# **NoBS Linear Algebra**
# and Vector Geometry

Jeffrey Wang

January 29, 2018 – May 20, 2018
version 2018.05.20.18:56
*First edition*

# Contents

# Author’s Notes

## 0.1 NoBS

NoBS, short for ”no bull$#!%”, strives for succinct guides that use simple, smaller, relatable concepts to develop a full understanding of overarching concepts.

## 0.2 What NoBS Linear Algebra covers

This guide succinctly and comprehensively covers most topics in an explanatory notes format for a college-level introductory Linear Algebra and Vector Geometry course.

## 0.3 What this study guide does

It explains all the concepts to you in an intuitive way so you understand the course material better.

If you are a mathematics major, it is recommended you read a proof-based book.

If you are not a mathematics major, this study guide is intended to make your life easier.

## 0.4 What this study guide does not do

This study guide is not intended as a replacement for a textbook. This study guide does not teach via proofs; rather, it teaches by concepts. If you are looking for a formalized, proof-based textbook, seek other sources.

## 0.5 Other study resources

NoBS Linear Algebra should by no means be your sole study material. For more conceptual, visual representations, 3Blue1Brown’s *Essence of Linear Algebra* video series on YouTube is highly recommended as a companion to supplement the information you learn from this guide.

Throughout this book, links to relevant *Essence of Linear Algebra* videos will be displayed for you to watch.

## 0.6 Dedication

To all those that helped me in life: this is for you.

Thank you to Edoardo Luna for introducing *Essence of Linear Algebra* to me. Without it, I would not have nearly as much intuitive understanding of linear algebra to write this guide.

## 0.7 Sources

This guide has been inspired by, and in some cases borrows, certain material from the following sources, which are indicated below and will be referenced throughout this guide by parentheses and their names when necessary:

- (Widmer) - Steven Widmer, University of North Texas, my professor from whom I took this course in Spring 2018.
- (Lay) - David C. Lay, Steven R. Lay, Judi J. McDonald, *Linear Algebra and Its Applications* (5th edition)
- Several concepts in Chapter 6 and 7 have text copied and modified from *NoBS Calculus* written by the same author as the author of *NoBS Linear Algebra* (Jeffrey Wang). For this reason, no direct attribution is given inline, but instead given here.
- Some inline citations do not appear here but next to their borrowed content.

Note that this study guide is written along the progression of *Linear Algebra and Its Applications* by Lay, Lay, and McDonald (5th edition). However, Chapter 7 is a supplemental chapter that deviates from the progression of Lay and covers topics such as the cross product.

## 0.8 Copyright and resale

# Chapter 1

# Systems of linear equations

## 1.1 Introduction

In linear algebra, we start with **linear equations**. In high school algebra, we learned about the linear equation $y = mx + b$. However, we will be considering linear equations with multiple variables that are ordered, which can be written in the following form:

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

The coefficient may be any real number ($a \in \mathbb{R}$). However, the variable must be to the $1$ power.

The following are linear equations:

- $x_1 = 5x_2 + 7x_3 - 12$ – Nothing looks out of the ordinary here.
- $x_2 = \sqrt{5}(9 - x_3) + \pi x_1$ – The coefficients can be non-integers and irrational.

However, the following are NOT linear equations:

- $3(x_1)^2 + 5(x_2) = 9$ – We can't have a quadratic.
- $9x_1 + 7x_2x_3 = -45$ – We also can't have variables multiplying by each other.
- $3\sqrt{x_1}+x_2 = 2$ – We furthermore can't have roots of variables, nor any other non-polynomial function.
- $\frac{x_3}{x_1} + x_2 = 3$ – Inverses are not allowed either.
- $\sin x_1 + \cos^3 x_2 - \ln x_3 + e^{-x_4} = -9$ – Obviously no transcendental functions are allowed.

A **system of linear equations** (or a linear system) is a collection of linear equations that share the same variables.

If an ordered list of values are substituted into their respective variables in the system of equations and each equation of the system holds true, then we call this collection of values a **solution**.

Say we have the ordered list of values $(1, 3, 2)$. This would be a solution of the following system of linear equations:

$$\begin{aligned} x_1 + x_2 + x_3 &= 6 \\ x_2 + x_3 &= 5 \\ x_3 &= 2 \end{aligned}$$

because if you substituted the solution for the respective variables, you get

$$\begin{aligned} 1+3+2 &= 6 \\ 3+2 &= 5 \\ 2 &= 2 \end{aligned}$$

which are all valid, true statements.

A system can have more than one solution. We call the collection of all solutions the **solution set**. If two systems have exactly the same solution set, they are actually the same system.

In high school algebra, you learned that the solution of two linear equations was the point at which they intersected; this still holds true, but in linear algebra, we'll be dealing with more generalized cases that might be too complicated to solve graphically.

There are three possibilities (and only three) to the number of solutions a system can have: 0, 1, and $\infty$.

A system is **inconsistent** if it has no solutions and it is **consistent** if it has at least one solution.

- No solutions - inconsistent
- Exactly one solution - consistent
- Infinitely many solutions - consistent

## 1.2 Matrices

### Matrix notation

An **m** $\times$ **n matrix** is a rectangular array (or, ordered listing) of numbers with $m$ rows and $n$ columns, where $m$ and $n$ are both natural numbers. We can use matrices to represent systems in a concise manner.

Given the system of linear equations:

$$\begin{aligned} 3x_1 + 5x_2 &= 9 \\ 7x_2 &= 56 \end{aligned}$$

We can rewrite this in matrix notation:

$$\left[\begin{array}{cc|c} 3 & 5 & 9 \\ 0 & 7 & 56 \end{array}\right]$$

Notice we have kept only the coefficients of the variables. Each row represents one equation and each column represents one variable. The last column is not a column of variables, but instead of the constants. We usually put a vertical line between the second-to-last and last column on the matrix to denote this.

The matrix we've just created contains both the coefficients and the constants. We call this kind of matrix an **augmented matrix**. If we only had the coefficients:

$$\begin{bmatrix} 3 & 5 \\ 0 & 7 \end{bmatrix}$$

then we call this the **coefficient matrix**.

You can apply many of the terminologies and the concepts we've established for systems onto matrices. For instance, a consistent system means a consistent matrix.

## Solving matrices

We're going to apply the same operations we used to solve systems of linear equations back in high school algebra, except this time in matrices.

The three **elementary row operations** for matrices are:

1. **Scaling** – multiplying all of the values on a row by a constant
2. **Interchange** – swapping the positions of two rows
3. **Replacement** – adding two rows together: the row that is being replaced, and a scaled version of another row in the matrix.

These row operations are reversible. In other words, you can undo the effects of them later in the process.

We consider two augmented matrices to be **row equivalent** if there is a sequence of elementary row operations that can transform one matrix into another. Two matrices $A$ and $B$ may appear different, but if you can go from matrix $A$ to $B$ using these elementary row operations, then $A \sim B$. In fact, there are infinitely many row equivalent matrices.

Linking back to the previous section: if two augmented matrices are row equivalent, then they also have the same solution set, which again confirms that they are the same matrix/system.

## Existence and uniqueness (part 1)

Two fundamental questions establish the status of a system/matrix:

1. **Existence** - does a solution exist? (i.e. Is this matrix consistent?)
2. **Uniqueness** - if a solution exists, is this solution unique?

The rest of this chapter is dedicated to the methods used to finding whether a solution exists and is unique. The first half focuses on concepts and techniques that give us existence, and the second half focuses on uniqueness. Then, we'll combine them together in transformations.

For now, these questions are answered by using row operations to change the matrix into something we'll call **triangular** form. (This will be one of many ways to show existence.) This means that all values of the matrix below the "diagonal line" are zero. We can make them zero through row operations.

This matrix is not in triangular form:

$$\left[\begin{array}{ccc|c} 2 & 3 & 4 & 2 \\ 7 & 2 & 3 & 4 \\ 4 & 19 & 9 & 9 \end{array}\right]$$

This matrix is in triangular form and consistent:

$$\left[\begin{array}{ccc|c} 3 & 5 & 7 & 23 \\ 0 & 2 & 3 & 45 \\ 0 & 0 & 9 & 34 \end{array}\right]$$

This matrix, while in triangular form, is NOT consistent:

$$\left[\begin{array}{ccc|c} 7 & 8 & 9 & 10 \\ 0 & 11 & 8 & 5 \\ 0 & 0 & 0 & 3 \end{array}\right]$$

This is because the last row says $0 = 3$, which is false. Therefore, there are no solutions for that matrix, so therefore it is inconsistent.

Consistency = existence. They are the same thing.

## Row reduction and echelon form

Now that we're beginning to delve into solving matrices using row operations, let's develop an algorithm and identify a pattern that will let us find the solution(s) of a matrix.

This pattern was identified in the previous section. We called it by an informal name ("triangular form") but this is actually not its formal name. The pattern is called **echelon form** (or **row echelon form**, pronounced "esh-uh-lohn"). The official definition of echelon form is a matrix that has the following properties:

1. Any rows of all zeros must be below any rows with nonzero numbers.
2. Each "leading entry" of a row (which is the first nonzero entry, officially called a **pivot**) is in a column to the right of the leading entry/pivot of the row above it.
3. All entries in a column below a pivot must be zero.

While echelon form is helpful, we actually want to reduce a matrix down to **reduced echelon form** (or reduced *row* echelon form) to obtain the solutions of the matrix. There are two additional requirements for a matrix to be in reduced echelon form:

4. The pivot in each nonzero row is $1$. (All pivots in the matrix must be $1$.)
5. Each pivot is the only nonzero entry of the column.

(Widmer)

While there are infinitely many echelon forms for a certain matrix, there is only one unique reduced echelon form for a certain matrix.

Note that sometimes, reduced echelon form is not necessary to solve a problem in linear algebra. Ask yourself if you really need reduced echelon form. Usually, the answer is no.

A matrix in echelon form is an **echelon matrix**, and if in reduced echelon form, then it's a **reduced echelon matrix**.

**Pivots** are the leading nonzero entries of the row. A **pivot position** is a position of a certain pivot. A **pivot column** is a column that contains a pivot position.

For instance, in the following matrix:

$$\left[\begin{array}{cccc|c} \boxed{3} & 5 & 8 & 7 & 23 \\ 0 & \boxed{2} & 10 & 3 & 45 \\ 0 & 0 & 0 & \boxed{9} & 34 \end{array}\right]$$

3 (located at row 1, column 1), 2 (at row 2, column 2), and 9 (at row 3, column 4) are the pivots (with their respective pivot positions).

**Gaussian elimination – the row reduction algorithm**

This is the process through which you can reduce matrices. It always works, so you should use it!

Forward phase (echelon form):

1. Start with the leftmost nonzero column. This is a pivot column. The pivot position is the topmost entry of the matrix.

2. Use row operations to make all entries below the pivot position 0.

3. Move to the next column. From the previous pivot, go right one and down one position. Is this number zero? If so, move to the next column, but only move right one (don't move down one) and repeat the process. Otherwise, don't move. This is the next pivot. Repeat the steps above until all entries below pivots are zeros.

Backward phase (reduced echelon form):

1. Begin with the rightmost pivot.

2. Use row operations to make all entries above the pivot position 0.

3. Work your way to the left.

4. After every entry above the pivots are 0, see if any pivots are not 1. If so, use scaling to make it 1.

(borrowed from Widmer)

## Evaluating solutions of a linear system

Reduced echelon form gives us the exact solutions of a linear system. We also found solutions by reducing our matrices to regular echelon form, plugging in values for the variables, and obtaining a solution. These are both valid ways to solve for a linear system. Now, let's interpret what these results mean.

Each column of a matrix corresponds to a variable. In our current nomenclature, we call these variables $x_1, x_2, x_3$, and so on. Accordingly, the first column of the matrix is associated with variable $x_1$ and so on.

If there is a pivot in a column (i.e. a column is a pivot column), then this variable has a value assigned to it. In an augmented matrix in reduced echelon form, the value is in the corresponding row of the last column. This variable is called a **basic variable**.

What if a column does not have a pivot within it? This indicates that the value of this column's variable does not affect the linear system's solution. Therefore, it is called a **free variable**.

The **parametric description** of a solution set leverages the above definitions. It is another way to represent the solution set.

For the following reduced echelon form augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & 0 & 8 & 0 & 5 \\ 0 & 1 & 10 & 0 & 7 \\ 0 & 0 & 0 & 1 & 3 \end{array}\right]$$

its parametric description is

$$\begin{cases} x_1 = 5 \\ x_2 = 7 \\ x_3 \text{ is free} \\ x_4 = 3 \end{cases}$$

### Existence and uniqueness (part 2) of matrices

If the bottom row of an augmented matrix is all zero EXCEPT for the rightmost entry, then the matrix will be inconsistent.

Visually, if the bottom row of an augmented matrix looks like this:

$$\begin{bmatrix} 0 & \dots & 0 & \vert & b \end{bmatrix}$$

then this system is inconsistent.

Furthermore, if a system is consistent and does not contain any free variables, it has exactly one solution. If it contains one or more free variables, it has infinitely many solutions.

## 1.3 Vectors and vector equations

Back when we discussed the different kinds of solutions of a linear system, we assigned each variable of the system $x_1, x_2, \dots, x_n$ its own column. Whatever number is its subscript is whichever column it represented in the matrix.

In fact, we'll now define these individual columns to be called **vectors**. Specifically, we call them **column vectors**: they have $m$ number of rows but only one column. They are matrices of their own right.

**Row vectors** also exist, and have $n$ columns but just one row. They are rarely used, and usually, the term vector refers to column vectors. (From now on, we will refer to column vectors simply as vectors unless otherwise specified.)

We represent the set of all vectors with a certain dimension $n$ by $\mathbb{R}^n$. Note that $n$ is the number of rows a vector has. For instance, the vector $\vec{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ has two rows, so it is a vector within the set $\mathbb{R}^2$.

Two vectors are equal if and only if:

- they have the same dimensions
- the corresponding entries on each vector are equal

### Vector operations

The two fundamental vector operations are vector addition and scalar multiplication.

**Vector addition**: $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 1+3 \\ 2+7 \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$

**Scalar multiplication**: $5 \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \cdot 3 \\ 5 \cdot 2 \end{bmatrix} = \begin{bmatrix} 15 \\ 10 \end{bmatrix}$

Vectors are commutative, associative, and distributive. This means we can use these operations together as well.

### Vectors and matrices

A set of vectors can themselves form a matrix by being columns of a matrix. For instance, if we are given $\vec{v} = \begin{bmatrix}1\\3\\4\end{bmatrix}, \vec{u} = \begin{bmatrix}9\\5\\2\end{bmatrix}, \vec{w} = \begin{bmatrix}7\\8\\6\end{bmatrix}$, and they are all within the set $A$, then they can be represented as a matrix:

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

### Combining vectors

We can use vector operations to form new vectors. In fact, given a set of vectors, we are usually able to use these vector operations to represent infinitely many vectors. For instance, given the vector $\vec{v} = \begin{bmatrix}1\\0\end{bmatrix}$ and $\vec{u} = \begin{bmatrix}1\\1\end{bmatrix}$, we can use just $\vec{v}$ and $\vec{u}$ to create any vector in $\mathbb{R}^2$, the set of all vectors with two real number entries. For instance, if we want to make $\vec{b} = \begin{bmatrix}8\\17\end{bmatrix}$, we can write it as $-9\begin{bmatrix}1\\0\end{bmatrix} + 17\begin{bmatrix}1\\1\end{bmatrix} = -9\vec{v} + 17\vec{u} = \begin{bmatrix}8\\17\end{bmatrix}$.

