}$ such that $\boldsymbol{x_1} = \begin{bmatrix} k_1 \\ 0 \end{bmatrix}$ and $\boldsymbol{x_2} = \begin{bmatrix} k_2 \\ 0 \end{bmatrix}$. Let $k \in \mathbb{R}$. Then $k\boldsymbol{x_1} = k \begin{bmatrix} k_1 \\ 0 \end{bmatrix} = \begin{bmatrix} kk_1 \\ 0 \end{bmatrix}$. We know that since $k \in \mathbb{R}$ and $k_1 \in \mathbb{R}$, then $kk_1 \in \mathbb{R}$. Therefore, by definition, $k\boldsymbol{x_1} \in A$. Thus, our set is closed under scalar multiplication. If we add the two vectors, we obtain $\boldsymbol{x_1} + \boldsymbol{x_2} = \begin{bmatrix} k_1 \\ 0 \end{bmatrix} + \begin{bmatrix} k_2 \\ 0 \end{bmatrix} = \begin{bmatrix} k_1 + k_2 \\ 0 + 0 \end{bmatrix} = \begin{bmatrix} k_1 + k_2 \\ 0 \end{bmatrix}$. Notice that $k_1 + k_2 \in \mathbb{R}$, and that the second element is equal to 0. Therefore, by definition, we have $\boldsymbol{x_1} + \boldsymbol{x_2} \in A$. As such, we have shown that the set A is closed under the operations of scalar multiplication and vector addition, and hence it is a vector space.

4 Vector Spaces

In the previous section, we introduced the concept of a vector a space, which is a set of vectors such that their scalar multiples and the sum of any two vectors in the set are also in the set. However, we have yet to explain how one can characterize a vector space. In order to characterize a vector space, we typically find a few different vectors that we can use to *generate* other vectors in the set by multiplying them by scalars and then adding them together (if need be). Consider our example from before. We had the vector space $A = \left\{ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \middle| x_1 \in \mathbb{R} \right\}$. In order to characterize this vector space, we notice that the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ can be used to generate all the possible vectors in this space. If we take the vector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and multiply it by the scalar $k \in \mathbb{R}$, then we are able to obtain any vector that is in A. The set containing this vector, namely $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$ is referred to as the *basis*.

The basis is used to characterize the vector space. That is, the basis comprises of special types of vectors that can be multiplied by scalars and added together to form other vectors that are in the space. The number of vectors in this set is referred to as the *dimension* of the vector space. A fundamental fact is that if given a vector space A and B, and if $A \subset B$ and $\dim A = \dim B$, then $A = B$. For example. if we have a subset $A \subset \mathbb{R}^2$, and we have two vectors in the basis for the space A, then $A = \mathbb{R}^2$, otherwise, it does not. If $\dim A \neq \dim B$, then there must exist a vector in B that is not in A. We saw from our earlier example that $\dim A = 1$, where $A = \left\{ \begin{bmatrix} x_1 \\ 0 \end{bmatrix} \middle| x_1 \in \mathbb{R} \right\}$. We know that $A \subset \mathbb{R}^2$, but $\dim A \leq \dim \mathbb{R}^2$, and therefore there must be a vector in $\mathbb{R}^2$ that is not in A. This is easy to see for this example, just pick any vector where the second element is nonzero.

All vectors that are in the basis of a vector space must be *linearly independent*. We say that a set of vectors $\{\boldsymbol{x_0}, \boldsymbol{x_1}, \ldots, \boldsymbol{x_n}\}$ is *linearly dependent* if for any vector $\boldsymbol{x_i}$ in this collection, we can write $\boldsymbol{x_i} = \alpha_0 \boldsymbol{x_0} + \cdots + \alpha_{i-1} \boldsymbol{x_{i-1}} + \alpha_{i+1} \boldsymbol{x_{i+1}} + \cdots + \alpha_n \boldsymbol{x_n}$. For example, the set of vectors $\left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 10 \\ 10 \end{bmatrix} \right\}$ is linearly dependent. The reason is that given any of these vectors, we can write the given vector as a linear combination (that is, a sum of scalars multiplied by the other vectors) of the other vectors. We can see that $\begin{bmatrix} 10 \\ 10 \end{bmatrix} = \frac{10}{7} \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \frac{20}{7} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. Reworking this equation, we can also write $\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \frac{7}{10} \begin{bmatrix} 10 \\ 10 \end{bmatrix} - 2 \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ as well as $\begin{bmatrix} 3 \\ 2 \end{bmatrix} = \frac{7}{20} \begin{bmatrix} 10 \\ 10 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. This illustrates that if we are given a set of vectors, and if any of the vectors can be written as a linear combination of any of the other vectors, then we have shown that any of the vectors in the set can be written as a linear combination as any of the other vectors in the set. Therefore, to prove that a set is linearly dependent, all we must do is find one vector in the set under consideration that can be written as a linear combination of the other vectors.

Notice that in our previous example, if we were to remove the vector $\begin{bmatrix} 10 \\ 10 \end{bmatrix}$ from the set, leaving us with the set $\left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \end{bmatrix} \right\}$, then the result is a collection of vectors that are *linearly independent*. The reason is simple, we cannot write one vector in terms of the other. In order to do this, we would need to find a $k \in \mathbb{R}$ such that $\begin{bmatrix} 1 \\ 3 \end{bmatrix} = k \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. We see that we would need to find a k such that both of the equations are true:

$$1 = 3k$$
$$3 = 2k$$

If we solve the first equation, we obtain $k = \frac{1}{3}$. But if we plug this into the second equation, then we obtain $3 = 2\frac{1}{3} = \frac{2}{3}$, which is of course false. If we solve the second equation, then $k = \frac{3}{2}$, but then plugging this into the first equation gives us $1 = 3\frac{3}{2} = \frac{9}{2}$, which also is not true. Geometrically, we can also see that it is impossible to find a k such that $\begin{bmatrix} 1 \\ 3 \end{bmatrix} = k \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. Recall what we mentioned earlier, that scalar multiplications of vectors are "extensions" of the same vector. If we were able to find a k such that our above equations are true, then geometrically, the two vectors should be on the same line, which we see from Figure 2 is just not possible, as these vector lie on entirely different lines. Therefore, it is impossible to write one vector as a linear combination of the other vector(s), and hence, this collection is *linearly independent*.

When given a collection of vectors, we can take linear combinations of them to form new vectors. The resulting collection of vectors, by definition, is a vector space. We typically call this vector space the *space spanned by* the original set of vectors given to us. More formally, if we are given a set of K vectors $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_K}$ that are linearly independent, then the set of vectors $\{\boldsymbol{x} = \alpha_1 \boldsymbol{x_1} + \alpha_2 \boldsymbol{x_2} + \ldots \alpha_K \boldsymbol{x_K} | \forall \alpha_1, \ldots, \alpha_K \in \mathbb{R}\}$ is referred to as the *vector space spanned by* $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \ldots, \boldsymbol{x_K}$. We say that the dimension of this vector space is $\dim K$, since by construction, we assumed the K vectors were linearly independent and can be used to find another vector in this space.

4.1 Spaces of a Matrix

Given any matrix, we can consider the columns of a matrix to be individual vectors. If we consider the collection of vectors that are in the columns of a given matrix, then we can obtain a vector space that is obtained by those columns. This space is referred to as the *column space*. If we take the transpose of the matrix, we obtain new columns, which themselves have a space. Since the columns in the transpose of a matrix are the rows in the original matrix, we refer to the vector space spanned by the columns of a transpose as the *row space*.