We can also put these vectors into matrices in the form $\begin{bmatrix}\vec{v} & \vec{u} & | & \vec{b}\end{bmatrix}$:

$$\left[\begin{array}{cc|c}1 & 1 & 8\\0 & 1 & 17\end{array}\right]$$

We can then derive from this matrix these equations and conclusions:

$$x_2 = 17$$

$$x_1 + x_2 = 8$$

$$x_1 = 8 - 17 = -9$$

Therefore, we have just determined that the solution of this matrix will give us the coefficients (or weights) of the vectors in order to give us this vector $\vec{b}$. Using vector operations to represent another vector is called a linear combination, which we will explore in the next section.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 1: Vectors:
`https://www.youtube.com/watch?v=fNk_zzaMoSs`

## 1.4 Linear combinations

A **linear combination** is the use of vector operations to equate a vector $\vec{y}$ to the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \in \mathbb{R}^n$ with coefficients, or **weights** $c_1, c_2, \ldots, c_p$.

$$\vec{y} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_p\vec{v}_p$$

An example of a linear combination is $3\vec{v}_1 - 2\vec{v}_2$.

## Span

The possible linear combinations that can result from a set of vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ is called the **span**.

The span is simply where this linear combination can reach. The span of $\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is simply the line parallel to $\vec{v}_1$. Notice you can't make much else out of that except scaling it. However, if you add $\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ to the mix, you can now create a linear combination to get ANY value in $\mathbb{R}^2$ or the 2D plane. That means these two vectors' possibilities *spans* $\mathbb{R}^2$. For instance, $\begin{bmatrix} 9 \\ 20 \end{bmatrix} = 9\vec{v}_1 + 2\vec{v}_2$.

We say that a vector $\vec{y}$ is in the span of $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, etc. through the notation $\vec{y} \in \text{Span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.

The formal definition of span: if $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \in \mathbb{R}^n$, then the set of all linear combinations of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ is denoted by $\text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ and is called the **subset of $\mathbb{R}^n$ spanned by** $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$:

$$\text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\} = \{c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_p\vec{v}_p : c_1, c_2, \ldots, c_p \text{ are scalars.}\}$$

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 2: Linear combinations, span, and basis vectors:
`https://www.youtube.com/watch?v=k7RM-ot2NWY`
(Note that basis vectors will not be covered until Chapter 2.)

## Existence and uniqueness (part 3) of linear combinations

If there is a solution to the vector equation $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_p\vec{v}_p = \vec{b}$, then there is a solution to its equivalent augmented matrix: $\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_p & \vec{b} \end{bmatrix}$. The solution to the equivalent augmented matrix are in fact the weights of the vector equation. That means we can get $x_1, x_2, \ldots, x_p$ from reducing the equivalent augmented matrix.

**Example**: Let's say we have two vectors $\vec{a}_1 = \begin{bmatrix} 2 \\ 6 \\ 7 \end{bmatrix}$ and $\vec{a}_2 = \begin{bmatrix} -1 \\ -5 \\ -4 \end{bmatrix}$. Can the vector $\vec{b} = \begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix}$ be generated as a a linear combination of $\vec{a}_1, \vec{a}_2$?

Let's consider what this means. Basically, is there a way we can find the following to be true?

$$x_1 \begin{bmatrix} 2 \\ 6 \\ 7 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ -5 \\ -4 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix}$$

We have to add these vectors $\vec{a}_1$ and $\vec{a}_2$ together. Obviously you can't just add them together. So, can we multiply the vectors and then add them together in some form to get $\vec{b} = \begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix}$? Maybe. The question is, what are the coefficients needed for this to happen?

We can solve for the coefficients by putting $\vec{a}_1$ and $\vec{a}_2$ in a matrix as columns and then $\vec{b}$ as the final column. Sound familiar? That's because this is an augmented matrix in the form $\left[\vec{a}_1 \quad \vec{a}_2 \quad | \quad \vec{b}\right]$:

$$\begin{array}{ccc} \vec{a}_1 & \vec{a}_2 & \vec{b} \end{array}$$

$$\left[\begin{array}{cc|c} 2 & -1 & 8 \\ 6 & -5 & 32 \\ 7 & -4 & 24 \end{array}\right]$$

We reduce this matrix and find the solutions from there. Here's the reduced echelon form of the above augmented matrix:

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

The last column contains the answers. The first entry of the last row is $x_1 = 2$ and the second entry of the last row is $x_2 = -4$. Now, let's insert these values back into the linear combination:

$$x_1 \begin{bmatrix} 2 \\ 6 \\ 7 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ -5 \\ -4 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix}$$

$$2 \begin{bmatrix} 2 \\ 6 \\ 7 \end{bmatrix} - 4 \begin{bmatrix} -1 \\ -5 \\ -4 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix}$$

$$\begin{bmatrix} 4 \\ 12 \\ 14 \end{bmatrix} + \begin{bmatrix} 4 \\ 20 \\ 16 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix}$$

$$\begin{bmatrix} 4+4 \\ 12+20 \\ 14+16 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix}$$

$$\begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ 24 \end{bmatrix}$$

Hey, these values $x_1 = 2$ and $x_2 = -4$ indeed hold true. That means yes, we can generate a linear combination from $\vec{a}_1$ and $\vec{a}_2$ for $\vec{b}$.

However, the question is whether $\vec{b}$ can be generated as a linear combination of $\vec{a}_1$ and $\vec{a}_2$. Because the augmented matrix we made from the vectors was consistent, we can conclude that $\vec{b}$ can be generated as a linear combination of $\vec{a}_1$ and $\vec{a}_2$. We did not need to do anything past the reduced echelon form. We just needed to see whether the augmented matrix was consistent.

For this same example, is $\vec{b}$ in the span of $\{\vec{a}_1, \vec{a}_2\}$?

The span refers to the possible combinations of a linear combination. Since $\vec{b}$ is a possible linear combination of $\{\vec{a}_1, \vec{a}_2\}$, then yes, $\vec{b}$ is in the span of $\{\vec{a}_1, \vec{a}_2\}$.

To recap, $\vec{b}$ generated as a linear combination of $\vec{a}_1, \vec{a}_2$ is the same question as whether $\vec{b}$ is in the span of $\vec{a}_1, \vec{a}_2$. They are both proven by seeing whether the augmented matrix of $\begin{bmatrix}\vec{a}_1 & \vec{a}_2 & \vec{b}\end{bmatrix}$ is consistent.

## 1.5 Basic matrix multiplication

To prepare for our next section, we will first jump to a limited scope of matrix multiplication where we are multiplying a matrix by a vector.

We can multiply matrices by each other. This operation multiplies the entries in a certain pattern. We can only multiply two matrices by each other if the first matrix's number of columns is equal to the second matrix's number of rows.

To be clear, let's say we have the first matrix and second matrix:

- The number of columns of the first matrix must equal the number of rows of the second matrix.
- The resulting matrix will have the number of rows that the first matrix has and the number of columns that the second matrix has.

For instance, we can multiply:

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

but not

$$\begin{bmatrix}1 & 8\end{bmatrix}\begin{bmatrix}4\\2\\3\end{bmatrix}$$

So how do we multiply matrices? Let's start with the simple example that worked.

$$\begin{bmatrix}1 & 2 & 3\end{bmatrix}\begin{bmatrix}4\\5\\6\end{bmatrix} = \begin{bmatrix}1\cdot 4 + 2\cdot 5 + 3\cdot 6\end{bmatrix} = \begin{bmatrix}32\end{bmatrix}$$

So we took the first row's entry and multiplied it by the first column's entry, then add to it the second row's entry by the second column's entry, etc.

Now what if we had two rows on the first matrix?

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

We do what we did the first time, except this time we put the second row's products on another row.

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\end{bmatrix}\begin{bmatrix}7\\8\\9\end{bmatrix} = \begin{bmatrix}1\cdot 7+2\cdot 8+3\cdot 9\\4\cdot 7+5\cdot 8+6\cdot 9\end{bmatrix} = \begin{bmatrix}50\\122\end{bmatrix}$$

For now, we will not worry about cases where there are more columns on the second matrix.

## 1.6 The matrix equation $A\vec{x} = \vec{b}$

Recall that a linear combination is the use of some vectors manipulated with vector operations and coefficients (called weights) to equate to another vector.

$$x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_p\vec{a}_p = \vec{b}$$

In the previous section's example, we set the vectors $\vec{a}_1, \vec{a}_2$ to be the following and $\vec{b}$ to be the following:

$$x_1\begin{bmatrix}2\\6\\7\end{bmatrix} + x_2\begin{bmatrix}-1\\-5\\4\end{bmatrix} = \begin{bmatrix}8\\32\\24\end{bmatrix}$$

Instead of writing $x_1, x_2$ as coefficients, we can actually put them in their own vector and use matrix multiplication to achieve the same thing.

$$\begin{bmatrix}2 & -1\\6 & -5\\7 & 4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}8\\32\\24\end{bmatrix}$$

The first matrix is $A$, the coefficient matrix. The vector in the middle is the collection of all weights, and it's called $\vec{x}$. The vector on the right, the linear combination of $A\vec{x}$, is called $\vec{b}$.

This equation's general form is $A\vec{x} = \vec{b}$ and is the **matrix equation**.

In general:

$$A\vec{x} = \begin{bmatrix}\vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_p\end{bmatrix}\begin{bmatrix}x_1\\x_2\\\vdots\\x_p\end{bmatrix} = x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_p\vec{a}_p = \vec{b}$$

This is a great way to represent linear combinations in a simple, precise manner.

### Existence and uniqueness (part 4): Four ways to represent a consistent linear system

So far, we have learned four ways to represent a consistent system (where a solution exists):

1. **Pivots**: The coefficient matrix $A$ has a pivot position in each row.

2. **Linear combination**: $\vec{b}$ is a linear combination of the columns of $A$.

3. **Span**: $\vec{b}$ is in the span of the columns of $A$.

4. **Matrix equation**: If $A\vec{x} = \vec{b}$ has a solution.

## 1.7 Homogeneous linear systems

A **homogeneous linear system** is a special kind of linear system in the form $A\vec{x} = \vec{b}$ where $\vec{b} = \vec{0}$.
In other words, a homogeneous linear system is in the form $A\vec{x} = \vec{0}$.
What gives about them? They have some special properties.

- All homogeneous systems have at least one solution. This solution is called the **trivial solution** and it's when $\vec{x} = \vec{0}$. Of course $A \cdot \vec{0} = \vec{0}$ is true; that's why we call it trivial!

- But the real question is, is there a case where a homogeneous system $A\vec{x} = \vec{0}$ when $\vec{x} \neq \vec{0}$? If such a solution exists, it's called a **nontrivial solution**.

- How do we know if there is a nontrivial solution? This is only possible when $A\vec{x} = \vec{0}$ has a **free variable**. When a free variable is in the system, it allows $\vec{x} \neq 0$ while $A\vec{x} = 0$.

- Therefore, we can say whenever a nontrivial solution exists for a homogeneous system, it has infinitely many solutions. When only the trivial solution exists in a homogeneous system, the system has a unique solution (uniqueness).

### Parametric vector form

When we have free variables, we describe basic variables with relation to the free variables. We group the weights of the free variables to the basic variables through parametric vector form.

If I have:
$x_1 = 3x_2 + 4x_3$
$x_2 = 2x_3$
$x_3$ free

Then we will do necessary replacements to arrive at:
$x_1 = 10x_3$
$x_2 = 2x_3$
$x_3 = x_3$

If we combine them in $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ then we get $\vec{x} = \begin{bmatrix} 10x_3 \\ 2x_3 \\ x_3 \end{bmatrix}$. Factor out the $x_3$ and we get $\vec{x} = x_3 \begin{bmatrix} 10 \\ 2 \\ 1 \end{bmatrix}$. Now, we have represented the solution $\vec{x}$ to the matrix equation $A\vec{x} = \vec{b}$ in parametric vector form $x_3 \begin{bmatrix} 10 \\ 2 \\ 1 \end{bmatrix}$.

### Non-homogeneous linear systems in terms of homogeneous linear systems

What linear system is a non-homogeneous linear system? Basically, it's anything where $A\vec{x} \neq \vec{0}$. So, basically most matrix equations.

What's fascinating is that non-homogeneous linear systems with infinitely many solutions can actually be represented through homogeneous linear systems that have nontrivial solutions. This is because non-homogeneous linear systems are in fact homogeneous linear systems with a *translation* by a certain constant vector $\vec{p}$, where $\vec{x} = \vec{p} + t\vec{v}$. $t\vec{v}$ is where all of the free variables are located.

And guess what? Since this system has infinitely many solutions, the translation is taken out when $A\vec{x} = \vec{0}$. That means, for a system with infinitely many solutions (i.e. one that can be written in parametric vector form), the nonconstant vector $\vec{v}$ (the one with the free variables), without the constant vector $\vec{p}$, is a possible solution!

## 1.8 Linear independence

Now, within our linear combination, could we have duplicate vectors? Yes. Graphically speaking, if two vectors are multiples of each other, they would be parallel. The whole point of linear combinations is to make new vectors. What use is two vectors that are multiples of each other? That's redundant.

This is where the concept of **linear independence** comes in. Are there any redundant variables in the linear combination? Or, are there any redundant rows in a reduced matrix? If so, there are redundancies and therefore we say the combination is linearly dependent. If nothing is redundant, then it's linearly independent.

Mathematically, however, we have a formal way of defining linear independence. If the linear combination $x_1\vec{v}_1 + x_2\vec{v}_2 + \ldots x_p\vec{v}_p = \vec{0}$ has *only the trivial solution*, then it's linearly independent. Otherwise, if there exists a nontrivial solution, it must be linearly dependent.

Why is that? If there are redundant vectors, then there exists weights for them to cancel each other out. Let's say we have two vectors $\vec{v}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}, \vec{v}_2 = \begin{bmatrix}2\\4\\6\end{bmatrix}$. Clearly, $\vec{v}_2$ is a multiple of $\vec{v}_1$. So we can say $2\vec{v}_1 = \vec{v}_2$. We can rearrange this to get $2\vec{v}_1 - 1\vec{v}_2 = \vec{0}$. Therefore, per our formal definition, $\{\vec{v}_1, \vec{v}_2\}$ is linearly dependent.

If a set of vectors are linearly dependent, then there are infinitely many ways to make a combination. Therefore, there would be infinitely many solutions for $A\vec{x} = \vec{b}$. However, if a set of vectors are linearly independent, then there would only be at most one way to form the solution for $A\vec{x} = \vec{b}$.

This leads us to a conceptual connection that is very important: if a set of vectors are linearly independent, at most one solution can be formed. This means linear independence implies uniqueness of solution.

### Linear independence of a set of one vector

A set with one vector contained within is linearly independent unless this vector is the zero vector. The zero vector is linearly dependent but has only the trivial solution.

$\begin{bmatrix}1\\2\\3\end{bmatrix}$ is linearly independent. $\begin{bmatrix}0\\0\end{bmatrix}$ is linearly dependent and has only the trivial solution.

### Linear independence of a set of two vectors

Given two vectors $\vec{v}, \vec{u}$, then this set is linearly independent as long as $\vec{v}$ is not in the span of $\vec{u}$. It could also be the other way around, but the same result still holds true.

### Linear independence of a set of multiple vectors

We have several rules for determining linear dependence of a set of multiple vectors. If one rule holds true, then the set is linearly dependent.

- If at least one of the vectors in the set is a linear combination of some other vectors in the set, then the set is linearly dependent. (Note that not all have to be a linear combination of each other for this to hold true. Just one needs to be a linear combination of others.)
- If there are more columns than rows in a coefficient matrix, that system is linearly dependent. Why? Recall that the number of entries a vector has is equivalent to the number of dimensions it contains. If there are more vectors than dimensions, the vectors will be redundant within these dimensions. For instance, $\vec{v}_1 = \begin{bmatrix}1\\2\end{bmatrix}$ only spans a line in $\mathbb{R}^2$. If we add $\vec{v}_2 = \begin{bmatrix}0\\1\end{bmatrix}$ to the set, we can now span all of $\mathbb{R}^2$ because $\vec{v}_1$ is not parallel to $\vec{v}_2$. But if we add $\vec{v}_3 = \begin{bmatrix}-1\\0\end{bmatrix}$, it doesn't add anything new to the span. We're still in $\mathbb{R}^2$ but with a redundant vector. Furthermore, $\vec{v}_1 = 2\vec{v}_2 - 1\vec{v}_3$, so clearly that would be a linearly dependent set of vectors.
- If a set contains the zero vector $\vec{0}$, then the set is linearly dependent. It's easy to create the zero vector with any vector; just make the weight 0, lol.

## 1.9 Linear transformations

We have studied the matrix equation $A\vec{x} = \vec{b}$. This equation is kind of like $y = mx + b$, which is just a relation. These relations can be represented as functions in regular algebra as $f(x) = mx + b$. Now, we will transition to the equivalent of functions in linear algebra, called linear transformations.

A **linear transformation**, also called a **function** or **mapping**, is where we take the vector $\vec{x}$ and manipulate it to get the result $T(\vec{x})$. This is the same concept as turning $x$ into $f(x)$. How do we manipulate it? We have several options, but for now, we will stick with the coefficient matrix $A$. Using the coefficient matrix for manipulation means we are using a special form of linear transformation called a **matrix transformation**. (Note of clarification: If we talk about $A$, then we are referring to a matrix transformation. Otherwise, we are referring to a general linear transformation.)

For the coefficient matrix $A$ with size $m \times n$ (where $m$ is the number of rows and $n$ is the number of columns), the number of columns $n$ it has is the **domain** ($\mathbb{R}^n$) while the number of rows $m$ it has is the **codomain** ($\mathbb{R}^m$). The codomain is different from the range.

Specifically, the transformed vector $\vec{x}$ itself, $T(\vec{x})$, is called the **image** of $\vec{x}$ under $T$. Note that the image is like the $\vec{b}$ in $A\vec{x} = \vec{b}$, and in fact, $T(\vec{x}) = \vec{b}$.

Something to keep in mind: matrix equations can have no solutions, one solution, or infinitely many solutions. If no $\vec{x}$ can form an image, it means there's no solutions. If only one $\vec{x}$ forms an image, then there is only one solution. But if we can find more than one $\vec{x}$ (i.e. there's probably a basic variable) then we have infinitely many solutions and not a unique solution.

Why isn't the codomain the range? Well, the range is the places where the solution to this matrix exists. The image is the actual vector $\vec{x}$ transformed, while the codomain is just the new domain where the image resides after this transformation has shifted. Not all of the codomain is in the range of the transformation. Remember, this is linear algebra, and we're shifting the dimensions here with these transformations, so we have to differentiate between which dimension we're in before and after. If $A\vec{x} = \vec{b}$, then $\vec{b}$ is in the range of the transformation $\vec{x} \mapsto A\vec{x}$.

By the way, the notation to represent a transformation's domain to codomain is:

$$T : \mathbb{R}^n \mapsto \mathbb{R}^m$$

Note that $\mapsto$ is read as "maps to", so that's why we'll call use the terms transformations and mappings interchangeably. It's just natural!

## Properties of linear transformations

It's important to note that matrix transformations are a specific kind of linear transformation, so you can apply these properties to matrix transformations, but they don't solely apply to matrix transformations. Also, don't forget later on that these properties hold true for other kinds of linear transformations. With that in mind, the following properties apply to all linear transformations:

- $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$, for $\vec{u}, \vec{v} \in$ Domain$\{T\}$
- $T(c\vec{u}) = cT(\vec{u})$, for $c \in \mathbb{R}, \vec{u} \in$ Domain$\{T\}$
- $T(\vec{0}) = \vec{0}$ - oh yeah, that's right. Those "linear functions" $y = mx + b$ with $b \neq 0$ you learned about in Algebra I? They're not actually "linear functions" in linear algebra.

So why do these properties even matter? In problems, you are given $T(\vec{u})$ but not $\vec{u}$ itself (i.e. $\vec{b}$ and not $\vec{x}$ in $A\vec{x} = \vec{b}$), so you must isolate $T(\vec{u})$ from these operations to do anything meaningful with it.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 3: Linear transformations and matrices:
`https://www.youtube.com/watch?v=kYB8IZa5AuE`

## 1.10 The matrix of a linear transformation

In the previous section, we mainly dealt with $\vec{x}$ in the matrix equation $A\vec{x} = \vec{b}$. Why is $A\vec{x} = \vec{b}$ a transformation? Because, every linear transformation that satisfies $\mathbb{R}^m \mapsto \mathbb{R}^n$ is a matrix transformation. (Note: This means only linear transformations with $\mathbb{R}^m \mapsto \mathbb{R}^n$ is a matrix transformation. What was said before about not all linear transformations being matrix transformations *still holds true*.) And because they are matrix transformations, we can use the form $\vec{x} \mapsto A\vec{x}$ to represent our transformations. In this section, we will shift the focus from $\vec{x}$ to $A$.

Now, when we map $\vec{x} \mapsto A\vec{x}$, we have technically do have a coefficient in the front of the lone $\vec{x}$. So really, we should say $I\vec{x} \mapsto A\vec{x}$. What matrix $I$ always satisfies $\vec{x} = I\vec{x}$? The **identity matrix** $I_n$, which is a matrix size $n \times n$ with all zeroes except for ones down the diagonal.

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \text{etc.}$$

Why does this matter? Well, in $I\vec{x} \mapsto A\vec{x}$, $\vec{x}$ is on both sides. So really, the transformation represents a change from $I$, the identity matrix, to $A$. We shall name $A$ the **standard matrix**.

The columns of the identity matrix, which we shall call $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$, correlate to the columns of the standard matrix, which are just transformations of the identity matrix's columns (i.e. $A = [T(\vec{e}_1)\, T(\vec{e}_2)\, \dots\, T(\vec{e}_n)]$).

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Footnote, Nonsquare matrices as transformations between dimensions:
`https://www.youtube.com/watch?v=v8VSDg_WQlA`

### Geometric transformations in $\mathbb{R}^2$

As an aside, some of these standard matrices perform geometric transformations in one of these classes:

- **Reflections**
- **Contractions and expansions**
- **Shears** - keeping one component constant while multiplying another component by a scalar
- **Projections** - "shadows" onto a lower dimension (in this case, $\mathbb{R}^1$)

### Another vector representation format

We can write vectors another way: $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = (x_1, x_2)$. This comes in handy when we want to write a transformation as a tuple, like we do for a function: $T(x_1, x_2) = (x_1 + x_2, 3x_2) = \begin{bmatrix} x_1 + x_2 \\ 3x_2 \end{bmatrix}$.

$$T(x_1, x_2) = (x_1 + x_2, 3x_2)$$

is the same thing as

$$T(\vec{x}) = A\vec{x} = \vec{b}$$

is the same thing as

$$T(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}) = \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ 3x_2 \end{bmatrix}$$

(Note that a tuple is the generic term for an ordered pair with more than two values. For instance, $(1, 3, 4)$ is a 3-tuple. And technically, $(2, 5)$ is an ordered pair, also called a 2-tuple.)

## Existence and uniqueness (part 5): Onto and one-to-one

Two ways we can characterize transformations are whether they are **onto mappings**, **one-to-one mappings**, both, or neither. They help us determine the existence and uniqueness of solution(s) for transformations.

Here are each of their formal definitions:

A mapping $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ is **onto** $\mathbb{R}^m$ if for each $\vec{b}$ in the codomain $\mathbb{R}^m$, there exists a $\vec{x}$ in the domain $\mathbb{R}^n$ so that $T(\vec{x}) = \vec{b}$.

A mapping $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ is **one-to-one** if each $\vec{b}$ in the codomain $\mathbb{R}^m$ is the image (or result) of at most one $\vec{x}$ in the domain $\mathbb{R}^n$.

These definitions connect transformations back to the matrix equation and existence and uniqueness.

If a mapping is onto, that means every value in its domain maps to something else on the new codomain. ("No Value Left Behind!") That means there shall always exist a solution $\vec{x}$ for $T(\vec{x}) = \vec{b}$, right? Well, $T(\vec{x}) = A\vec{x}$. And if $T(\vec{x})$ always has a solution, so does $A\vec{x}$. Remember what $A\vec{x}$ always having a solution means? The columns of its matrix $A$ span the codomain $\mathbb{R}^m$.

If a mapping $T(\vec{x}) = \vec{b}$ is one-to-one, which means it has one unique solution, then similarly, $A\vec{x} = \vec{b}$ must also have a unique solution. And that means the columns of $A$ are linearly independent, which is true because that's also a way to say a matrix's columns form a unique solution.

Now, practically speaking, we need to use a few evaluating techniques to determine whether transformations are onto, one-to-one, both, or neither.

Onto:

- Is there a pivot in each row? If yes, it IS onto.
- Do the columns of $A$ span the codomain $\mathbb{R}^m$? If yes, it IS onto.
- Is $m < n$ (i.e. the number of rows is less than the number of columns)? If yes, it is NOT onto.

One-to-one:

- Is there a pivot in each column? If yes, it IS onto.

- Are the columns of $A$ linearly independent? If yes, it IS one-to-one.
- Is $m > n$ (i.e. the number of rows is greater than the number of columns)? If yes, it is NOT one-to-one.
- Are there any free variables? If yes, it is NOT one-to-one.

In summary:

- **Onto** mapping (also called **surjective**): does a solution exist? If yes, then the transformation is onto. More precisely, $T$ maps the domain $\mathbb{R}^m$ onto the codomain $\mathbb{R}^n$.
- **One-to-one** mapping (also called **injective**): are there infinitely many solutions? (i.e. are there any free variables in the equation?) If yes, then the transformation is NOT one-to-one.

If a transformation is one-to-one, then the columns of $A$ will be linearly independent.
If there is a pivot in every row of a standard matrix $A$, then $T$ is onto.
If there is a pivot in every column of a standard matrix $A$, then $T$ is one-to-one.

## 1.11 Summary of Chapter 1: Ways to represent existence and uniqueness

The following concepts are related to or show **existence** of a solution:

- If a pivot exists in every row of a matrix.
- If a system is consistent.
- If weights exist for a linear combination formed from the columns of the coefficient matrix to equate the solution.
- If the solution is in the span of a set of vectors, usually the set of vectors in a linear combination or the columns of the coefficient matrix.
- If there exists a solution for $A\vec{x} = \vec{b}$.
- If a transformation is onto or surjective.

The following concepts are related to or show **uniqueness** of a solution:

- If a pivot exists in every column of a matrix.
- If there are no free variables (solely basic variables) in a solution.
- If a homogeneous linear system $A\vec{x} = \vec{0}$ has only the trivial solution where $\vec{x} = 0$ is the only solution to $A\vec{x} = \vec{0}$.
- If a solution can be expressed in parametric vector form.
- If a set of vectors is linearly independent.
- If a transformation is one-to-one or injective.

# Chapter 2

# Matrix algebra

Now, we'll begin to formally explore the algebra behind matrices beyond just their reductions and transformations. Let's start by formally defining what a matrix and its components are.

## 2.1 Matrix operations

Let $A$ be an $m \times n$ matrix. The scalar entry of the $i$th row and the $j$th column of $A$ is denoted $a_{ij}$ and is called the $(i, j)$ entry. (Notice this is lowercase "a" because an entry will be the lowercase letter of the matrix letter.) The columns are denoted as $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n$. (Widmer)

$$A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \ldots & \vec{a}_n \end{bmatrix} = \begin{bmatrix} \vec{a}_{11} & \ldots & \vec{a}_{1j} & \ldots & \vec{a}_{1n} \\ \vdots & \ddots & \vdots & & \vdots \\ \vec{a}_{i1} & \ldots & \vec{a}_{ij} & \ldots & \vec{a}_{in} \\ \vdots & & \vdots & \ddots & \vdots \\ \vec{a}_{m1} & \ldots & \vec{a}_{mj} & \ldots & \vec{a}_{mn} \end{bmatrix}$$

$$\vec{a}_j = \begin{bmatrix} \vec{a}_{1j} \\ \vec{a}_{2j} \\ \vdots \\ \vec{a}_{mj} \end{bmatrix}$$

The **diagonal entries** of $A$ are the ones whose indices $i$ equal $j$ ($i = j$, i.e. $\vec{a}_{11}, \vec{a}_{22}, \ldots, \vec{a}_{i=j}$).

A **diagonal matrix** is a matrix whose non-diagonal entries are all zero. The diagonal entries can be either nonzero or zero. The identity matrix $I_n$ is a diagonal matrix.

If a matrix's entries are all zero, it's called the **zero matrix** and is denoted by $0$ just like the scalar. Its size fluctuates based on context and is implied. Note that the zero matrix is also a diagonal matrix.

The rest of this section will explore three matrix operations: addition, scalar multiplication, and matrix multiplication.

### Addition and scalar multiplication

The first two matrix operations are addition and scalar multiplication.

Addition of two matrices is only possible if they share the same size. (If they are not the same size, the operation is undefined.) Simply add up their corresponding entries to form a new, same-sized matrix.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} + \begin{bmatrix} 3 & 4 & 5 \\ 1 & 2 & 6 \end{bmatrix} = \begin{bmatrix} 4 & 6 & 8 \\ 5 & 7 & 12 \end{bmatrix}$$

Scalar multiplication is simply multiplying all of the entries of a matrix by a scalar value.

$$3 \begin{bmatrix} 2 & 4 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 12 & 9 \\ 9 & 6 & 3 \end{bmatrix}$$

Now that we've established these simple concepts, let's start talking about their algebraic properties. Central to our understanding of these operations are:

1. what criteria makes matrices equal to each other (equivalence)

2. what algebraic properties these matrix operations have (commutative, associative, and distributive)

So first, let's talk about equivalence. Two matrices are **equivalent** if they 1. share the same size and 2. each of their corresponding entries are equal to each other. That's easy.

Matrix addition and scalar multiplication have commutative, associative, and distributive properties.

For example, let $A, B, C$ be matrices of the same size, and let $x, y$ be scalars.

- **Associative**:
  $(A + B) + C = A + (B + C)$ - order in which you perform matrix addition does not matter.
  $x(yA) = (xy)A$ - same for multiplication

- **Commutative**:
  $A + B = B + A$ - the order of the operands (A and B) around the operator (+) does not matter
  $xA = Ax$ - same thing

- **Distributive**:
  $x(A + B) = xA + xB$
  $(x + y)A = xA + yA$

## Matrix multiplication

The last matrix operation we will learn about is matrix multiplication. Instead of multiplying a scalar by a matrix, we multiply two matrices by each other.

Why is this significant? When we make multiple linear transformations in a row, like:

$$\vec{x} \mapsto B\vec{x} \mapsto A(B\vec{x})$$

then this is really equivalent to

$$\vec{x} \mapsto AB\vec{x}$$

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 4: Matrix multiplication as composition:
`https://www.youtube.com/watch?v=XkY2DOUCWMU`

**Matrix multiplication procedure expansion**

We already learned some basics of matrix multiplication in section 1.5. Now, we will expand it from multiplying a matrix by a vector to a matrix by another matrix!

The result of a matrix multiplied by a column vector is actually always going to be a column vector. Why? Because the number of columns a product matrix has (the result of the multiplication) is actually always going to be the number of columns the second matrix has. A column vector is a matrix with one column. Therefore, the result was always a column matrix.

However, now we simply modify the multiplication procedure by doing the same thing as before, only now we also multiply all of the first matrix by the second column of the second matrix and put the values in a new column instead of adding them onto the first column's values.