The number of linearly independent columns in a matrix is referred to as the *column rank*, while the number of linearly independent rows in a matrix is referred to as the *row rank*. These numbers also refer to the dimension of the column space and the row space. It is true (I will leave it as an exercise for you to prove it) that the row rank is always equal to the column rank, and generally, we just simply refer to the "rank" of a matrix. That is, the rank of a matrix is the number of linearly independent columns in the matrix. If the rank of the matrix is equal to the number of columns in the matrix, then we say that the matrix has *full rank*. These types of matrices have applications in the fields of statistics, machine learning, and Econometric analysis, since they allow for us to compute statistical estimates of coefficients.

As an example, consider the matrix $\begin{bmatrix} 3 & 4 & 6 \\ 5 & 2 & 3 \\ 8 & 0 & 4 \end{bmatrix}$. This matrix has the following vectors in the columns: $\left\{ \begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ 3 \\ 4 \end{bmatrix} \right\}$. In the next section, we will show how to determine if vectors are linearly independent. In this example, it is fairly easy to see this. Consider the vector $\begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix}$. We want to see if there exists an $\alpha_1, \alpha_2 \in \mathbb{R}$ such that $\begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix} = \alpha_1 \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 6 \\ 3 \\ 4 \end{bmatrix}$.

We see that the only way that is is possible is if $\alpha_2 = 2$. When this is the case, then the last number (4) in the last vector becomes 8. This is the only value of α_2 that will satisfy the last element's equality, since every number α_1 is multiplied by 0 on the last element of the first vector. However, if $\alpha_2 = 2$, then $\alpha_1 = -\frac{1}{2}$, since we need $5 = \alpha_1(2) + \alpha_2(3) = \alpha_1(2) + (2)(3) = \alpha_1(2) + 6$, and solving for α_1 gives us the correct value for the second element in the vector. However, we notice that $\alpha_1(4) + \alpha_2(6) = (-\frac{1}{2})(4) + (2)(6) = -2 + 12 = 10 \neq 3$. Therefore, we cannot write the first vector as a linear combination of the other two, and hence, the vectors must all be linearly independent.

This illustrates that the rank of our matrix is equal to 3 (since we have three linearly independent column vectors in our matrix). We also notice that the matrix is full rank, since there are 3 columns in the matrix, and the matrix has rank 3. Determining whether or not columns of a matrix are linearly independent, and how many and which ones are linearly independent, is difficult to accomplish using this method. We hence need to resort to another method that is more generic so we can determine the ranks of a matrix, as well as the basis for the column space (or any vector space spanned by a given set of vectors).

4.1.1 Matrix Manipulation and Computing Ranks

When given a matrix, we need to determine how many, and which, columns in the matrix are linearly independent. One could use a trial and error approach to determine this, however, this is too cumbersome. Instead, what we try to do is take the matrix and reduce it to a particular form that will tell us which columns can serve as the basis for the column space, as well as be able to compute the rank of the matrix.

When given a matrix, we can make certain manipulations to the matrix which will "preserve" it in a special way. For example, we can switch rows, or switch columns. We can multiply an entire row by a single number, and we can add a number or multiples of a row to another row. These are called *elementary row operations*, and they allow us to put a matrix in a "better" form.

This "better form" is usually one that results in the columns (or some of the columns) being the same as *basis vectors*. Recall that a basis vector is a vector that is in the form of $\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$. The goal is to reduce a matrix from it's current form to a form that "looks like" a matrix where there are basis vectors in the columns. The resulting columns that are basis vectors are the columns that are linearly independent, with the other columns (if there are any) that are not basis vectors are vectors that can be written as a linear combination of the basis vectors.

We will illustrate this with a simple example. Suppose we have the matrix $\begin{bmatrix} 6 & 3 & 2 \\ 5 & 1 & 8 \end{bmatrix}$. We want to reduce this matrix in a way that will give us basis vectors. Recall that we do this by using *elementary row operations*. First, we divide the entire first row by the first number in the first column, to obtain: $\begin{bmatrix} \frac{6}{6} & \frac{3}{6} & \frac{2}{6} \\ 5 & 1 & 8 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 5 & 1 & 8 \end{bmatrix}$. Now we multiply each element in the first row by the negative of the second element in the first column (5), and take the result and add it to the second row: $\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 5-(5) & 1-\frac{5}{2} & 8-\frac{5}{3} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & -\frac{3}{2} & \frac{19}{3} \end{bmatrix}$. Now we divide the entire second row by the number $-\frac{3}{2}$ to obtain: $\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{-\frac{3}{2}}{-\frac{3}{2}} & \frac{\frac{19}{3}}{-\frac{3}{2}} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & 1 & -\frac{38}{9} \end{bmatrix}$. Now we can multiply the negative of the first number in the second column by the last row and add it to the first row. This will remove the first element in the second column: $\begin{bmatrix} 1-(\frac{1}{2})(0) & \frac{1}{2}-(\frac{1}{2})(1) & \frac{1}{3}-(\frac{1}{2})(-\frac{38}{9}) \\ 0 & 1 & -\frac{38}{9} \end{bmatrix} = \begin{bmatrix} 1-0 & \frac{1}{2}-(\frac{1}{2}) & \frac{1}{3}+\frac{19}{9} \\ 0 & 1 & -\frac{38}{9} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{22}{9} \\ 0 & 1 & -\frac{38}{9} \end{bmatrix}$. We notice that there are no more row operations we can perform to reduce the last column down to a basis vector. This last matrix gives us a lot of information. First, it tells us that the columns in the matrix are linearly dependent (that is, we can obtain a column in the matrix by finding linear combinations of the other columns). It also tells us that this matrix is **not** full rank. The rank of the original matrix is equal to the number of linearly independent columns. The number of columns in the reduced matrix that are unique basis vectors is the rank of the matrix. Here, we see that we have two such columns, namely $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. This tells us that the vectors $\left\{ \begin{bmatrix} 6 \\ 5 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\}$ in the original matrix are linearly independent, since they eventually reduce down to two known linearly independent vectors. Since the last column in our reduced matrix is **not** a basis vector, this tells us that the third column in the original matrix can be found as a linear combination of the first two columns.

If we are curious to see which two constants $\alpha_1, \alpha_2 \in \mathbb{R}$ can be multiplied to our basis vectors in order to obtain the third column, we can set up some equations:

$$\alpha_1 \begin{bmatrix} 6 \\ 5 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$$

This is equivalent to writing:

$$\begin{bmatrix} 6 & 3 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \end{bmatrix}$$

We then can form what is called an *augmented matrix*, which looks like this:

$$\left[\begin{array}{cc|c} 6 & 3 & 2 \\ 5 & 1 & 8 \end{array}\right]$$

When we conduct our row operations on the augmented matrix, we are in essence solving for the values of $\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}$. Earlier, we reduced this matrix to:

$$\left[\begin{array}{cc|c}1 & 0 & \frac{22}{9}\\ 0 & 1 & -\frac{38}{9}\end{array}\right]$$

and this is equivalent to writing

$$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} = \begin{bmatrix}\frac{22}{9}\\ -\frac{38}{9}\end{bmatrix}.$$

But we know that $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} = \begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}$. Therefore, we have $\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix} = \begin{bmatrix}\frac{22}{9}\\ -\frac{38}{9}\end{bmatrix}$

You can test this by replacing the values for the coefficients into our linear combination from earlier:

$$\alpha_1\begin{bmatrix}6\\5\end{bmatrix} + \alpha_2\begin{bmatrix}3\\1\end{bmatrix} = (\frac{22}{9})\begin{bmatrix}6\\5\end{bmatrix} + (-\frac{38}{9})\begin{bmatrix}3\\1\end{bmatrix} = \begin{bmatrix}(\frac{22}{9})6\\ (\frac{22}{9})5\end{bmatrix} + \begin{bmatrix}(-\frac{38}{9})3\\ (-\frac{38}{9})1\end{bmatrix} = \begin{bmatrix}\frac{132}{9}\\ \frac{110}{9}\end{bmatrix} + \begin{bmatrix}-\frac{114}{9}\\ -\frac{38}{9}\end{bmatrix} = \begin{bmatrix}\frac{18}{9}\\ \frac{72}{9}\end{bmatrix} = \begin{bmatrix}2\\8\end{bmatrix}$$

Generally, when we are given a matrix and reduce it to a form where there are columns that are basis vectors, we have the matrix in what is referred to as *reduced row-echelon form*. The procedure we followed is called the *Gauss-Jordan Elimination* algorithm. The algorithm works by identifying *pivots* in the matrix, and using the pivots to reduce each column down to a basis vector. The procedure, generally, works as such:

1. Start with the element in the 1st row and the 1st column in the matrix. Divide the entire first row by this number.

2. For each element in the first column that is below the first element in the first column, multiply the first row by the negative of that number. Add the resulting row to the current row, this will turn the value to 0 in the first column of that row.

3. Move to the next column, repeat the same process. Now, repeat the process for the rows above the second row in the second column. That is continue to eliminate the numbers in the column until they are all zero, with the exception of the row that has the 1 in it.

4. Repeat this process until you have a value of 1 on the bottom row. At this point, you are done, and if there are remaining columns, then these columns are not in the basis for the column space. Otherwise, you continue so long as you have more columns. When you run out of columns, you are done.

Sometimes, you will be unable to obtain a 1 on the diagonal since there will already be a 0 in it's place.. In situations like this, you can swap columns so that a non-zero value is in it's place.

We illustrate the algorithm again, in detail, with the following matrix: $\begin{bmatrix}2 & 2 & 5 & 2\\ 4 & 0 & 1 & 4\\ 3 & 9 & 3 & 1\\ 1 & 1 & 3 & 2\end{bmatrix}$

First, we start with the first column. We notice that the first element in the first row in the first column is 2. So we divide the entire first row by 2:

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

Notice that this has allowed us to easily use the number one as a *pivot*. We can now eliminate the lower numbers in the column (namely, 4,3,and 1) using the number 1. To convert the 4 in the second row of the first column, we multiply the first row by the number (-4) and add the two rows:

$$\begin{bmatrix} 1 & 1 & \frac{5}{2} & 1 \\ 4-(4)(1) & 0-(4)(1) & 1-(\frac{5}{2})(4) & 4-(4)(1) \\ 3 & 9 & 3 & 1 \\ 1 & 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{5}{2} & 1 \\ 0 & -4 & -9 & 0 \\ 3 & 9 & 3 & 1 \\ 1 & 1 & 3 & 2 \end{bmatrix}$$

Now we want to convert the value of 3 in the 3rd row in the first column to 0. We do this by again leveraging the pivot in the first row of the first column. To make the conversion, we multiply the first row by the negative of the number we would like to eliminate. So, we multiply the first row by the number (-3), and then add this row to the third row in the matrix:

$$\begin{bmatrix} 1 & 1 & \frac{5}{2} & 1 \\ 0 & -4 & -9 & 0 \\ 3-(3)(1) & 9-(3)(1) & 3-(3)(\frac{5}{2}) & 1-(3)(1) \\ 1 & 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{5}{2} & 1 \\ 0 & -4 & -9 & 0 \\ 0 & 6 & -\frac{9}{2} & -2 \\ 1 & 1 & 3 & 2 \end{bmatrix}$$

Finally, we want to remove the value of 1 that is in the first column in the last row. This is easy to do since it is a value of 1. All we need to do is subtract the 1st row from the last row, which will convert the 1 in the last row into a 0:

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

Now we have the first column reduced down to a standard basis vector. We would like to do the same for the second column, but this time with the "1" in the second position of the vector instead of the first. We will follow our procedure all over again with a slight small change. First, let us do the same as before: divide the entire second row by the value that is in our pivot position (which is now the second row and second column). This value is -4, so we will divide out the entire second row by -4:

$$\begin{bmatrix} 1 & 1 & \frac{5}{2} & 1 \\ \frac{0}{-4} & \frac{-4}{-4} & \frac{-9}{-4} & \frac{0}{-4} \\ 0 & 6 & -\frac{9}{2} & -2 \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \frac{5}{2} & 1 \\ 0 & 1 & \frac{9}{4} & 0 \\ 0 & 6 & -\frac{9}{2} & -2 \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}$$

Now we will use the second row, second column as the pivot to remove the values in the second column at rows 1,3, and 4. To remove the 1 from the first row in the second column, we subtract the second row from the first row, which gives us:

$$\begin{bmatrix} 1-0 & 1-1 & \frac{5}{2}-\frac{9}{4} & 1-0 \\ 0 & 1 & \frac{9}{4} & 0 \\ 0 & 6 & -\frac{9}{2} & -2 \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{1}{4} & 1 \\ 0 & 1 & \frac{9}{4} & 0 \\ 0 & 6 & -\frac{9}{2} & -2 \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}$$

Lucky for us, the last row in the second column is already 0, so we only have one more row to eliminate, which is the 6 in the second column. We again multiply the negative of this number by the row where the pivot is, and add this to the row we are trying to eliminate. Doing this leads us to:

$$\begin{bmatrix} 1 & 0 & \frac{1}{4} & 1 \\ 0 & 1 & \frac{9}{4} & 0 \\ 0-(6)(0) & 6-(6)(1) & -\frac{9}{2}-(6)(\frac{9}{4}) & -2-(6)(0) \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{1}{4} & 1 \\ 0 & 1 & \frac{9}{4} & 0 \\ 0 & 0 & -\frac{36}{2} & -2 \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{1}{4} & 1 \\ 0 & 1 & \frac{9}{4} & 0 \\ 0 & 0 & -18 & -2 \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}$$

We have now finished the second column, but we still have two more columns left. We now move to the third column, where our pivot point will be in the third row of the third column (-18). We again divide this entire row by the number in the pivot:

$$\begin{bmatrix} 1 & 0 & \frac{1}{4} & 1 \\ 0 & 1 & \frac{9}{4} & 0 \\ \frac{0}{-18} & \frac{0}{-18} & \frac{-18}{-18} & \frac{-2}{-18} \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{1}{4} & 1 \\ 0 & 1 & \frac{9}{4} & 0 \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}$$