**Matrix multiplication: requirements and properties**

Now, in order for matrix multiplication to occur, first the number of columns the first matrix has must equal the number of rows the second matrix has. The resulting product matrix will have the number of rows the first has and the number of columns that the second matrix has.

Let's say we multiply a $3 \times 5$ matrix $A$ by a $5 \times 7$ one $B$. $5 = 5$, so we can move on. The product matrix, $AB$, will have a size of $3 \times 7$.

But what if we multiplied it in the reverse order, $BA$? $5 \times 7$ multiplied by $3 \times 5$? Well, $7 \neq 3$ and therefore we actually cannot multiply in the order $BA$.

This is very important. The order in which matrices are multiplied do matter. Therefore, matrix multiplication is not communitative. Generally speaking, $AB \neq BA$.

Here are the properties of matrix multiplication:

- **Associative**:
  $(AB)C = A(BC)$ - order in which you perform matrix multiplication does not matter.
- **Distributive**:
  $(A + B)C = AC + BC$
  $A(B + C) = AB + AC$
  $x(AB) = (xA)B = A(xB)$
- **Identity**: $I_m A = A = AI_n$
- ~~**Commutative**:~~ Lol jk it's not commutative remember? In general, $AB \neq BA$.

Some things to watch out for because of the special properties of matrix multiplication!

1. In general, $AB \neq BA$.
2. Cancelation laws do not apply to matrix multiplication. If $AB = AC$, this does not mean $B = C$ is necessarily true. It might be, rarely, but that's not the rule in general.

3. If $AB = 0$, that does not mean $A$ is 0 nor is $B$ 0. They could both be nonzero.

One final thing. You can multiply matrices to the power of themselves (**matrix exponentiation**) if they have the same number of rows and columns. In that case, $A^k = \prod_1^k A$ (meaning $A$ multiplied $k$ times), where $k \in \mathbb{Z}^+$.

## 2.2 Transpose and inverse matrices

### Transpose

Sometimes, it may be convenient to switch the rows with the columns by "flipping" the matrix over its main diagonal. This means the first row will become the first column and vice versa. This flipping is called a **transpose**.

Formal definition (Widmer): If $A$ is an $m \times n$ matrix, then the transpose of $A$ is an $n \times m$ matrix denoted $A^T$ whose columns are formed from the corresponding rows of $A$.

$$\text{row}_i(A) = \text{col}_i(A^T)$$

So basically, we are making the original matrix's rows into the transpose's columns. Or, equivalently, we are making the original matrix's columns into the transpose's rows.

#### Examples of transposes

The following are borrowed from the Wikipedia article on transposes.

$$\begin{bmatrix} 1 & 2 \end{bmatrix}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

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

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

#### Properties of transposes

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(rA)^T = r(A^T)$, where $r$ is a scalar
- $(AB)^T = B^T A^T$
  Note that the order of the matrices have been reversed. This is because the number of rows and columns in the transpose matrix have been swapped, so the order must swap too.

## Inverse matrices

Let's say we have a square matrix $A$, size $n \times n$. When we multiply $A$ by another matrix, it becomes the identity matrix $I_n$. What is this other matrix? It's the **inverse matrix**. If a matrix $A$ and another $B$ produce the following results:

$$AB = BA = I_n$$

then they are inverses of each other/they are **invertible**. The inverse of matrix $A$ is denoted $A^{-1}$.

Inverses are not always possible. For one, if a matrix is not square, then it cannot have an inverse. But in certain scenarios, a square matrix may not have an inverse. These matrices are called **singular** or **degenerate** and will be further discussed later.

Also, both $AB$ and $BA$ must be true for $A$ to be the inverse matrix of $B$ and vice versa. Since matrix multiplication is not commutative, there are instances where $AB$ might produce the identity matrix but not $BA$. In those instances, $A$ and $B$ are not considered inverses of each other.

### Determining the inverse of a $2 \times 2$ matrix

There exists a computational formula to find the inverse of a $2 \times 2$ matrix. While it does not apply to any other sizes, we will later find that computing larger inverses can be done by breaking matrices up into multiple $2 \times 2$ matrices, and then using this formula.

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

In the matrix, $a$ and $d$ have swapped positions while $b$ and $c$ have swapped signs. Also, the denominator is called something special that will be of great significance in the coming chapters—it's called the **determinant**, and for a $2 \times 2$ matrix, it is calculated by $ad - bc$. Notice that when $\det A = ad - bc = 0$, the denominator would be zero and that would make the inverse impossible to calculate. This means that a matrix has no inverse (i.e. it is singular) if and only if its determinant is equal to zero (i.e. $\det A = 0$).

A matrix $A$ is only invertible when $\det A = ad - bc \neq 0$.

### Inverses and matrix equations

In the last chapter, we solved the matrix equation $A\vec{x} = \vec{b}$ through various methods mainly boiling down to Gaussian row reduction. But now, we can introduce another, more intuitive way to solve for it. Why can't we "cancel out" the $A$ on the left side and move it to the right side, in a way?

With inverses, we can! If $A$ is an invertible $n \times n$ matrix, then $A\vec{x} = \vec{b}$ has the unique solution $\vec{x} = A^{-1}\vec{b}$. Note the order of the latter part. It is $A^{-1}\vec{b}$, not ~~$\vec{b}A^{-1}$~~. (It has been struck out because it is wrong and we don't want you getting the wrong idea here.)

We can write $A\vec{x} = \vec{b}$ as $\vec{x} = A^{-1}\vec{b}$ if $A$ is invertible.

This is kind of like division in scalar algebra (regular algebra you did in high school). Remember though, that division is not a defined operation in matrix algebra, so this is one of the "analogs" we can come up with.

**Properties of invertible matrices**

- If $A$ is an invertible matrix, then $A^{-1}$ is invertible and $(A^{-1})^{-1} = A$.
- If $A$ and $B$ are invertible $n \times n$ matrices, then $AB$ is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.
- If $A$ is an invertible matrix, $A^T$ is invertible and $(A^T)^{-1} = (A^{-1})^T$.
  This property is derived from an inverse's inherent commutative property.

## Inverse matrices and row equivalence

This subsection will explore the connection between matrix multiplication and row equivalence.

Up to this point, we have identified that invertible matrices have the following property: $AA^{-1} = A^{-1}A = I$ is always true. That is, an invertible matrix multiplied by its inverse matrix is always equal to the identity matrix, regardless of the order. This means we can find the inverse matrix using matrix multiplication.

However, we can also use the three elementary row operations (replacement, swap, and scale) to substitute for a series of matrix multiplications like $A^{-1}A = I$ (which are really transformations). If we can use these row operations to get from $A$ to $I$, that means $A$ is row equivalent to the identity matrix. We learned about row equivalence back in the first half of Chapter 1.

A square matrix $A$ is row equivalent to the identity matrix if and only if $A$ is invertible. These two things mean the same thing!

Using this profound connection, we can also start considering each elementary row operation as a matrix that is multiplied onto the matrix $A$. We call these matrices elementary matrices.

**Elementary matrices**

An **elementary matrix** is a matrix obtained by performing a single elementary row operation on the identity matrix.

For instance, the following are elementary matrices:

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$$

$E_1$ is a replacement made of the identity matrix $I_3$. Row 3 was replaced with -4 times Row 1 plus Row 3.

$$E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

In $E_2$, a swap was made. Rows 1 and 2 were swapped with each other.

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

In $E_3$, scaling was done on row 3 by a factor of 9.

Since elementary matrices are row equivalent with identity matrices (by definition), they must also be invertible, per the profound connection we just made. Therefore, $E_1E_1^{-1} = E_2E_2^{-1} = E_3E_3^{-1} = I_3$.

Why are elementary matrices important? Because the matrix multiplication $E_1A$ is the same as the row operation "$A : R_3 \to R_3 + (-4)R_1$". This is how we represent row operations through matrix operations.

**Row equivalence of elementary matrices**

Interestingly enough, the same sequence of row operations/multiplication made on $A$ to get to $I$ will get $I$ to $A^-1$. More precisely, this means $A \to I$ is row equivalent to $I \to A^{-1}$.

This means if you have an augmented matrix $[A|I]$ and you perform a sequence of elementary row operations on it to get $A$ on the left turned into $I$, you will simultaneously turn $I$ on the right to $A^{-1}$, which is $A$'s inverse.

$$[A|I] \to [I|A^{-1}]$$

Then, if we consider this sequence of elementary row operations as a sequence of matrices multiplying by each other, we can use elementary matrices to represent a sequence of row operations as: $E_p \dots E_2E_1$. Notice that the last row operation $E_p$ is on the left and the first row operation $E_1$ is on the right. That is because order matters in matrix multiplication and the invertible matrix $A$ will be to the right of $E_1$.

We can conclude that a sequence of elementary matrices $E_p \dots E_2E_1$ that allows $E_p \dots E_2E_1A = I$ to be true will also make $E_p \dots E_2E_1I = A^{-1}$. Therefore:

$$E_p \dots E_2E_1 = A^{-1}$$

## 2.3 Characteristics of invertible matrices: the Invertible Matrix Theorem

As we have seen, invertible matrices have so many properties. What's important to conclude from all of these properties is that a matrix being invertible makes it have a **unique solution** in our all-important fundamental questions of linear algebra: existence and uniqueness.

If $A$ is a square matrix (that's very important!), and if any of the following statements are true, it follows that the rest must also be true. We usually use statement 3 to see whether a matrix is invertible or not.

1. $A$ is an invertible matrix.
2. $A$ is row equivalent to $I_n$, the $n \times n$ identity matrix.
3. $A$ has $n$ pivot positions. (i.e. there is a pivot in every row and every column).
4. The homogeneous matrix equation $A\vec{x} = \vec{0}$ has only the trivial solution.
5. The columns of $A$ form a linearly independent set.
6. The linear transformation $x \mapsto Ax$ is one-to-one.
7. The equation $A\vec{x} = \vec{b}$ has exactly one solution $\forall \vec{b} \in \mathbb{R}^n$. ($\forall$ means "for all".)
8. The columns of $A$ span $\mathbb{R}^n$.

9. The linear transformation $x \mapsto Ax$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

10. There is an $n \times n$ matrix $C$ so that $CA = I_n$.

11. There is an $n \times n$ matrix $D$ so that $AD = I_n$. $C = D$ always.

12. $A^T$ is an invertible matrix.

To reiterate, this theorem only applies to square matrices! Regular non-square rectangular matrices do NOT obey this theorem.

(This theorem will be expanded with new statements in later sections.)

### Invertible linear transformations

In algebra, we can say that a function $g(x)$ is the inverse of $f(x)$ if $g(f(x)) = (g \circ f)(x) = x$. We can say the same thing for transformations in linear algebra.

A linear transformation $T : \mathbb{R}^n \mapsto \mathbb{R}^n$ is invertible if $\exists$ (”there exists”) a function $S : \mathbb{R}^n \mapsto \mathbb{R}^n$ such that

$$S(T(\vec{x})) = \vec{x} \text{ and } T(S(\vec{x})) = \vec{x}\,, \forall \vec{x} \in \mathbb{R}^n$$

From this, we can say that if $T(\vec{x}) = A\vec{x}$, then $S(\vec{x}) = A^{-1}\vec{x}$.

Also, $T$ and $S$ are one-to-one transformations because they are invertible.

## 2.4 Partitioned matrices

(This section will be briefly covered.)

It is possible to partition matrices vertically and horizontally to create smaller partitions. This is desirable to computationally optimize matrix operations for sparse matrices, which are mainly comprised of zeroes but contain a few data points at places. Instead of computing everything one by one, areas of zeroes can be skipped.

## 2.5 Subspaces

In Chapter 1, we saw that linear combinations can produce combinations over a wide range of space. However, not all linear combinations span the entirety of $\mathbb{R}^n$. For instance, $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a linear combination in $\mathbb{R}^3$ but its span only forms a 2D plane in 3D space. The 2D plane is an example of a subspace.

The formal definition of a subspace requires knowing a property called *closure* of an operation. An operation is something like addition, multiplication, subtraction, division, etc. These operations are performed upon numbers that are members of certain sets like the reals, for instance, or sometimes just integers. If we add integers, we're going to get more integers. That is, the addition operation obeys closure for integers, or more correctly, the set of all integers $\mathbb{Z}$ has closure under addition.

The same is true for multiplication. If you multiply any two integers, you will still remain in the set of all integers because the result will still be some integer.

Can we say the same for division? No. Dividing 1 by 2 will yield 0.5, which is not within the set of all integers. Therefore, the set of all integers does NOT have closure under division.

Now that we have defined closure, let's continue to defining what a **subspace** is; it's a set $H$ that is a subset of $\mathbb{R}^n$ with two important characteristics: it contains the zero vector and is closed under addition and scalar multiplication.

In other words, for a set to be a subspace, the following must be true:

- The zero vector is in $H$.
- For each $\vec{u}, \vec{v} \in H$, $\vec{u} + \vec{v} \in H$ as well.
- For each $\vec{u}, \vec{v} \in H$, $c\vec{u} \in H$ too.

(To be clear, vectors $\vec{u}$ and $\vec{v}$ are members of the set $H$.)

To prove whether a set defined by a span is a subspace (i.e. $H = \text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$), use linear combinations to prove the definitions true. Do the adding and scalar multiplication operations still hold true?

An important edge case to consider is whether the line $L$, which does not pass through the origin, is in the subspace of $\mathbb{R}^2$. Looking at the first condition of subspace, we can see that the line $L$ does not contain the zero vector because the zero vector is at the origin. Therefore, $L$ is not in the subspace of $\mathbb{R}^2$.

## Column space

We looked at whether a set defined by a span is a subspace. Turns out, if we have a matrix $A$, the linear combinations that its columns form are a specific kind of subspace called the **column space**. The set is denoted by $\text{Col}\,A$.

If $A = [\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n]$ is an $m \times n$ matrix, then $\text{Col}\,A = \text{Span}\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\} \subseteq \mathbb{R}^m$. ($\subset$ means "is a proper subset of" while $\subseteq$ means "is an ('improper') subset of".) Note it is $\mathbb{R}^m$ because there are $m$ many entries in each vector. So each vector is $m$ big, not $n$ big. The number of entries determines what $\mathbb{R}$ the span will be a subset of.

So basically, column space is the possible things that $\vec{b}$ can be.

"Is $\vec{b} \in \text{Col}\,A$?" is just another way of asking the question "Can the columns of $A$ form a linear combination of $\vec{b}$?" or "Is $\vec{b}$ in the span of the columns of A?". To answer this question, just find $\vec{x}$ in the matrix equation $A\vec{x} = \vec{b}$.

## Null space

Whereas the column space refers to the possibility of what $\vec{b}$ can be, null space refers to what possible solutions are available for the homogeneous system $A\vec{x} = \vec{0}$. In other words, null space is the possibility of what $\vec{x}$ can be.

When $A$ has $n$ columns (being an $m \times n$ matrix), $\text{Nul}\,A \subseteq \mathbb{R}^n$.

The null space of an $m \times n$ matrix $A$ is a subspace of $\mathbb{R}^n$. Therefore, the set of all solutions that satisfies $A\vec{x} = \vec{0}$ of $m$ homogeneous linear equations in $n$ unknowns is a subspace of $\mathbb{R}^n$.

### Relating column space and null space together

$$A\vec{x} = \vec{b}$$

$$\text{Nul } A \uparrow \qquad \uparrow \text{Col } A$$

Column space has to do with the span of $\vec{b}$ while null space has to do with the span of $\vec{x}$.

Column space is a subspace of $\mathbb{R}^m$ while null space is a subspace of $\mathbb{R}^n$.

Let's work through an example to relate both of these concepts together. Let $A = \begin{bmatrix} 1 & -3 & -4 \\ -4 & 6 & -2 \\ -3 & 7 & 6 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 3 \\ 3 \\ -4 \end{bmatrix}$. Is $\vec{b} \in \text{Col } A$?