Now we use this to eliminate the $\frac{1}{4}$ in the first row by multiplying the third row by $-\frac{1}{4}$ and adding it to the first row: $\begin{bmatrix} 1-(\frac{1}{4})(0) & 0-(\frac{1}{4})(0) & \frac{1}{4}(\frac{1}{4})(1) & 1-(\frac{1}{4})(\frac{1}{9}) \\ 0 & 1 & \frac{9}{4} & 0 \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \frac{35}{36} \\ 0 & 1 & \frac{9}{4} & 0 \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}$ Now we eliminate the second row, which has the value of $\frac{9}{4}$, by multiplying the third row by $-\frac{9}{4}$ and adding it to the second row:

$$\begin{bmatrix} 1 & 0 & 0 & \frac{35}{36} \\ 0-(\frac{9}{4})(0) & 1-(\frac{9}{4})(0) & \frac{9}{4}-(\frac{9}{4})(1) & 0-(\frac{9}{4})(\frac{1}{9}) \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \frac{35}{36} \\ 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & \frac{1}{2} & 1 \end{bmatrix}$$

Finally, we remove the $\frac{1}{2}$ in the last row of the third column. We do this by again multiplying the third row with the value of $-\frac{1}{2}$ and adding this to the last row:

$$\begin{bmatrix} 1 & 0 & 0 & \frac{35}{36} \\ 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{9} \\ 0-(\frac{1}{2})(0) & 0-(\frac{1}{2})(0) & \frac{1}{2}-(\frac{1}{2})(1) & 1-(\frac{1}{2})(\frac{1}{9}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \frac{35}{36} \\ 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & 0 & \frac{17}{18} \end{bmatrix}$$

Now we move to the last column, which has the pivot in the location of the fourth column in the fourth row, which has a value of $\frac{17}{18}$. First, we divide the entire row by this number, which gives us: $\begin{bmatrix} 1 & 0 & 0 & \frac{35}{36} \\ 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{9} \\ \frac{0}{\frac{17}{18}} & \frac{0}{\frac{17}{18}} & \frac{0}{\frac{17}{18}} & \frac{\frac{17}{18}}{\frac{17}{18}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \frac{35}{36} \\ 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Now we once again want to remove the non-pivot elements of our vector. These are the elements in the fourth column in rows 1, 2, and 3, respectively. To do so, we first multiply the last row by the negative of the value in the last column we seek to remove, and add the result to the row. We shorten up the illustration here:

$$\begin{bmatrix} 1-(\frac{35}{36})(0) & 0-(\frac{35}{36})(0) & 0-(\frac{35}{36})(0) & \frac{35}{36}-(\frac{35}{36})(1) \\ 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0+(\frac{1}{4})(0) & 1+(\frac{1}{4})(0) & 0+(\frac{1}{4})(0) & -\frac{1}{4}+(\frac{1}{4})(1) \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{9} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0-(\frac{1}{9})(0) & 0-(\frac{1}{9})(0) & 1-(\frac{1}{9})(0) & \frac{1}{9}-(\frac{1}{9})(1) \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We notice that we get the identity matrix as a result. Furthermore, all of the vectors were transformed to the standard basis vectors. This tells us that our original 4 columns in the original matrix were linearly independent. Since there are 4 linearly independent vectors, the rank of the matrix is 4, and since this is exactly the same number of columns we have in our matrix, this tells us that the matrix has *full rank*.

4.2 Determinants

When we have a square matrix, there is a special number that will help us tremendously in some of our computations. This number is referred to as the *determinant*, and it has many applications when we are employing matrix algebra. Geometrically, the determinant represents the volume of a structure in n-dimensional space. We will forgo further description, since this requires a bit of knowledge of *linear transformations*, which is a bit outside the scope of our discussion. Now, there are many different ways to compute the determinant. We will cover one of the more general methods. Let $\boldsymbol{A}$ be a matrix. Then the determinant of the matrix, denoted as $\det A = |A|$ is found by $\det A = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma_i}$

In plain English, this says to take the summation over every possible *permutation* of n numbers, where the sgn(σ) function returns either 1 or -1 based on the permutation and the $\prod$ says to take the product over the elements in the matrix at row i and column $\sigma(i)$. We will demonstrate this with an example. Suppose we had the 2x2 matrix $\begin{bmatrix} 6 & 3 \\ 5 & 1 \end{bmatrix}$. We first need to find all possible permutations of the numbers $\{1,2\}$. That is, a permutation is a function from a set of n numbers to itself which indicates how to "order" the numbers. If $n = 2$, we only have two possible orderings, namely (1 2) and (2 1). Therefore, our summation will have only two terms since there are only two permutations.

Next, we need to compute the sgn function for each permutation. This is done by using the equation sgn $(\sigma) = \frac{P(x_{\sigma(1)}, x_{\sigma(2)})}{P(x_1, x_2)}$, where $P(x_1, x_2, \ldots, x_n) = \prod_{i<j}(x_i - x_j)$. Here, we would have $P(x_{\sigma(1)}, x_{\sigma(2)}) = (x_{\sigma(1)} - x_{\sigma(2)})$ and $P(x_1, x_2) = (x_1 - x_2)$. Therefore we have $\text{sgn}(\sigma) = \frac{P(x_{\sigma(1)}, x_{\sigma(2)})}{P(x_1, x_2)} = \frac{(x_{\sigma(1)} - x_{\sigma(2)})}{(x_1 - x_2)}$. For the first permutation (1 2), we have $\sigma(1) = 1$ and $\sigma(2) = 2$. There-

fore, we would have $\text{sgn}((12)) = \frac{(x_{(12)(1)} - x_{(12)(2)})}{(x_1 - x_2)} = \frac{(x_1 - x_2)}{(x_1 - x_2)} = 1$. For the other permutation, we would have $\sigma(1) = 2$ and $\sigma(2) = 1$, so $\text{sgn}((21)) = \frac{(x_{(21)(1)} - x_{(21)(2)})}{(x_1 - x_2)} = \frac{(x_2 - x_1)}{(x_1 - x_2)} = -1$. Now, following the equation, we can compute the determinant. We have
$det(A) = \sum_{\sigma \in \{(12),(21)\}} \text{sgn}(\sigma) a_{1,\sigma(1)} a_{2,\sigma(2)} = \text{sgn}((12)) a_{1,1} a_{2,2} + \text{sgn}((21)) a_{1,2} a_{2,1} = (1)(6)(1) + (-1)(3)(5) = 6 - 15 = -9.$

As another example, let us take the 3x3 matrix $\boldsymbol{A} = \begin{bmatrix} 5 & 3 & 1 \\ 1 & 2 & 3 \\ 6 & 4 & 6 \end{bmatrix}$. First, we list out all the possible permutations for $n = 3$:

Permutations of $\{1,2,3\}$
(123)
(132)
(213)
(231)
(312)
(321)

The number in the first position of the notation (abc) means that $\sigma(1) = a$,the second position means that $\sigma(2) = b$, and the third position means that $\sigma(3) = c$. For example, the permutation (132) means that $\sigma(1) = 1, \sigma(2) = 3, \sigma(3) = 2$, and this is how the function is constructed. In general, if we have n numbers, there will be $n!$ possible permutations. In our case, $n = 3$, so there are a total of $3! = (3)(2)(1) = 6$ unique permutations (each one we have accounted for above).

Now, we need to compute $\text{sgn}(\sigma)$ for each possible permutation. Following the same rules as before, we know that for $n = 3$, $\text{sgn}\ (\sigma) = \frac{P(x_{\sigma(1)}, x_{\sigma(2)}, x_{\sigma(3)})}{P(x_1, x_2, x_3)} = \frac{(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)}$. Furthermore, we know that the terms we are going to multiply for each permutation will be $a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,sigma(3)}$. Taking these into account, we can list the computations below in the table:

σ	Assignments	$\text{sgn}(\sigma)$	$a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,sigma(3)}$
(123)	$\sigma(1) = 1, \sigma(2) = 2, \sigma(3) = 2$	$\frac{(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = \frac{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = 1$	$a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,sigma(3)} = a_{1,1} a_{2,2} a_{3,3} = (5)(2)(6) = 60$
(132)	$\sigma(1) = 1, \sigma(2) = 3, \sigma(3) = 2,$	$\frac{(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = \frac{(x_1 - x_3)(x_1 - x_2)(x_3 - x_2)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = -1$	$a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,sigma(3)} = a_{1,1} a_{2,3} a_{3,2} = (5)(3)(4) = 60$
(213)	$\sigma(1) = 2, \sigma(2) = 1, \sigma(3) = 3$	$\frac{(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = \frac{(x_2 - x_1)(x_2 - x_3)(x_1 - x_3)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = -1$	$a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,sigma(3)} = a_{1,2} a_{2,1} a_{3,3} = (3)(1)(6) = 18$
(231)	$\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$	$\frac{(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = \frac{(x_2 - x_3)(x_2 - x_1)(x_3 - x_1)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = 1$	$a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,sigma(3)} = a_{1,2} a_{2,3} a_{3,1} = (3)(3)(6) = 54$
(312)	$\sigma(1) = 3, \sigma(2) = 1, \sigma(3) = 2$	$\frac{(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = \frac{(x_3 - x_1)(x_3 - x_2)(x_1 - x_2)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = 1$	$a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,sigma(3)} = a_{1,3} a_{2,1} a_{3,2} = (1)(1)(4) = 4$
(321)	$\sigma(1) = 3, \sigma(2) = 2, \sigma(3) = 1$	$\frac{(x_{\sigma(1)} - x_{\sigma(2)})(x_{\sigma(1)} - x_{\sigma(3)})(x_{\sigma(2)} - x_{\sigma(3)})}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = \frac{(x_3 - x_2)(x_3 - x_1)(x_2 - x_1)}{(x_1 - x_2)(x_1 - x_3)(x_2 - x_3)} = -1$	$a_{1,\sigma(1)} a_{2,\sigma(2)} a_{3,sigma(3)} = a_{1,3} a_{2,2} a_{3,1} = (1)(2)(6) = 12$

Using the table we constructed, we can compute the determinant:
$\det A = (1)(60) + (-1)(60) + (-1)(18) + (1)(54) + (1)(4) + (-1)(12) = 28$

4.3 Systems of Linear Equations

We now have covered enough material to discuss a very important application of Linear Algebra. Namely that of solving systems of linear equations. We encounter systems of linear equations quite frequently in many business applications, which is why having an understanding of it is of the utmost importance. A *system of linear equations* is a collection of equations that are linear. A function of n variables is said to be *linear* if it can be put into the following format:

$f(x_1, x_2, \ldots, x_n) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$.

Each $\alpha_i \in \mathbb{R}$. Therefore, a system of linear equations is a collection of equations where each equation is essentially a linear function in n variables. Generally, we can write a system of linear equations as:

$$\begin{aligned} \alpha_{0,1}x_1 + \alpha_{0,2}x_2 + \cdots + \alpha_{0,n}x_n &= b_0 \\ \alpha_{1,1}x_1 + \alpha_{1,2}x_2 + \cdots + \alpha_{1,n}x_n &= b_1 \\ &\vdots \\ \alpha_{m,1}x_1 + \alpha_{m,2}x_2 + \cdots + \alpha_{m,n}x_n &= b_m \end{aligned}$$

Each equation represents a hyper-plane if n-dimensional space. A *solution* to the system is a collection of values assigned to the variables $x_1, x_2, \ldots, x_n$ such that every equation above is true. If any of the equations are false for a given assignment of values, then the assignment itself is **not** a solution to the system. Geometrically, solutions represent intersections of the lines and hyper-planes.

There are only three possibilities for systems of linear equations. Either the system has no solution, exactly one unique solution, or an infinite number of solutions. Given the system above, we can simplify our notation using our newfound matrix notation if we let

$\boldsymbol{A} = \begin{bmatrix} \alpha_{0,1} & \alpha_{0,2} & \ldots & \alpha_{0,n} \\ \alpha_{1,1} & \alpha_{1,2} & \ldots & \alpha_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \ldots & \alpha_{m,n} \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$, and $\boldsymbol{b} = \begin{bmatrix} b_0 \\ b_1 \\ \ldots \\ b_m \end{bmatrix}$

then we can represent our system of equations in the matrix format as $\boldsymbol{Ax} = \boldsymbol{b}$. Obviously we can see this is much easier to handle than all of those equations. The next question is, how do we solve for $\boldsymbol{x}$? There are a few ways.

The first method we can take is to consider the augmented matrix $\left[\begin{array}{cccc|c} \alpha_{0,1} & \alpha_{0,2} & \ldots & \alpha_{0,n} & b_0 \\ \alpha_{1,1} & \alpha_{1,2} & \ldots & \alpha_{1,n} & b_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \ldots & \alpha_{m,n} & b_m \end{array}\right]$ and perform Gauss-Jordan Elimination on it. The resulting vector on the right of the line, as we illustrated earlier, would be the solution to the system. Another method is to first compute the *inverse* of the matrix A, and then multiply both sides by the inverse. An *inverse* to a square matrix $\boldsymbol{A}$ is a matrix $\boldsymbol{A^{-1}}$ such that $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the nxn identity matrix. If we are able to find an inverse matrix to $\boldsymbol{A}$, then we can easily solve the system:

$\boldsymbol{Ax} = \boldsymbol{b}$
$\boldsymbol{A^{-1}Ax} = \boldsymbol{A^{-1}b}$
$\boldsymbol{Ix} = \boldsymbol{A^{-1}b}$
$\boldsymbol{x} = \boldsymbol{A^{-1}b}$

An inverse matrix A^{-1} can be found by solving the system of equations $\boldsymbol{AA^{-1}} = \boldsymbol{I}$. A solution to this can be found by augmenting the matrix $\boldsymbol{A}$ with an nxn identity matrix and performing Gauss-Jordan Elimination. The goal is to turn the left of the matrix to the identity matrix, and the matrix on the right will be the inverse of the matrix.

$$\begin{bmatrix} \alpha_{0,1} & \alpha_{0,2} & \dots & \alpha_{0,n} & | & 1 & 0 & \dots & 0 \\ \alpha_{1,1} & \alpha_{1,2} & \dots & \alpha_{1,n} & | & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & | & \vdots & \vdots & \vdots & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \dots & \alpha_{m,n} & | & 0 & 0 & \dots & 1 \end{bmatrix}$$

We will forego an illustration of this process. An important feature of matrix inverses is their existence. Not all Matrices have an inverse. There is a fairly easy way to test if a matrix has an inverse. If the matrix is a square matrix, and it's determinant is equal to 0, then the matrix does **not** have an inverse. On the other hand, if the determinant is non-zero, then it does have an inverse. Another important property of inverses is that if we have two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ where their inverses,$\boldsymbol{A}^{-1}$ and $\boldsymbol{B}^{-1}$ exist, then $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

5 Eigenvalues and Eigenvectors

Suppose we have a number $\lambda \in \mathbb{R}$ and a square matrix $\boldsymbol{A}$. Then we would like to solve the equation $\boldsymbol{Ax} = \lambda \boldsymbol{x}$. This equation basically says that when we multiply the vector $\boldsymbol{x}$ by the matrix $\boldsymbol{A}$, we obtain a scalar extension or contraction of the same vector $\boldsymbol{x}$. If there is a solution to this equation, then λ is said to be an eigenvalue, and $\boldsymbol{x}$ is said to be an eigenvector.