This question is really asking "is there an $\vec{x}$ so that $A\vec{x} = \vec{b}$?" The reduced echelon form of $A = \begin{bmatrix} 1 & 0 & 5 & -9/2 \\ 0 & 1 & 3 & -3/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. We can see not only that the system is consistent, but it has infinitely many solutions. $\boxed{\text{So yes, } \vec{b} \in \text{Col } A}$.

The solution to $A\vec{x} = \vec{0}$ is $\vec{x} = x_3 \begin{bmatrix} -5 \\ -3 \\ 1 \end{bmatrix}$. So $\text{Nul } A = \text{Span} \left\{ \begin{bmatrix} -5 \\ -3 \\ 1 \end{bmatrix} \right\}$. Furthermore, $\text{Col } A = \text{Span} \left\{ \begin{bmatrix} 1 \\ -4 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ 6 \\ 7 \end{bmatrix} \right\}$. We neglected the last vector because it turns out $\begin{bmatrix} -4 \\ 2 \\ 6 \end{bmatrix} \in \text{Span} \left\{ \begin{bmatrix} 1 \\ -4 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \\ 7 \end{bmatrix} \right\}$, so it'd be redundant to include it.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 6: Inverse matrices, column space and null space:
`https://www.youtube.com/watch?v=uQhTuRlWMxw`

## 2.6 Basis

We can use a few vectors to form many different vectors with a linear combination. Specifically, we want these vectors to all be unique in their own right when forming the span, so the vectors need to be linearly independent of each other. We will call this a *basis*.

A **basis** for a subspace $H \subseteq \mathbb{R}^n$ is a linearly independent set of vectors in $H$ that span to give $H$.

The set of vectors $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ is a basis for $H$ if two conditions are true:

1. $\mathcal{B}$ is linearly independent.

2. $H = \text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$

Restated succinctly, if $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p \in \mathbb{R}^n$, then $\text{Span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p\}$ is a subspace of $\mathbb{R}^n$.

This means everything in the subspace $H$ can be made with a basis. Moreover, there is only one way (a unique way) to make such a linear combination with this basis.

## Standard basis

The most common basis combination that spans $\mathbb{R}^2$ is $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$. This is an example of the **standard basis** $\mathcal{B}_{std} = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$. In fact, we usually derive a standard basis for a certain size $n$ from the identity matrix $I_n$ because its columns are linearly independent and spans all of $\mathbb{R}^n$ in a standard way that's very intuitive to us.

We can think of the standard basis as components of a vector. When we use $\hat{\imath}, \hat{\jmath}, \hat{k}$ to denote directions, we are really just using them to represent $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$ respectively.

$$\hat{\imath} = \begin{bmatrix}1\\0\\0\end{bmatrix}, \hat{\jmath} = \begin{bmatrix}0\\1\\0\end{bmatrix}, \hat{k} = \begin{bmatrix}0\\0\\1\end{bmatrix}$$

When we write our vectors, we usually do so by assigning weights to $\hat{\imath}, \hat{\jmath}, \hat{k}$ to represent the x-component, y-component, and z-component. This, in reality, is just a linearly independent linear combination comprised of the standard basis. This means $\text{Span}\{\hat{\imath}, \hat{\jmath}, \hat{k}\} \in \mathbb{R}^3$.

## Nonstandard basis

Usually, we have to find a basis for a subspace that isn't so simple as the identity matrix.

For a null space, find the solution to $\vec{x}$ for $A\vec{x} = \vec{0}$. The free variables form a solution in parametric vector form. The vectors associated with each free variable form the basis for the null space.

For a column space, the basic variables are linearly independent whereas the free are not. So we can use the pivot columns of a matrix to form a basis for the column space of A. Note we use the pivot columns, not the reduced columns with pivots in them. There is a difference! Use the columns from the original, unreduced matrix to form a basis for $\text{Col}\, A$.

In summary:

- **Null space** comes from the vectors associated with the parametric vector solution to $A\vec{x} = \vec{0}$. That basis will have $n$ many entries in its vectors.
- **Column space** comes from the pivot columns. That basis will have $m$ many entries in its vectors.

## Basis examples

These examples will help you better understand basis. (They are borrowed from Dr. Widmer's lecture notes.)

1. Find a basis for the null space of matrix $A = \begin{bmatrix}-3 & 6 & -1 & 1 & -7\\ 1 & -2 & 2 & 3 & -1\\ 2 & -4 & 5 & 8 & -4\end{bmatrix} \sim \begin{bmatrix}1 & -2 & 0 & -1 & 3\\ 0 & 0 & 1 & 2 & -2\\ 0 & 0 & 0 & 0 & 0\end{bmatrix}$.

   (by the way, $\sim$ means "row equivalent to". The matrix before $\sim$ was not reduced whereas the matrix after was reduced.)

From this, we can see that $\vec{x} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}$. This is a linearly independent set that spans, making it a basis. So we can say that a basis for $\text{Nul}\, A$ is $\left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right\}$.

2. Find a basis for the column space of $B = \begin{bmatrix} 1 & 0 & -4 & 3 & 0 \\ 0 & 1 & 5 & -7 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. (This matrix just so happens to already be reduced.)

   The columns that contain basic variables are columns 1, 2, and 5. Therefore, they form a basis for $\text{Col}\, B$: $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$.

   Important note: The last row is all zeroes. However, just because every entry in the fourth row is zero does not mean the entries are in $\mathbb{R}^3$; in fact, they are actually in $\mathbb{R}^4$.

3. Find a basis for the column space of $A = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

   We can see that the pivot columns are 1, 2, and 5. However, the reduced matrix's columns are not the actual vectors that form a basis. Notice that it is impossible to form a nonzero value for the 5th row using the reduced columns. However, this must be true. Therefore, we use the unreduced 1st, 2nd, and 5th columns to form our basis.

   $$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -9 \\ 2 \\ 1 \\ -8 \end{bmatrix} \right\}.$$

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 2: Linear combinations, span, and basis vectors:
`https://www.youtube.com/watch?v=k7RM-ot2NWY`

## 2.7 Coordinate vector

A basis is a set of vectors that can represent any vector in subspace it spans. Because a basis is linearly independent, that means any vector that's formed by a linear combination of this basis has one and only one set of weights to form it (unique set of weights). So, let's store these weights in a single vector called the **coordinate vector**. The coordinate vector is always in relation to this basis, otherwise it wouldn't be valid.

Suppose the set $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_p\}$ is a basis for a subspace $H$. For each $\vec{x} \in H$, the coordinates of $\vec{x}$ relative to the basis $\vec{B}$ are the weights $c_1, c_2, \ldots, c_p$ so that $\vec{x} = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_p\vec{b}_p$ and the vector in $\mathbb{R}^p$, $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{bmatrix}$ is called the coordinate vector of $\vec{x}$ (relative to $\mathcal{B}$) or $\mathcal{B}$—coordinate vector of $\vec{x}$. Then, the following is true:

$$\mathcal{B}[\vec{x}]_{\mathcal{B}} = \vec{x}$$

Coordinate vectors will be further explored in Chapter 4.

## 2.8 Dimension, rank, and nullity

From the previous section, we learned that basis gives us the fewest number of vectors needed to span a subspace. So the space that the basis spans is actually not always going to be the number of columns in a matrix $n$, especially not when $n \neq$ # of pivots.

The **dimension** of a nonzero subspace $H$ is the number of vectors present in a basis (and therefore *any* basis as well) for H. It's denoted by $\dim H$.

Dimension is very intuitive to the definition that you're familiar with. Dimension, just like when you think of 2-D, 3-D, etc., is how many dimensions a linear combination can span. If it's just a line, then 1. If it's a plane, then 2. And so on.

Here are some examples of dimensions:

- By definition, the dimension of the subspace $\{\vec{0}\}$ is defined to be 0. It's equivalent to a point in space. A point is 0-D.
- A plane through the zero vector in $\mathbb{R}^3$ has dimension 2.
- A line through the zero vector in $\mathbb{R}^2$ has dimension 1. In fact, it could be in $\mathbb{R}^n$ and the dimension would still be 1.

Now, dimension is for subspaces, but the equivalent concept for matrices is called **rank**. In fact, the rank of a matrix $A$ is the dimension of the column space of $A$. It's denoted by $\operatorname{rank} A$, so $\operatorname{rank} A = \dim \operatorname{Col} A$. (To reduce confusion, you can put parentheses around the $\operatorname{Col} A$ like so: $\dim(\operatorname{Col} A)$.)

Some texts call the equivalent $\dim \operatorname{Nul} A$ the **nullity** of $A$. We will be using this notation too (because it's more efficient than writing $\dim \operatorname{Nul} A$).

Since rank is based on the number of vectors in a column space, and column space is based on the number of basic variables in a matrix, then we can say that $\operatorname{nullity} A$ (the opposite of rank) is the number of free variables in a matrix. It easily follows that the total number of columns $n$ is $\operatorname{rank} A + \operatorname{nullity} A$. This is called the **Rank–Nullity Theorem**.

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

Similarly, the **Basis Theorem** is an abstract version of the Rank Theorem. Let $H$ be a subspace of $\mathbb{R}^n$ that has a dimension of $p$. any linearly independent set of exactly $p$ elements in $H$ is automatically a basis for H. Also, any set of $p$ elements that spans $H$ is automatically a basis for $H$.

Let's break this theorem down.

1. The first sentence says that $\dim H = p$, which implies that the basis is comprised of $p$ basis vectors. This makes sense; have any less and you can't cover all of them. So $n \leq p$.
2. Furthermore, have any more ($n > p$) and the set of vectors become linearly dependent and therefore cannot be a basis anymore.
3. A linearly independent set with exactly $p$ elements has the *perfect* amount of columns to form a basis, because it's a square matrix!

Also, for all invertible matrices, $\operatorname{rank} A = n$ and $\operatorname{nullity} A = 0$.

From the conclusions we have reached in this chapter, we can add upon our Invertible Matrix Theorem.

Here's the original stipulations:

1. $A$ is an invertible matrix.
2. $A$ is row equivalent to $I_n$, the $n \times n$ identity matrix.
3. $A$ has $n$ pivot positions. (i.e. there is a pivot in every row and every column).
4. The homogeneous matrix equation $A\vec{x} = \vec{0}$ has only the trivial solution.
5. The columns of $A$ form a linearly independent set.
6. The linear transformation $x \mapsto Ax$ is one-to-one.
7. The equation $A\vec{x} = \vec{b}$ has exactly one solution $\forall \vec{b} \in \mathbb{R}^n$. ($\forall$ means "for all".)
8. The columns of $A$ span $\mathbb{R}^n$.
9. The linear transformation $x \mapsto Ax$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.
10. There is an $n \times n$ matrix $C$ so that $CA = I_n$.
11. There is an $n \times n$ matrix $D$ so that $AD = I_n$. $C = D$ always.
12. $A^T$ is an invertible matrix.

The following are our new additions (that imply that a square matrix is invertible and imply the other stipulations):

13. The columns of $A$ form a basis for $\mathbb{R}^n$.
14. $\operatorname{Col} A = \mathbb{R}^n$.
15. $\operatorname{rank} A = n$.
16. $\operatorname{Nul} A = \{\vec{0}\}$.
17. $\operatorname{nullity} = 0$.

# Chapter 3

# Determinants

Earlier in Chapter 2, we introduced the concept of a determinant, but we did not go into detail about what it exactly is. In this chapter, we will explore how to calculate the determinant in various ways and explore its properties and applications.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 5: The determinant:
`https://www.youtube.com/watch?v=Ip3X9LOh2dk`

(Throughout this chapter, we may not be able to explain the various meanings of the determinant like this video is able to. We highly recommend you watch this video if you are not sure about the significance of the determinant. This chapter is mainly focused on the computation of the determinant.)

## 3.1 Introduction to determinants

### Pre-definition notations and conventions

A determinant is defined for a $2 \times 2$ square matrix (where $n = 2$) to be $ad - bc$.

For a $2 \times 2$ matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies \det A = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

We use $\det$ to stand for the determinant-taking function upon a matrix. We use vertical bars to denote we're taking the determinant of a matrix.

Determinants are only defined for square matrices.

In the definition of the determinant, we use the notation $A_{ij}$ to denote a specific kind of submatrix that is the same as $A$, except the $i^{\text{th}}$ row and $j^{\text{th}}$ column are removed. This is called a **minor**. (Specifically, because we only remove one row and one column, a first minor.) For instance, given the matrix $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, $A_{11} = \begin{bmatrix} e & f \\ h & i \end{bmatrix}$, $A_{23} = \begin{bmatrix} a & b \\ g & h \end{bmatrix}$, and $A_{31} = \begin{bmatrix} b & c \\ e & f \end{bmatrix}$.

## Definition of the determinant

For cases where $n \geq 2$, we define the determinant of an $n \times n$ matrix $A = [a_{ij}]$ to be

$$\begin{aligned}\det A &= a_{11}\det A_{11} - a_{12}\det A_{12} + \cdots + (-1)^{n-1}a_{1n}\det A_{1n}\\ &= \sum_{j=1}^{n}(-1)^{1+j}a_{1j}\det A_{1j}\end{aligned}$$

Notice that the summation *alternates signs* on each successive term. This is represented by $(-1)^{n-1}$.

This is the most common method to calculate a determinant. However, because this can get computationally heavy for larger matrices, we will now explore ways to more efficiently calculate determinants.

## Laplace expansion (cofactor expansion)

One way we can reduce the number of calculations is to select other rows or columns to act as minors. In the definition above, we only use the first row. There are many cases where the first row is not as efficient as using other rows. Whenever a row contains many zeroes, it will allow the coefficient in the front to be zero, cancelling out that term. We want to do less work, so this is desirable. In this case, we will calculate something called the *cofactor* and use a theorem called Laplace expansion to calculate the determinant.

Given $A = [a_{ij}]$, the $(i, j)$-**cofactor** of $A$ is the number $C_{ij}$ and is:

$$C_{ij} = (-1)^{i+j}\det A_{ij}$$

Using this cofactor definition, we can use a generalized formula for a determinant:

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

This formula is the cofactor expansion across the first row of A. What we have done is take the original definition and squeeze $\det A_{ij}$ and the alternating sign term into $C_{ij}$. But what about calculating the cofactor expansion for any row or column? We shall use the **Laplace expansion theorem** (cofactor expansion across any row or column):

The determinant of an $n \times n$ matrix $A$ can be computed by a cofactor expansion across any row or down any column. For each $1 \leq i \leq n$ and $1 \leq j \leq n$:

$$\begin{aligned}\det A &= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \text{ (for rows)}\\ &= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \text{ (for columns)}\end{aligned}$$

Keep in mind that the alternation of signs is now determined by the cofactor term!

## Sign patterns for terms of a determinant summation

In the definition of a cofactor, the sign is determined by the term $(-1)^{i+j}$. This means

- if $i + j$ is **even**, the term will be positive.
- if $i + j$ is **odd**, the term will be negative.

On a matrix, it can be visually determined without calculating based on this pattern:

$$\begin{array}{c c} & \begin{array}{ccccc} + & - & + & - & + \end{array} \\ \begin{array}{c} + \\ - \\ + \\ - \\ + \end{array} & \begin{bmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \\ p & q & r & s & t \\ u & v & w & x & y \end{bmatrix} \end{array}$$

Multiply the sign of the row by the sign of the column. $a$ would be positive ($+ \times + = +$, $b$ would be negative ($+ \times - = -$), $f$ would be negative ($- \times + = -$), and $g$ would be positive ($- \times - = +$). Keep this in mind if you go across an arbitrary row or column! The term might start as a negative instead of as a positive. (But at least it will still always alternate signs.)