Eigenvectors and eigenvalues have very interesting properties that allow us to easily find and represent information regarding our matrix. For example, it can be used to find the determinant of the matrix, as well as understand solutions to optimization problems. Working on this problem, we can find the values and vectors in the following manner. We notice that if we subtract the vector $\lambda\boldsymbol{x}$ from both sides, we obtain $\boldsymbol{Ax} - \lambda\boldsymbol{x} = \boldsymbol{0}$. Notice, however, that $\lambda\boldsymbol{x} = \lambda\boldsymbol{Ix}$, and so we can use the distributive property to pull out the $\boldsymbol{x}$: $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$. But, the identity matrix $\boldsymbol{I}$ is just the scalar matrix with zeros everywhere and with the number λ on the diagonal.

The way that we can solve for the eigenvalues involves a clever trick. Since $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$, this means that if λ is indeed a solution, then it must be true that $\det((\boldsymbol{A} - \lambda\boldsymbol{I})) = \det(\boldsymbol{0}) = 0$. Hence, we can use the formula for the determinant to obtain a polynomial in terms of the variable λ. This polynomial is referred to as the *characteristic polynomial*, and the roots of it are the possible values for the *eigenvalues*. Once we obtain the values for the eigenvalues, we can then solve the equation above to find the eigenvectors.

To illustrate this process, let us consider the matrix $\boldsymbol{A} = \begin{bmatrix} 2 & 6 & 1 \\ 5 & 3 & 6 \\ 1 & 1 & 2 \end{bmatrix}$. Working with the equation given above, we have:

$$
\begin{aligned}
&\det\left(\boldsymbol{A} - \lambda \boldsymbol{I}\right) = 0\\
&\det\left(\begin{bmatrix} 2 & 6 & 1\\ 5 & 3 & 6\\ 1 & 1 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}\right) = 0\\
&\det\left(\begin{bmatrix} 2 & 6 & 1\\ 5 & 3 & 6\\ 1 & 1 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda \end{bmatrix}\right) = 0\\
&\det\left(\begin{bmatrix} 2-\lambda & 6 & 1\\ 5 & 3-\lambda & 6\\ 1 & 1 & 2-\lambda \end{bmatrix}\right) = 0
\end{aligned}
$$

Using the equation for the determinant, we have:

σ	Assignments	sgn(σ)	$a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,sigma(3)}$
(123)	$\sigma(1)=1, \sigma(2)=2, \sigma(3)=2$	$\frac{(x_{\sigma(1)}-x_{\sigma(2)})(x_{\sigma(1)}-x_{\sigma(3)})(x_{\sigma(2)}-x_{\sigma(3)})}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = \frac{(x_1-x_2)(x_1-x_3)(x_2-x_3)}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = 1$	$a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,sigma(3)} = a_{1,1}a_{2,2}a_{3,3} = (2-\lambda)(3-\lambda)(2-\lambda)$
(132)	$\sigma(1)=1, \sigma(2)=3, \sigma(3)=2,$	$\frac{(x_{\sigma(1)}-x_{\sigma(2)})(x_{\sigma(1)}-x_{\sigma(3)})(x_{\sigma(2)}-x_{\sigma(3)})}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = \frac{(x_1-x_3)(x_1-x_2)(x_3-x_2)}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = -1$	$a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,sigma(3)} = a_{1,1}a_{2,3}a_{3,2} = (2-\lambda)(6)(1) = 6(2-\lambda)$
(213)	$\sigma(1)=2, \sigma(2)=1, \sigma(3)=3$	$\frac{(x_{\sigma(1)}-x_{\sigma(2)})(x_{\sigma(1)}-x_{\sigma(3)})(x_{\sigma(2)}-x_{\sigma(3)})}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = \frac{(x_2-x_1)(x_2-x_3)(x_1-x_3)}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = -1$	$a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,sigma(3)} = a_{1,2}a_{2,1}a_{3,3} = (6)(5)(2-\lambda) = 30(2-\lambda)$
(231)	$\sigma(1)=2, \sigma(2)=3, \sigma(3)=1$	$\frac{(x_{\sigma(1)}-x_{\sigma(2)})(x_{\sigma(1)}-x_{\sigma(3)})(x_{\sigma(2)}-x_{\sigma(3)})}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = \frac{(x_2-x_3)(x_2-x_1)(x_3-x_1)}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = 1$	$a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,sigma(3)} = a_{1,2}a_{2,3}a_{3,1} = (6)(6)(1) = 36$
(312)	$\sigma(1)=3, \sigma(2)=1, \sigma(3)=2$	$\frac{(x_{\sigma(1)}-x_{\sigma(2)})(x_{\sigma(1)}-x_{\sigma(3)})(x_{\sigma(2)}-x_{\sigma(3)})}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = \frac{(x_3-x_1)(x_3-x_2)(x_1-x_2)}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = 1$	$a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,sigma(3)} = a_{1,3}a_{2,1}a_{3,2} = (1)(5)(1) = 5$
(321)	$\sigma(1)=3, \sigma(2)=2, \sigma(3)=1$	$\frac{(x_{\sigma(1)}-x_{\sigma(2)})(x_{\sigma(1)}-x_{\sigma(3)})(x_{\sigma(2)}-x_{\sigma(3)})}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = \frac{(x_3-x_2)(x_3-x_1)(x_2-x_1)}{(x_1-x_2)(x_1-x_3)(x_2-x_3)} = -1$	$a_{1,\sigma(1)}a_{2,\sigma(2)}a_{3,sigma(3)} = a_{1,3}a_{2,2}a_{3,1} = (1)(3-\lambda)(1) = (3-\lambda)$

Hence:

$$
\begin{aligned}
\det\left(\begin{bmatrix} 2-\lambda & 6 & 1\\ 5 & 3-\lambda & 6\\ 1 & 1 & 2-\lambda \end{bmatrix}\right) &= (2-\lambda)(3-\lambda)(2-\lambda) - 6(2-\lambda) - 30(2-\lambda) + 36 + 5 - (3-\lambda)\\
&= -\lambda^3 + 7\lambda^2 + 21\lambda - 22 = 0
\end{aligned}
$$

It can be shown via computation in a computer (or, through some very complex algebraic derivations) that the roots of this equation, rounded, are $\lambda_1 = 0.84, \lambda_2 = -2.89, \lambda_3 = 9.05$. The next step is to solve the system with these roots. First, lets solve it for $\lambda = 0.84$:

$$
\begin{aligned}
(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} &= \boldsymbol{0}\\
\begin{bmatrix} 2-\lambda & 6 & 1\\ 5 & 3-\lambda & 6\\ 1 & 1 & 2-\lambda \end{bmatrix} &= \boldsymbol{0}\\
\begin{bmatrix} 2-0.84 & 6 & 1\\ 5 & 3-0.84 & 6\\ 1 & 1 & 2-0.84 \end{bmatrix} &= \boldsymbol{0}\\
\begin{bmatrix} 1.16 & 6 & 1\\ 5 & 2.16 & 6\\ 1 & 1 & 1.16 \end{bmatrix} &= \begin{bmatrix} 0\\ 0\\ 0 \end{bmatrix}
\end{aligned}
$$