So the signs of each entry would look like this:

$$\begin{array}{c c} & \begin{array}{ccccc} + & - & + & - & + \end{array} \\ \begin{array}{c} + \\ - \\ + \\ - \\ + \end{array} & \begin{bmatrix} + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \\ - & + & - & + & - \\ + & - & + & - & + \end{bmatrix} \end{array}$$

### Triangular matrix determinant calculation

Using Laplace expansion, we can find that if all elements but the first of a column of a matrix is zero, then only that first element's determinant of its minor matrix will matter. If we repeat this pattern, we get a triangular matrix where all of the entries below the main diagonal are zero.

Well, that means that on triangular matrices, we only need to multiply the *entries along the main diagonal* to get the determinant.

If $A$ is a triangular matrix, then $\det A$ is the <u>product of the entries on the main diagonal</u> of $A$.

Keep in mind that triangular matrices can either be upper triangular matrices (where all of the nonzero entries are at the top right of the matrix) or lower triangular matrices (where all the nonzero entires are at the bottom left of the matrix).

## 3.2 Properties of determinants

Performing elementary row operations on a matrix will change the determinant of a matrix in different ways, depending on which type of operation was made. Thankfully, these changes will affect the determinant in a systematic way such that we can correct for the differences.

Let $A$ be a square matrix and $B$ be the resulting matrix after the elementary row operation has been performed. Then the elementary row operations will affect the determinant in the following ways:

- **Replacement** (replacing a row's entries with each former entry plus another row's entry scaled): this causes no change. $\det B = \det A$.
- **Interchange** (swapping the position of two rows): because the signs of the determinant alternate every row, a swap will alternate the determinant's sign. $\det B = -\det A$.

- **Scaling** (multiplying all of the entries of a row by a scalar): if you multiplied a row by $k$, then the new determinant is $k$ times the old determinant. $\det B = k \det A$.

**Caution!** When performing row operations, we really are doing $\det A \to \frac{1}{k} \det B$. Keep in mind the difference between the definition of scaling and how it is applied in row reduction.

Another property is relating the numerical value of the determinant with the actual invertibility of the matrix. If a matrix's determinant is zero, it is singular (i.e. not invertible). If it's nonzero, it's invertible. If the determinant function is not defined, then it's not a square matrix.

$$\det A = 0 \iff A \text{ is singular}$$

The determinant of a matrix's transpose is the same as the original matrix. Therefore, we can actually perform "elementary column operations" on a matrix and get the same determinant.

$$\det A = \det A^T$$

One last property is that:

$$\det(AB) = (\det A)(\det B)$$

### Summary of determinant properties

- Replacement: $\det B = \det A$
- Interchange: $\det B = -\det A$
- Scaling: $\det B = k \det A$
- $\det A \neq 0 \iff A$ invertible
- $\det A = \det A^T$
- $\det(AB) = (\det A)(\det B)$

## 3.3 Cramer's rule

For the matrix equation $A\vec{x} = \vec{b}$, you can solve for each of $\vec{x}$'s entries $x_i$ by using Cramer's rule. While it is computationally ineffective, it works. Let $A_i(\vec{b})$ be the matrix $A$ but the $i$th column is replaced with $\vec{b}$. Then the $i$th entry of $\vec{x}$ can be determined with the following formula:

$$x_i = \frac{\det A_i(\vec{b})}{\det A}$$

# Chapter 4

# Vector spaces

Vector spaces are an abstract concept that's similar to a subspace, which we defined earlier. In fact, this is where linear algebra ties into *abstract algebra*, which is a higher-order class of algebra. In this chapter, we'll link everything we have learned so far with this more general concept of vector spaces, and introduce some new, pertinent things along the way.

## 4.1 Introduction to vector spaces and their relation to subspaces

An integral part of abstract algebra is grouping structures, each with its own unique characteristics. A **vector space** is a kind of structure; it's a nonempty set (let's call it $V$) of objects (called vectors) on which are defined two operations: addition and scalar multiplication. Vector spaces are subject to the following axioms that make vector spaces what they are:

1. Just like subspaces, $V$ is closed under addition. That means adding two vectors in $V$ will always produce a vector that's still in $V$.
2. $V$ is also closed under scalar multiplication.
3. The addition operation is communicative: $\vec{u}+\vec{v} = \vec{v}+\vec{u}$ - it doesn't matter which operand ($\vec{u}$ and $\vec{v}$ are operands) comes first.
4. The addition operation is associative. $(\vec{u}+\vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$ - it doesn't matter in which order you process the operations.
5. There is a zero vector $\vec{0} \in V$ so that $\vec{u}+\vec{0} = \vec{u}$. (In abstract algebra, this is known as the identity element.)
6. For each $\vec{u} \in V$, there is a vector $-\vec{u} \in V$ so that $\vec{u} + (-\vec{u}) = \vec{0}$.
7. The scalar multiplication operation is distributive. $c(\vec{u}+\vec{v}) = c\vec{u}+c\vec{v}$ and $(c+d)\vec{u} = c\vec{u} + d\vec{u}$.
8. The scalar multiplication operation is associative. $c(d\vec{u}) = (cd)\vec{u}$.
9. The identity property for scalar multiplication is that $1\vec{u} = \vec{u}$.

Where do vector spaces come in? Previously, when we defined linear transformations, we were actually being too specific on what space it occurred in. A linear mapping is a transformation from one *vector space* to another. So that means working with vector spaces is like working in $\mathbb{R}^n$: now, instead of using $\mathbb{R}^n$, we use vector spaces and all the properties that come with it. (In fact, an $n$-dimensional vector space is an isomorphism of $\mathbb{R}^n$ — i.e. it is a one-to-one mapping.)

The following are vector spaces:

- $M_{m,n}$, the set of $m \times n$ matrices, whose entries are in $\mathbb{R}$.
- $M_n$, the set of all $n \times n$ matrices, whose entries are in $\mathbb{R}$.
- $\mathbb{R}^n$.
- $\mathbb{P}_n$ – polynomials of at most degree $n$ in variable $x$ with coefficients in the reals ($\mathbb{R}$).
- $\mathbb{P}_\infty$ – power series in $x$ of the form $\sum_{n=0}^{\infty} a_n x^n$.
- Let $\mathbb{S}$ be the set of doubly infinite sequences, $\{y\} = (\dots, y_{-2}, y_{-1}, y_0, y_1, y_2, \dots)$.
- Let $F = \{f : \mathbb{R} \mapsto \mathbb{R}\}$ be the set of real-valued functions with operations $(f + g)(x) = f(x) + g(x)$ and $(rf)(x) = r \cdot f(x)$.

## Subspaces in relation to vector spaces

We already defined what a subspace is. Now, we can say that subspaces are really just vector spaces living in other vector spaces. They exist in other vector spaces. It's like a house within a house.

Which of the following are subspaces? Let's use the definition we established back in Chapter 2.

- $\{\vec{0}\}$ in a vector space $V$? Yes.
- $\mathbb{P}_n \subset \mathbb{P}_\infty$? Yes.
- Invertible $n \times n$ matrices in $M_n$? Well, the zero matrix has no pivots. Furthermore, neither matrix addition nor matrix multiplication are closed operations. No.
- Upper triangular matrices $n \times n$ in $M_n$? Zero matrix is technically an upper triangular matrix. And it's closed under addition and scalar multiplication. Yes.
- Is $\mathbb{R}^2$ a subspace of $\mathbb{R}^3$? No, they're considered entirely different subspaces. You can't add something in $\mathbb{R}^2$ to $\mathbb{R}^3$. No.
- Is $H = \{ \begin{bmatrix} r \\ s \\ 0 \end{bmatrix} : r, s \in \mathbb{R}\}$ a subspace of $\mathbb{R}^3$? Isomorphisms (which we have not defined yet) has that zero, but it has 3 entries, so it's in $\mathbb{R}^3$. Yes.
- A line not through the origin in $\mathbb{R}^2$, or a plane not through the origin in $\mathbb{R}^3$? They do not include their respective zero vectors. No.
- The set of all points in the form $(3s, 2+5s)$ in $\mathbb{R}^2$? That additive factor makes the zero vector outside of this set. No.

Now, subspaces are simply parts of a vector space. For instance, a plane going through the origin in $\mathbb{R}^3$ would be considered a subspace of $\mathbb{R}^3$. (We will later establish that $\mathbb{R}^n$ is effectively a vector space.)

Now, how do we create subspaces? We usually start with linear combinations. It turns out that the following theorem provides a good foundation:

If $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p$ are vectors in a vector space $V$, then Span$\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\}$ is a subspace of $V$.

This is basically saying that you can use some vectors in a specific vector space to form a linear combination, and all the possible combinations creates a subspace of the vector space the vectors were in.

The easiest way to show something is a subspace of a space is to show it is a *spanning set* of a space. We call Span$\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\}$ the subspace spanned by (or generated by) $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p$. Given a subspace $H$ of $V$, a spanning set (or generating set) of $H$ is a set $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\}$ so that $H = $ Span$\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\}$.

A basis is a spanning set, but not all spanning sets are bases. (A basis is a linearly independent spanning set, while spanning sets are not necessarily linearly independent.)

For instance, say we have a matrix $A = [a_1\ a_2\ a_3\ a_4\ a_5]$ where columns 1, 2, and 3 contain pivots. That means a basis for $\text{Col}\, A$ could be formed by $\mathcal{B} = \{a_1, a_2, a_3\}$. However, a spanning set could contain $\{a_1, a_2, a_3, a_4, a_5\}$. Or $\{a_1, a_2, a_3, a_4\}$. These are both linearly dependent, but notice they all contain the linearly independent columns of the matrix that are necessary to establish the basis.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 11: Abstract vector spaces:
`https://www.youtube.com/watch?v=TgKwz5Ikpc8`

## 4.2 Null spaces, column spaces, and linear transformations

Suppose $A$ is an $m \times n$ matrix.

- $\text{Col}\, A$ is the span of the columns of $A$. $\text{Col}\, A \subseteq \mathbb{R}^m$.
- $\text{Nul}\, A$ contains all vectors $\vec{x}$ such that $A\vec{x} = \vec{0}$. $\text{Nul}\, A \subseteq \mathbb{R}^n$.

For the matrix (linear) transformation $\vec{x} \mapsto A\vec{x}$, we go from $\mathbb{R}^n$ (the domain) to $\mathbb{R}^m$ (the codomain). The codomain is the space where the range (the solution of the transformation, where the images lies) lives in. In a typical transformation that you're used to (like $f(x)$), the domain is equivalent to the codomain, so there is no need to specify both. Now that we're in linear algebra, you do need to specify both.

The range of the transformation is the column space ($\text{Col}\, A$) in the codomain. (Image)

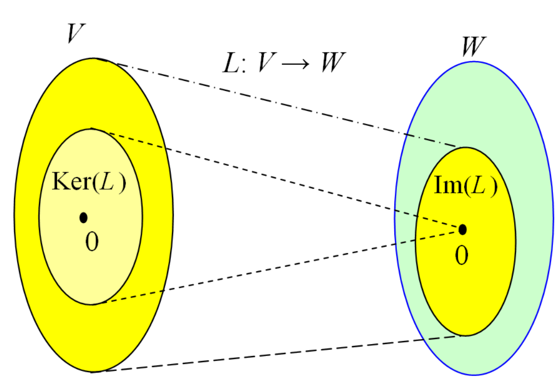
The subspace that is mapped to $\vec{0}$ is the null space in the domain. (Kernel)

For vector spaces $V$ and $W$, a linear transformation $T : V \mapsto W$ is a rule that assigns to each vector $\vec{x} \in V$ a unique vector $T(\vec{x}) \in W$ so that:

1. $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}), \forall \vec{u}, \vec{v} \in V$
2. $T(c\vec{u}) = cT(\vec{u}), \forall \vec{u} \in V$ and all scalars $c$.

- The kernel of $T$ is the set of all $\vec{u} \in V$ so that $T(\vec{u}) = \vec{0}$.
- The range of $T$ is the set of all vectors in $W$ of the form $T(\vec{x})$ for some $\vec{x} \in V$. The range is the set of all images. In other words, the range is the set of the images of mapping.

The following graphic should give you a good idea of the above definition relating images and kernels. All vectors in the kernel of the transformation $L$ are members of $V$ that map to $\vec{0}$, while some vectors in $V$ map to the image of the transformation $L$ in the vector space $W$. (From Wikimedia Commons: `https://commons.wikimedia.org/wiki/File:KerIm_2015Joz_L2.png`)



## 4.3 Spanning sets

If we have a set of vectors $S$ that are able to span all vectors in a vector space $V$, then we call $S$ a **spanning set**. In contrast to a basis, a spanning set may not necessarily be linearly independent (i.e. there may be redundant vectors in it).

Two important stipulations arise from a spanning set:

1. If some $\vec{v}_k \in S$ is a linear combination of the remaining vectors in $S$ (i.e. it's redundant), then $S$ is linearly dependent because of $\vec{v}_k$, so the set $S'$ formed by removing $\vec{v}_k$ still spans $\vec{v}_k$.

2. If $H$ is not simply the zero vector ($H \neq \{\vec{0}\}$), then some subset of $S$ is a basis for $H$. (From this spanning set, if we take some or all of the vectors from it, it'll form a basis.)

## 4.4 Coordinate systems

In this section, we will learn that since bases are *unique* representations for the vectors that it spans, we can simply take a vector containing relevant weights (i.e. a coordinate vector), with their order corresponding with the vectors in a basis, and multiplying them together will return that same vector $\vec{x}$.

Remember from chapter 2, we established what a coordinate vector was. Now we extend this to vector spaces too: If we have a basis $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ which is a basis for vector space $V$, then for each vector in $V$ ($\vec{x} \in V$), there's a unique set of scalars $c_1, c_2, \ldots, c_n$ (weights) so that when they're combined with their respective basis vectors, they form $\vec{x}$.

$$\vec{x} = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_n\vec{b}_n$$

Not only do these weights exist, but these weights are unique. This means using this specific basis, there is only one set of scalars to represent $\vec{x}$. This is the stipulation of the **unique representation theorem**.

A vector that stores these unique weights, as we have established in chapter 2, is called a **coordinate vector**. It turns out that we can transform our vector $\vec{x}$ into the coordinate vector relative to a certain basis. The linear transformation $\vec{x} \mapsto [\vec{x}]_\mathcal{B}$ is called the coordinate mapping determined by $\mathcal{B}$. This allows the following to be true:

$$P_\mathcal{B}[\vec{x}]_\mathcal{B} = \vec{x}$$

where $P_\mathcal{B} = \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \ldots & \vec{b}_n \end{bmatrix}$, the matrix whose columns comprise of the basis vectors in $\mathcal{B}$.

Fun fact: Under the standard basis $\mathcal{E} = \{\vec{e}_1, \vec{e}_2\}$, which comprise the columns of the identity matrix $I$, all coordinate vectors are equivalent to the vectors they form with the standard basis $\mathcal{E}$.

This linear transformation $\vec{x} \mapsto [\vec{x}]_\mathcal{B}$ is one-to-one and onto. Everything in vector space $V$ (which is where $\vec{x}$ resides) also exists on $W$ (which is where $[\vec{x}]_\mathcal{B}$ resides). That means this transformation is easily reversible, and all operations for $V$ are also applicable to $W$ when transformed accordingly.

When we have a one-to-one transformation from $V$ onto $W$, we call this an **isomorphism**. If two vector spaces have an isomorphism between them, they're *isomorphic*.

Isomorphisms provide immense benefits. We can now relate each and every vector operation and calculation in $V$ to those in $W$.

- Since isomorphisms are one-to-one and onto, they are invertible.
- Since they are linear transformations, we can work with $[\vec{x}]_{\mathcal{B}}$ in $\mathbb{R}^n$ the same way we work with $\vec{x}$ in $V$. Vector spaces isomorphic to each other are functionally the same space. We can therefore relate these vector spaces with $\mathbb{R}^n$.

We have now established that we can extend this relationship not just between vector spaces, but also from a vector space to $\mathbb{R}^n$. If a basis $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$ for the vector space $V$ has a coordinate mapping $\vec{x} \mapsto [\vec{x}]_{\mathcal{B}}$, then this mapping is one-to-one from $V$ onto $\mathbb{R}^n$. (i.e. the mapping is an isomorphism between $V$ and $\mathbb{R}^n$.)

$$V \simeq \mathbb{R}^n$$

For instance, if we are given a basis $\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$ for a subspace $H$, and we want to see whether $\vec{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix} \in H$. If so, find $[\vec{x}]_{\mathcal{B}}$.

Reducing the augmented matrix $[\vec{v}_1 \vec{v}_2 | \vec{x}]$ gives us $\begin{bmatrix} 3 & -1 & 3 \\ 6 & 0 & 12 \\ 2 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$. Since this matrix is consistent, we can establish $\vec{x} \in H$. Furthermore, we can see that $x_1 = 2$ and $x_2 = 3$, so the coordinate vector relative to basis $\mathcal{B}$ is $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Note that there isn't a $x_3$ variable. This means the coordinate vector is in $\mathbb{R}^2$ even though the basis vectors and $\vec{x}$ are in $\mathbb{R}^3$.

What can we conclude from this example? The dimension of $H$ is 2, aka $\dim H = 2$, because there are only two entries in the coordinate vector. Furthermore, this establishes $H \simeq \mathbb{R}^2$ ($H$ is isomorphic to $\mathbb{R}^2$ but living in $\mathbb{R}^3$).

In $\mathbb{R}^3$, any subspace $H$ will have either dimension 0, 1, 2, or 3. If $\dim H = 1$, then this is a copy of $\mathbb{R}$ in $\mathbb{R}^3$ (an isomorphism). $H \simeq \mathbb{R}$ but $H \subseteq \mathbb{R}^3$.

## 4.5 Dimension of a vector space

We can generalize the definition of linear independence to vector spaces too, and use this to form the definition of a dimension space of a vector space. Below are some fundamental theorems and facts concerning this subject matter.

1. Let's say a vector space $V$ has a basis $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_n\}$. Then any set in $V$ more than $n$ vectors must necessarily be linearly dependent.

2. If a vector space $V$ has a basis of $n$ vectors, then every basis of the same $V$ must also contain exactly $n$ vectors.

   This makes sense. Otherwise, if there are any fewer, it would only be able to span a certain subspace of $V$, but not all of $V$. If there were more vectors and all were still linearly independent, then that's an even larger $V$. But if there were more vectors and the newly added vector were linearly dependent to the original set, then that's not a basis anymore (since it's not linearly independent), only a spanning set.

3. If $V$ is not spanned by a finite set, then $V$ is **infinite-dimensional**.

4. If $H$ is a subspace of a finite-dimensional vector space $V$, then any linearly independent set in $H$ can become a basis of $H$ if necessary. Also, $H$ will be finite-dimensional and $\dim H \leq \dim V$.

5. Basis Theorem: If $V$ is a $p$-dimensional vector space (where $p \geq 1$), then any linearly independent set of exactly $p$ elements in $V$ is automatically a basis for $V$. Any set of exactly $p$ elements that spans $V$ is also automatically a basis for $V$.

   This means we can show that two vector spaces are the exact same thing if they have the same span. For instance, let's say we have a $3\times 7$ matrix $A$ with 3 pivots. Is its column space equal to $\mathbb{R}^3$? Usually, the answer would be no. However, we need to carefully consider this case. We know that $\text{Col}\, A \subseteq \mathbb{R}^3$ (it's definitely a subspace of $\mathbb{R}^3$). But we also know there's 3 pivots in the matrix. This means the dimension of the column space (or rank) is 3. The dimension of $\mathbb{R}^3$ is also 3. Since they have the same dimension, coupled with the fact that $\text{Col}\, A$ is a subspace of $\mathbb{R}^3$, we can therefore conclude $\text{Col}\, A = \mathbb{R}^3$.

## 4.6 Rank of a vector space's matrix

We can carry over the same definitions from Chapter 2 over what a rank of $\mathbb{R}^2$ is, and now apply it to any vector space $V$ through isomorphisms. Now, let's introduce a new concept called a **row space**. It's the set of the possible linear combinations of the rows (instead of columns) of a matrix $A$. The row space of $A$ is equivalent to the column space of $A$ transpose.

$$\text{Row}\, A = \text{Col}\, A^T$$

Finding a basis for our row space $\text{Row}\, A$ is easy. Simply row reduce $A$ to reduced row echelon form. Then use the nonzero rows of the reduced row echelon matrix (i.e. the ones with pivots, or pivot rows) to form your basis.

While the row space of $A$ is the same as $A$ reduced, it is important to note that you cannot always use the rows of $A$ unreduced to form a basis for $\text{Row}\, A$. You should only use the rows of $A$ reduced to form the basis, but knowing that they are the same because row reductions do not affect the row space.

In fact, let's generalize the sharing of row spaces: if two matrices $A$ and $B$ are row equivalent, then their row spaces are actually equivalent. If $B$ is in reduced echelon form, the nonzero rows of $B$ will form a basis for both $\text{Row}\, A$ and $\text{Row}\, B$. (This means all invertible matrices have the same row space.)

$$A \sim B \implies \text{Row}\, A = \text{Row}\, B$$

Furthermore, because there are the same number of pivot columns and pivot rows, we can conclude that the dimension of the column space ($\text{rank}\, A$) is equivalent to the dimension of the row space.

$$\text{rank}\, A = \dim \text{Col}\, A = \dim \text{Row}\, A$$

The implication of this rule is that sometimes, a matrix will be required to have a null space. In fact, a matrix's minimum null space's size is $\operatorname{nullity} A = \dim \operatorname{Nul} A = |m - n|$, where $m$ and $n$ refer to a matrix being $m \times n$.

So can a $6 \times 8$ matrix have $\operatorname{nullity} A = 1$? Nope. The range of rank $A$ is from 0 to 6. $\operatorname{rank} A \leq 6$ (dimension of both the row space and column space) but per the rank–nullity theorem, the rank would have to be 7 ($\operatorname{rank} = 7$). Well, that's not possible. So no, the nullity (dimension of the null space) cannot be less than 2.

## 4.7 Change of basis

There are infinitely many bases for a certain vector space $V$, and sometimes one won't cut it. Furthermore, each coordinate vector is uniquely relative to its basis $\mathcal{B}$ to represent a certain vector $\vec{x}$. So if we want to represent the same vector $\vec{x}$ with a different set of bases, this is what we do: we transform the coordinate vector relative to the original basis and change it to the coordinate vector relative to the new basis. It's kind of like change of variables in calculus. We do this through a linear transformation. And to represent this linear transformation, we use a matrix. This matrix is called the **change of coordinates matrix** (or the *change of basis matrix*) and is represented by $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$, where $\mathcal{B}$ is the original basis and $\mathcal{C}$ is the new basis. This change of coordinates matrix allows the following to be true:

$$[\vec{x}]_{\mathcal{C}} = \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}[\vec{x}]_{\mathcal{B}}$$

The columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are the $\mathcal{C}$ coordinate vectors of the vectors in $\mathcal{B}$:

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} [\vec{b}_1]_{\mathcal{C}} & [\vec{b}_2]_{\mathcal{C}} & \dots & [\vec{b}_n]_{\mathcal{C}} \end{bmatrix}$$

If we are going from one basis to another, they must represent the same subspace. So therefore, they must have the same subspace. Since a change of basis is a linear transformation, and linear transformations are really going from $\mathbb{R}^m \mapsto \mathbb{R}^n$, then $m$ must be the same as $n$. So this means the change of coordinates matrix is always square. Furthermore, it's always invertible because the columns of $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$ are always linearly independent. Since it's invertible, there must be an inverse!

$$\underset{\mathcal{B}\leftarrow\mathcal{C}}{P} = (\underset{\mathcal{C}\leftarrow\mathcal{B}}{P})^{-1}$$

Note we are now going from $\mathcal{C}$ to $\mathcal{B}$ instead of $\mathcal{B}$ to $\mathcal{C}$.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 9: Change of basis:
`https://www.youtube.com/watch?v=P2LTAUO1TdA`

### Change of basis and the standard basis

A special case of the change of coordinates matrix is going from some nonstandard basis $\mathcal{B}$ to the standard basis $\mathcal{E}$. In that case,

$$\underset{\mathcal{E}\leftarrow\mathcal{B}}{P} = P_{\mathcal{B}}$$

This is because all vectors in $\mathbb{R}^n$ can be represented by coordinate vectors of the standard basis. $\vec{x} = [\vec{x}]_{\mathcal{E}}$. This is easy to see because $P_{\mathcal{E}} = \begin{bmatrix} 1 & 0 & \dots \\ 0 & 1 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} = I_n$, the identity matrix. Therefore $P_{\mathcal{E}}[\vec{x}]_{\mathcal{E}} = \vec{x}$ really means $I[\vec{x}]_{\mathcal{E}} = \vec{x}$, hence implying $[\vec{x}]_{\mathcal{E}} = \vec{x}$ in $\mathbb{R}^n$.

If you did not directly translate from one arbitrary basis to another, but rather translated from $\mathcal{B}$ to $\mathcal{E}$ (standard basis) and then from $\mathcal{E}$ to $\mathcal{C}$, then you would really be doing:

$$\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = (P_{\mathcal{C}})^{-1} P_{\mathcal{B}}$$

This is how we will be finding the change of coordinates matrix from one arbitrary (nonstandard) basis to another. We need to start with $P_{\mathcal{B}}$ and multiply by $(P_{\mathcal{C}})^{-1}$, as detailed in the next subsection.

## Finding the change of coordinates matrix

In order to get from $P_{\mathcal{B}}$ to $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$, we need to multiply $P_{\mathcal{B}}$ by $(P_{\mathcal{C}})^{-1}$ on the left side. This transformation can be done easily through reducing the augmented matrix $[P_{\mathcal{C}}|P_{\mathcal{B}}]$ to $[I|\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}]$ because doing so allows us to apply the same steps of going from $P_{\mathcal{C}}$ to the identity matrix to $P_{\mathcal{B}}$, effectively multiplying $P_{\mathcal{B}}$ by $(P_{\mathcal{C}})^{-1}$ to get $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$.

To find this change of coordinates matrix, we need to find $[\vec{b}_i]_{\mathcal{C}}$ for each $i = 1, 2, \dots, n$. So we need to set the matrix in the following method, and then reduce it down to:

$$\begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \dots & \vec{c}_n & | & \vec{b}_1 & \vec{b}_2 & \dots & \vec{b}_n \end{bmatrix} \sim \begin{bmatrix} I_n & \underset{\mathcal{C}\leftarrow\mathcal{B}}{P} \end{bmatrix}$$

where $\vec{c}_i$ are the basis vectors of $\mathcal{C}$ and the analogous is true for $\vec{b}_i$. For instance, if we have $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}, \mathcal{C} = \{\vec{c}_1, \vec{c}_2\}, \vec{b}_1 = \begin{bmatrix} -9 \\ 1 \end{bmatrix}, \vec{b}_2 = \begin{bmatrix} -5 \\ -1 \end{bmatrix}, \vec{c}_1 = \begin{bmatrix} 1 \\ -4 \end{bmatrix}, \vec{c}_2 = \begin{bmatrix} 3 \\ -5 \end{bmatrix}$, and we want to find the change of coordinates matrix from $\mathcal{B}$ to $\mathcal{C}$, we put these vectors into the augmented matrix and reduce the $\mathcal{C}$ basis vectors into the $2 \times 2$ identity matrix $I_2$.

$$\left[\begin{array}{cc|cc} 1 & 3 & -9 & -5 \\ -4 & -5 & 1 & -1 \end{array}\right] \sim \left[\begin{array}{cc|cc} 1 & 0 & 6 & 4 \\ 0 & 1 & -5 & -3 \end{array}\right]$$

So we get $[[\vec{b}_1]_{\mathcal{C}}[\vec{b}_2]_{\mathcal{C}}]$ on the right, which is $= \underset{\mathcal{C}\leftarrow\mathcal{B}}{P}$. Therefore $\underset{\mathcal{C}\leftarrow\mathcal{B}}{P} = \begin{bmatrix} 6 & 4 \\ -5 & -3 \end{bmatrix}$.

# Chapter 5

# Eigenvalues and eigenvectors

It turns out that for certain matrix equations or matrix transformations, that the same transformation can be applied to a certain vector using just a scalar number in place of a matrix. This is useful because instead of computing all those matrix calculations, we can replace it with one scalar multiplication.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this chapter:

Essence of Linear Algebra – Chapter 10: Eigenvectors and eigenvalues:
`https://www.youtube.com/watch?v=PFDu9oVAE-g`

## 5.1 Introduction to eigenvalues and eigenvectors

An **eigenvector** of an $n \times n$ matrix $A$ is a nonzero vector $\vec{x}$ so that $A\vec{x} = \lambda\vec{x}$ for some scalar $\lambda$. A scalar $\lambda$ is called an **eigenvalue** of $A$ if there is a nontrivial solution to $\vec{x}$ of $A\vec{x} = \lambda\vec{x}$. Such an $\vec{x}$ is called an eigenvector corresponding to $\lambda$.

For instance, given the matrix $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, $\vec{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, and $\vec{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, let's compute $A\vec{u}$ and $A\vec{v}$. $A\vec{u} = \begin{bmatrix} -5 \\ -1 \end{bmatrix}$ and $A\vec{v} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Note that $A\vec{v} = 2\vec{v}$. This means 2 is an eigenvalue of $A$, and the eigenvector corresponding to 2 is $\vec{v}$. Also, anything in the span of $\vec{v}$ will be able to use this eigenvalue of 2.

$$A\vec{x} = \lambda\vec{x}$$

How do we tell if a scalar is indeed an eigenvalue of a matrix? We take this equation above and we move both terms to one side. We get $A\vec{x} - \lambda\vec{x} = \vec{0}$. Now we want to factor out $\vec{x}$ per the distributive property of vectors. However, that would result in $A - \lambda$, which is not defined (you can't subtract a scalar from a matrix). So, we multiply $\lambda$ by the identity matrix, which changes nothing. We then can factor out $\vec{x}$ and we get:

$$(A - \lambda I)\vec{x} = \vec{0}$$

If this equation has a nontrivial solution, then $\lambda$ is an eigenvalue. Furthermore, if there is a nontrivial solution, there actually are infinitely many nontrivial solutions (i.e. infinitely many eigenvectors corresponding to $\lambda$). The subspace that contains these solutions is known as the **eigenspace** and is the null space of this matrix equation. Finding a basis for the eigenspace of the matrix (the **eigenvector basis**) corresponding to $\lambda$ is simply finding a basis of the null space of the matrix $A - \lambda I$.

Also, note that $\lambda$ is just one eigenvalue of the matrix. This means that this eigenspace corresponds to this specific eigenvalue. And this eigenspace can be represented by the span of a basis of its null space. Now, these eigenvectors are actually going to be minding their own business in their own eigenspaces. In fact, these eigenspaces' spans don't cross over!

This means that if we have eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ that correspond to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ of an $n \times n$ square matrix $A$, then the set of these eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_p$ is *linearly independent*. Furthermore, if there are $n$ many vectors in this set, then that set forms a basis for all possible eigenvectors of the matrix $A$. The set of all eigenvalues of a matrix is called the **spectrum** of the matrix.