We will perform Gauss-Jordan Elimination:

$$\begin{bmatrix} 1.16 & 6.00 & 1.00 \\ 5.00 & 2.16 & 6.00 \\ 1.00 & 1.00 & 1.16 \end{bmatrix} \rightarrow \begin{bmatrix} 1.00 & 5.17 & 0.86 \\ 5.00 & 2.16 & 6.00 \\ 1.00 & 1.00 & 1.16 \end{bmatrix} \rightarrow \begin{bmatrix} 1.00 & 5.17 & 0.86 \\ 0.00 & -23.70 & 1.69 \\ 1.00 & 1.00 & 1.16 \end{bmatrix}$$
$$\rightarrow \begin{bmatrix} 1.00 & 5.17 & 0.86 \\ 0.00 & -23.70 & 1.69 \\ 0.00 & -4.17 & 0.30 \end{bmatrix} \rightarrow \begin{bmatrix} 1.00 & 5.17 & 0.86 \\ 0.00 & 1.00 & -0.07 \\ 0.00 & -4.17 & 0.30 \end{bmatrix} \rightarrow \begin{bmatrix} 1.00 & 0.00 & 1.23 \\ 0.00 & 1.00 & -0.07 \\ 0.00 & -4.17 & 0.30 \end{bmatrix}$$
$$\rightarrow \begin{bmatrix} 1.00 & 0.00 & 1.23 \\ 0.00 & 1.00 & -0.07 \\ 0.00 & 0.00 & 0.00 \end{bmatrix}$$

At this point in the Gauss-Jordan Elimination, we notice that we have ran out of pivots to use! If we take the matrix and rewrite it in equation form, we see that we have:

$$\begin{aligned} \begin{bmatrix} 1.00 & 0.00 & 1.23 \\ 0.00 & 1.00 & -0.07 \\ 0.00 & 0.00 & 0.00 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &\rightarrow \\ x_1 + 1.23x_3 &= 0 \\ x_2 - 0.07x_3 &= 0 \rightarrow \\ x_1 &= -1.23x_3 \\ x_2 &= .07x_3 \end{aligned}$$

Here, x_3 is referred to as a *free variable*. That is, we can assign it to any number we like. We see that doing so leads us to the following:

$$\begin{aligned} \boldsymbol{x} &= \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= \begin{bmatrix} -1.23x_3 \\ .07x_3 \\ x_3 \end{bmatrix} \\ &= x_3 \begin{bmatrix} -1.23 \\ .07 \\ 1 \end{bmatrix} \end{aligned}$$

Therefore, the solution to the equation $(\boldsymbol{A} - \boldsymbol{I}\lambda)\boldsymbol{x} = \boldsymbol{0}$ for $\lambda = 0.84$ is any vector that is in the vector space spanned by the vector $\begin{bmatrix} -1.23 \\ .07 \\ 1 \end{bmatrix}$. This vector is referred to as an *eigenvector*, and the space that it spans, namely $\left\{ k \begin{bmatrix} -1.23 \\ .07 \\ 1 \end{bmatrix} | k \in \mathbb{R} \right\}$, is referred to as the *eigenspace*. If we continue on in the same procedure for the other two eigen values, we find that the eigenvectors are $\left\{ \begin{bmatrix} -25.14 \\ 20.32 \\ 1.00 \end{bmatrix}, \begin{bmatrix} 3.32 \\ 3.74 \\ 1.00 \end{bmatrix} \right\}$.

5.0.1 Eigenvalue Matrix Decomposition

Now, let $\Lambda = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 0.84 & 0 & 0 \\ 0 & -2.89 & 0 \\ 0 & 0 & 9.05 \end{bmatrix}$ and construct the matrix $\boldsymbol{C}$ with the columns as the eigen vectors associated with the eigen values in the matrix we just formed. That is, put the first eigen vector as the first column of $\boldsymbol{C}$ which was associated with $\lambda_1 = 0.84$, etc. We obtain $\boldsymbol{C} = \begin{bmatrix} -1.23 & -25.14 & 3.32 \\ 0.07 & 20.32 & 3.74 \\ 1 & 1 & 1 \end{bmatrix}$. If we find the inverse of $\boldsymbol{C}$ (do this by hand or using a computer), we obtain (rounded) $\boldsymbol{C}^{-1} = \begin{bmatrix} -0.09 & -0.16 & 0.90 \\ -0.02 & 0.03 & -0.03 \\ 0.11 & 0.13 & 0.13 \end{bmatrix}$ Interestingly enough, if you take the following product, observe what we obtain as a result:

$$
\begin{aligned}
\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^{-1} &= \begin{bmatrix} -1.23 & -25.14 & 3.32 \\ 0.07 & 20.32 & 3.74 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.84 & 0 & 0 \\ 0 & -2.89 & 0 \\ 0 & 0 & 9.05 \end{bmatrix} \begin{bmatrix} -0.09 & -0.16 & 0.90 \\ -0.02 & 0.03 & -0.03 \\ 0.11 & 0.13 & 0.13 \end{bmatrix} \\
&= \begin{bmatrix} -1.03 & 72.65 & 30.05 \\ 0.06 & -58.72 & 33.85 \\ 0.84 & -2.89 & 9.05 \end{bmatrix} \begin{bmatrix} -0.09 & -0.16 & 0.90 \\ -0.02 & 0.03 & -0.03 \\ 0.11 & 0.13 & 0.13 \end{bmatrix} \\
&= \begin{bmatrix} 1.95 & 6.25 & 0.80 \\ 4.89 & 2.63 & 6.22 \\ 0.98 & 0.96 & 2.02 \end{bmatrix} \\
&\approx \begin{bmatrix} 2 & 6 & 1 \\ 5 & 3 & 6 \\ 1 & 1 & 2 \end{bmatrix} \\
&= \boldsymbol{A}
\end{aligned}
$$

We had rounding error in our derivations, but that is ok, because we illustrated a very important rule. The rule we illustrated is that we can *decompose* our square matrix $\boldsymbol{A}$ into three other matrices, namely $\boldsymbol{C}, \boldsymbol{\Lambda}, \boldsymbol{C}^{-1}$. This is called the *eigenvalue decomposition* of a matrix. We will see soon that this decomposition will be **very** useful for us in some of our derivations.

5.0.2 Properties of Matrices and Eigenvalues

Eigenvalues have very beautiful properties, and they give us a lot of information about our matrix $\boldsymbol{A}$. We cover some of these now. In general, if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of matrix A, then $\det(A) = \prod_{i=1}^{n} \lambda_i = \lambda_1 \lambda_2 \ldots \lambda_n$. We can see this from the last example. The determinant of matrix

$A = \begin{bmatrix} 2 & 6 & 1 \\ 5 & 3 & 6 \\ 1 & 1 & 2 \end{bmatrix}$ is -22 (do this as an exercise to practice finding determinants). We can see that the product of the eigenvectors, $\lambda_1 \lambda_2 \lambda_3 = (.84)(-2.89)(9.05) = -22$ (we have rounding error).

Another interesting property is that the *trace* of a matrix $\boldsymbol{A}$ is equal to the sum of the eigenvalues. By definition, the trace of a matrix A is the sum of it's diagonal elements. We see from our previous example that $tr(\boldsymbol{A}) = 2 + 3 + 2 = 7$. We also notice that $.84 - 2.89 + 9.05 = 7$. Thus, in general, we have $tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$.

Eigenvalues also can help us determine the rank of a matrix. Recall from earlier that in order for us to determine the rank of a matrix, we need to conduct Gauss-Jordan Elimination to determine the number of standard basis vectors that result (which is equal to the rank). Alternatively, we could compute the eigenvalues for the matrix $\boldsymbol{A'A}$. Doing so leads us to an interesting property. The number of non-zero eigenvalues of $\boldsymbol{A'A}$ is equal to the rank of $\boldsymbol{A}$ We will state this one without proof or illustration, and I will leave it as an exercise for you to illustrate this.

We also notice that if we would like to find the inverse of a matrix A and we can determine it's eigenvalues, then the inverse is quite easy to find. Notice that using our decomposition we have $A^{-1} = (C\Lambda C^{-1})^{-1} = (C^{-1})^{-1}(\Lambda)^{-1}(C)^{-1} = C\Lambda^{-1}C^{-1}$. Since Λ is a diagonal matrix, it's inverse is very easy to find. In general, the inverse of a diagonal matrix $B = \begin{bmatrix} a_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & a_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & a_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{n-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & a_n \end{bmatrix}$ is just simply the matrix of the reciprocals: $B^{-1} = \begin{bmatrix} \frac{1}{a_1} & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{a_2} & 0 & \dots & 0 & 0 \\ 0 & 0 & \frac{1}{a_3} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{a_{n-1}} & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{a_n} \end{bmatrix}$

5.1 Powers of a Matrix

The eigenvalue decomposition of a matrix provides amazingly stunning results that allows us make computations that are not intuitive, but very useful. We will start with the simple power of a square matrix. Suppose we multiply a matrix $\boldsymbol{A}$ by itself. Then we would have $\boldsymbol{A}^2 = \boldsymbol{AA}$. However, as we take higher powers, say A^n, this computation can be quite burdensome. Thanks to the eigenvalue decomposition, however, we can easily find these powers. Let us reexamine the power of a matrix by recalling that $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^{-1}$:

$$\begin{aligned}\boldsymbol{A}^2 &= (\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^{-1})^2 \\ &= (\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^{-1})(\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^{-1}) \\ &= \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^{-1} \\ &= \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{I}\boldsymbol{\Lambda}\boldsymbol{C}^{-1} \\ &= \boldsymbol{C}\boldsymbol{\Lambda}^2\boldsymbol{C}^{-1}\end{aligned}$$

Now, since $\boldsymbol{\Lambda}$ is a diagonal matrix, it is easy to show that

$$\begin{aligned}\boldsymbol{\Lambda}^2 &= \begin{bmatrix}\lambda_1 & 0 & 0 & \dots & 0 & 0\\ 0 & \lambda_2 & 0 & \dots & 0 & 0\\ 0 & 0 & \lambda_3 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & \lambda_{n-1} & 0\\ 0 & 0 & 0 & \dots & 0 & \lambda_n\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & 0 & \dots & 0 & 0\\ 0 & \lambda_2 & 0 & \dots & 0 & 0\\ 0 & 0 & \lambda_3 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & \lambda_{n-1} & 0\\ 0 & 0 & 0 & \dots & 0 & \lambda_n\end{bmatrix}\\ &= \begin{bmatrix}\lambda_1^2 & 0 & 0 & \dots & 0 & 0\\ 0 & \lambda_2^2 & 0 & \dots & 0 & 0\\ 0 & 0 & \lambda_3^2 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & \lambda_{n-1}^2 & 0\\ 0 & 0 & 0 & \dots & 0 & \lambda_n^2\end{bmatrix}\end{aligned}$$

In general, this leads us to the general result that given any $n \in \mathbb{R}$, we have

$$\begin{aligned}\boldsymbol{\Lambda}^n &= \begin{bmatrix}\lambda_1 & 0 & 0 & \dots & 0 & 0\\ 0 & \lambda_2 & 0 & \dots & 0 & 0\\ 0 & 0 & \lambda_3 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & \lambda_{n-1} & 0\\ 0 & 0 & 0 & \dots & 0 & \lambda_n\end{bmatrix}\cdots\begin{bmatrix}\lambda_1 & 0 & 0 & \dots & 0 & 0\\ 0 & \lambda_2 & 0 & \dots & 0 & 0\\ 0 & 0 & \lambda_3 & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & \lambda_{n-1} & 0\\ 0 & 0 & 0 & \dots & 0 & \lambda_n\end{bmatrix}\\ &= \begin{bmatrix}\lambda_1^n & 0 & 0 & \dots & 0 & 0\\ 0 & \lambda_2^n & 0 & \dots & 0 & 0\\ 0 & 0 & \lambda_3^n & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & \lambda_{n-1}^n & 0\\ 0 & 0 & 0 & \dots & 0 & \lambda_n^n\end{bmatrix}\end{aligned}$$

which means $\boldsymbol{A}^n = \boldsymbol{C}\boldsymbol{\Lambda}^n\boldsymbol{C}^{-1}$ Given this result, then if the eigenvalues are all positive, we should be able to find the "square root" of a matrix, since $\sqrt{\boldsymbol{A}} = \boldsymbol{A}^{\frac{1}{2}} = \boldsymbol{C}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{C}^{-1}$

5.2 Definite Matrices

Suppose that $\boldsymbol{A}$ is a symmetric matrix. If this is the case, then there are special properties of the matrix that we can speak of that will be tremendously helpful to us in determining certain decisions. If it is true that $\forall \boldsymbol{x}, \boldsymbol{x'Ax} > 0$, then the matrix $\boldsymbol{A}$ is said to be *positive definite*. On the other hand, if $\forall \boldsymbol{x}, \boldsymbol{x'Ax} < 0$, then the matrix $\boldsymbol{A}$ is said to be negative definite.

At first, you may think that it would be impossible to prove that a matrix is positive (or negative) definite. After, how can we show that the term $\boldsymbol{x'Ax}$ will always result in a positive number for all vectors $\boldsymbol{x}$? As it turns out, the use of our eigenvalue decomposition is very useful for showing this. First of all, if $\boldsymbol{A}$ is symmetric, then it can be shown that the following is true about the matrix of it's eigenvectors: $\boldsymbol{C'} = \boldsymbol{C}^{-1}$. That is, it's transpose is equal to it's inverse. This is true only if $\boldsymbol{A}$ is a symmetric matrix. Given this, we can expand as such: $\boldsymbol{x'Ax} = \boldsymbol{x'C\Lambda C'x}$. Now, let $\boldsymbol{y} = \boldsymbol{Ax}$. Then $\boldsymbol{y'} = (\boldsymbol{Ax'}) = x'A'$. So we can make the substitution $\boldsymbol{x'Ax} = \boldsymbol{y'\Lambda y}$. However, since $\boldsymbol{\Lambda}$ is a diagonal matrix, it is easy to see that $\boldsymbol{y'\Lambda y} = \sum_{i=1}^{n} y_i^2 \lambda_i$. Hence, we know that regardless of the value for y_i (and hence, x_i), we know that the value of this sum will be positive if all the terms are positive, and all the terms are positive only if the eigenvalues are all positive.