So how do we actually find the eigenvalues without cycling through infinitely many lambdas? Note that $A - \lambda I$ is not an invertible matrix. Therefore, its determinant, $\det(A - \lambda I) = 0$. This means we can treat $\lambda$ as a polynomial and solve for its roots to find eigenvalues, like a quadratic. But we will go into detail over this process in the next section.

We can connect that $A - \lambda I$ having a determinant of $0$ means that we can add some new axioms to the Invertible Matrix Theorem:

- If zero is not an eigenvalue of the matrix, then the matrix is invertible.
- If $\det A \neq 0$, then the matrix is invertible.

First, let's take a look at a specific case. If we have a triangular matrix, then the eigenvalues are simply the entries on the diagonal. This comes from $\dim(A - \lambda I) = \vec{0}$'s roots being the eigenvalues of $A$. Remember that for a diagonal matrix, you simply multiply the diagonal entries to get the determinant? Same thing here.

Let's find the eigenvalues of this matrix $A - \lambda I = \begin{bmatrix} 3 & -6 & 8 \\ 0 & 0 & 4 \\ 0 & 0 & 2 \end{bmatrix}$.

We can see that $3, 0, 2$ are the diagonal entries. Therefore, they are the eigenvalues of this entry.

And that's all of the eigenvalues for this particular matrix! In fact, the maximum number of eigenvalues for an $n \times n$ square matrix is equivalent to $n$.

Can $0$ be an eigenvalue? Yes. Remember that only eigenvectors had to be nonzero, but this restriction does not apply to eigenvalues. Whenever $0$ is an eigenvalue of a matrix, then $A\vec{x} = \lambda\vec{x}$ turns into $A\vec{x} = 0\vec{x}$, or $A\vec{x} = \vec{0}$, which represents a homogeneous linear system. That means $\dim \operatorname{Nul} A > 0$ and $A$ is not invertible because its determinant is $0$ and it's not row equivalent to $I$.

Now, for the matrix $B = \begin{bmatrix} 4 & 0 & 0 \\ -2 & 1 & 0 \\ 7 & 3 & 4 \end{bmatrix}$, note that $4$ appears twice. Therefore, there are only two eigenvalues for this matrix: $1$ and $4$.

While it may be tempting to reduce matrices to echelon form to find eigenvalues, this actually won't work. When finding eigenvalues, it is important to note that row reductions will change a matrix's eigenvalues. In other words, a reduced matrix's eigenvalues will not be the same as the original matrix's eigenvalues except by pure coincidence.

## 5.2 The characteristic equation: finding eigenvalues

In the previous section, we found that if the following equation had any nontrivial solutions, then $\lambda$ would be an eigenvalue:

$$(A - \lambda I)\vec{x} = \vec{0}$$

This equation is called the **characteristic equation**. Note that as long as $\lambda$ is an eigenvalue, $A - \lambda I$ is never an invertible matrix. Therefore, its determinant equals zero: $\det(A - \lambda I) = 0$. This means we can treat $\lambda$ as a polynomial and solve for its roots to find eigenvalues, like a quadratic.

The **characteristic polynomial** is

$$\det(A - \lambda I)$$

which is an $n$-degree polynomial because $A$ is an $n \times n$ matrix. The eigenvalues of $A$ are the roots of the polynomial.

### Problem-solving process for triangular matrices

Let's first find the characteristic equation of the triangular matrix

$$A = \begin{bmatrix} 5 & -22 & 46 & -131 \\ 0 & 3 & 2 & 10 \\ 0 & 0 & 5 & 38 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Using the generic characteristic equation and the fact that the diagonal entries constitute the solutions to the characteristic equation, we can find that the characteristic equation for this is:

$$(5 - \lambda)(3 - \lambda)(5 - \lambda)(1 - \lambda) = 0$$

where the characteristic polynomial is

$$(5 - \lambda)(3 - \lambda)(5 - \lambda)(1 - \lambda) = \lambda^4 - 14\lambda^3 + 68\lambda^2 - 130\lambda + 75 = 0$$

Note that both forms are considered characteristic polynomials, but the first is preferable because it already contains the eigenvalues in the roots.

### Problem-solving process for general matrices

If we do not have a triangular matrix, then we need to do a bit more work. (The following example was borrowed from Rozenn Dahyot, Trinity College Dublin)

Say that we have the following non-triangular matrix:

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

Then we can establish that its characteristic equation looks like this:

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}\right) = \begin{vmatrix} 1-\lambda & -3 & 3 \\ 3 & -5-\lambda & 3 \\ 6 & -6 & 4-\lambda \end{vmatrix} = 0$$

Now, you just need to take the determinant of this matrix.

$$= (1-\lambda)\begin{vmatrix} -5-\lambda & 3 \\ -6 & 4-\lambda \end{vmatrix} - (-3)\begin{vmatrix} 3 & 3 \\ 6 & 4-\lambda \end{vmatrix} + 3\begin{vmatrix} 3 & -5-\lambda \\ 6 & -6 \end{vmatrix}$$

$$= 16 + 12\lambda - \lambda^3 = \lambda^3 - 12\lambda - 16$$

The possible roots are $\pm 1, \pm 2, \pm 4, \pm 8, \pm 16$. Dividing by 4 gives us 0. So we factor our $(\lambda - 4)$ to get $(\lambda - 4)(\lambda^2 + 4\lambda + 4) = (\lambda - 4)(\lambda + 2)^2$. So our two eigenvalues are $\boxed{4, -2}$, with -2 being repeated and having a multiplicity of 2.

The above example was a bit complicated to factor, but gives you a good idea of what the characteristic polynomial can be like to solve.

## 5.3 Similarity

Now that we've learned how to find eigenvalues, can we devise a way to relate matrices with the same eigenvalues and the same multiplicities together? Yes, and this is what we call similar matrices.

If $A$ and $B$ are $n \times n$ matrices, then $A$ and $B$ are **similar** to each other if there's an invertible matrix $P$ such that the following is true:

$$P^{-1}AP \implies \cancel{PP^{-1}}A\cancel{PP^{-1}} = PBP^{-1} \implies A = PBP^{-1}$$

Note the order of multiplying $P$ and $P^{-1}$ reverses when switching sides of the equation. This is because matrix multiplication is not commutative, meaning order matters when multiplying!

In fact, if $A$ and $B$ are similar, they have the same characteristic polynomials, which is why we know they have the same eigenvalues and the same multiplicity for each root of the characteristic polynomial. (Caution: This is a one-way statement. If $A$ and $B$ have the same characteristic polynomial, they still might not be similar to each other.)

If we are converting from $A$ to $PAP^{-1}$, then the transformation is known as a **similarity transformation**.

## 5.4 Diagonalization

Let's say we wanted to square a matrix $A$ and get $A^2 = AA$. Usually, it's a pain to calculate this, and there's no real formula that applies to squaring all matrices. Higher exponents $A^k$ are even more tedious to calculate. But if this matrix has zero values in every entry except along the diagonal, it's really fast to calculate. You would just take each diagonal entry and exponentiate it to a certain desired power.

For instance, if we have a diagonal matrix $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$, then matrix multiplication will show that $D^2 = \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}$. Therefore we can say that $D^k = \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}$.

Furthermore, calculating eigenvalues from those non-triangular matrices is a pain. But it's really easy for diagonal matrices, because it's just the values along the main diagonal. So why don't we find a way to relate those painful matrices to diagonal matrices?

We'll use our similarity definition for diagonalization. A square matrix $A$ is **diagonalizable** if and only if for some invertible $P$, it is similar to a diagonal matrix $D$ and the following is true:

$$A = PDP^{-1}$$

We can also rewrite this equation as

$$AP = PD$$

for computational ease so we can avoid finding the inverse of $P$.

Also, from our realization that diagonal matrices are a great way to compute $A^k$, we can conclude that:

$$A^k = PD^kP^{-1}$$

Now, let's see an example of a $A = PDP^{-1}$ equation:

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

It turns out that an $n \times n$ matrix will only be diagonalizable if and only if $A$ has $n$ linearly-independent eigenvectors. Thus, $A$ is diagonalizable if and only if there are $n$ vectors in the eigenvector basis, meaning it spans all of $\mathbb{R}^n$. The columns of $P$ actually comprise an eigenvector basis. The $i$th column of $P$ corresponds to the entry $d_{ii}$ ($i$th diagonal entry of $D$).

- The $i$th diagonal entry in $D$ is an eigenvalue of $A$.
- The $i$th column of $P$ is the eigenvalue's corresponding eigenvector.
- $D$'s diagonal entries are the spectrum of $A$.
- $P$'s columns comprise an eigenvector basis relative to the eigenvalues in $D$.

So in the above example $A = PDP^{-1}$ equation, $5$ is one eigenvalue of $A$ with corresponding eigenvector $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $3$ is the other eigenvalue of $A$ with corresponding eigenvector $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$.

## The process of diagonalizing matrices

Diagonalizing matrices can be done in six steps.

1. Find the eigenvalues of $A$.
2. Find the eigenvectors of $A$ using the eigenvalues.
3. Check if there are $n$ many eigenvectors.
4. Check whether they are all linearly independent.
5. Assemble $P$ using the eigenvectors you found in step 2.
6. Create $D$ using the eigenvalues you found in step 1.

## Example of diagonalizing a matrix

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

1. First, solve for the characteristic equation $\begin{vmatrix} 1-\lambda & 3 & 3 \\ -3 & -5-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{vmatrix} = 0$. We get $-\lambda^3 - 3\lambda^2 + 4$. Factoring this gives us $-(\lambda-1)(\lambda+2)^2$. Therefore our eigenvalues are $\lambda = -2, 1$, with the multiplicity of $-2$ being 2.

2. Next, reduce $A + 2\lambda$ and $A - 1\lambda$ to find their null spaces to extract their eigenvectors.
   $A + 2\lambda = \begin{bmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. We get $x_1 = -x_2 - x_3$, and the solution set is $x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. It is clear these two vectors are linearly independent because they form a basis for the null space.
   $A - 1\lambda = \begin{bmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. We get $x_1 = x_3$ and $x_2 = -x_3$. So the solution set is $x_3 \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$.

3. There are indeed 3 eigenvectors. We can proceed.

4. Put the eigenvectors in a matrix and reduce. $P = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ so yes, we can say the columns of $P$ are linearly independent. We can proceed.

5. We get $P = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$ from the previous step.

6. Respecting the order we already established in $P$, we get that $D = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Check if $AP = PD$ to verify you have diagonalized the matrix correctly. This is the same thing as checking if $A = PDP^{-1}$ is valid, but eliminates the need to compute an inverse.

**Note from author**: *We will omit covering material concerning the relation between eigenvectors and linear transformations.*

# Chapter 6

# Orthogonality, least squares, and symmetry

This chapter dives into making vectors orthogonal. Essentially, we want our vectors to be at right angles at each other, especially if they form a basis of a subspace. To do so, we'll first define what a dot product is, an inner product, length of a vector, and then distance of a vector. Then, we'll dive into the Gram–Schmidt process to orthogonalize a set of vectors, including taking a look at QR decomposition.

## 6.1 Dot product, length, and distance

### Dot product/Inner product

The **dot product** (also known as the **inner product**) is a vector operation defined as the summation of the product of each entry in a vector, with the result being a scalar (so it is also sometimes called the scalar product). It is a tool we will be using to analyze the orthogonality of a matrix. It is defined as $\vec{u} \cdot \vec{v} = \vec{u}^T \vec{v}$. Note that we must take the transpose of the first vector to satisfy matrix multiplication rules (the first operand's $n$ must equal the second operand's $m$).

There are two ways to calculate dot products. The first way is more relevant to computational linear algebra. Simply take $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$. The dot product is

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{i=1}^{n} u_i v_i = \vec{u}^T \vec{v} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Note that the dot product multiplies a $1 \times n$ matrix by a $n \times 1$ matrix, in that order. Therefore, it produces a $1 \times 1$ matrix, which we usually write as a scalar. Furthermore, you can only take the dot product of two matrices of the same size.

The other way to find it (graphically) is if you have the two vectors' magnitudes (defined soon) and the angle between the two vectors, you can calculate the dot product using the following formula:

$$\vec{u} \cdot \vec{v} = ||\vec{u}|| \, ||\vec{v}|| \cos\theta$$

If the dot product of two vectors equals 0, that means $\theta = 90°$ and therefore $\cos(90°) = 0$, so the dot product will be 0. This also works with component form.

The subspace formed by the dot product is called the *inner product space*.

The dot product has important geometric ramifications. We use the dot product to determine the length of a vector and the distance between two vectors. More importantly, however, when the dot product of two vectors is zero, then we can conclude they are perpendicular, or *orthogonal* to each other.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 7: Dot product and duality:
`https://www.youtube.com/watch?v=LyGKycYT2v0`

### Length/magnitude of a vector

When we plot a vector, the **length** (or **magnitude**) from its head to its tail is defined by

$$||\vec{u}|| = \sqrt{\vec{u} \cdot \vec{u}} = \sqrt{(u_1)^2 + (u_2)^2 + \ldots (u_n)^2}$$

Notice the similarities of this formula to the distance formula you learned in algebra. However, this is not the same as the distance between two vectors, which will be defined next.

The length/magnitude will always be nonnegative and will only be zero when $\vec{u} = \vec{0}$. Furthermore, the following stipulations arise:

- $||\vec{u}||^2 = \sqrt{\vec{u} \cdot \vec{u}}^2 = \vec{u} \cdot \vec{u}$
- $||c\vec{u}|| = \sqrt{c\vec{u} \cdot c\vec{u}} = \sqrt{c^2\vec{u} \cdot \vec{u}} = \sqrt{c^2}\sqrt{\vec{u} \cdot \vec{u}} = |c|||\vec{u}||$

### Unit vector

A **unit vector** is a vector with length 1. The expression

$$\frac{1}{||\vec{u}||}\vec{u}$$

gives us the unit vector in the same direction as $\vec{u}$.

### Distance between two vectors

In the one-dimensional real space $\mathbb{R}$, the distance between two scalars $a, b$ is the absolute value of their difference, $|a - b|$. The **distance between two vectors** in $\mathbb{R}^n$ is defined to be

$$\text{dist}(\vec{u}, \vec{v}) = ||\vec{u} - \vec{v}||$$

## 6.2 Introduction to orthogonality

This section lays the groundwork for making use of orthogonality-related concepts.

## Orthogonal vectors

By definition, two vectors $\vec{u}$ and $\vec{v}$ are orthogonal to each other if their inner product is equal to 0. However, any of these three equations would indicate that $\vec{u}$ is orthogonal to $\vec{v}$.

$$\left\{ \begin{array}{c} \vec{u} \cdot \vec{v} = 0 \\ \text{dist}(\vec{u}, \vec{v}) = \text{dist}(\vec{u}, -\vec{v}) \\ ||\vec{u} + \vec{v}||^2 = ||\vec{u}||^2 + ||\vec{v}||^2 \end{array} \right\} \implies \vec{u} \perp \vec{v}$$

## Orthogonal complements

If we have a trimmed version of the standard basis $\mathcal{E}_{xy} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$, then we have a 2D plane (let's call it the vector space $W$). But how would we expand this to all of $\mathbb{R}^3$? We simply add a vector to this basis with a nonzero value in the third entry. But let's say we want all of our basis vectors to be orthogonal to each other, because it's neater that way. And let's say we only want unit vectors in our bases. Then there is only one choice: $\vec{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. This vector part of a set of vectors that are also orthogonal to $W$ (the 2D plane), called an *orthogonal complement* of $W$.

If a vector $\vec{x}$ is orthogonal to every vector in the subspace $W \subseteq \mathbb{R}^n$, then $\vec{x}$ is orthogonal to the entire subspace $W$. The set of all vectors $\vec{x}$ that are orthogonal to $W$ is called the **orthogonal complement** of $W$, denoted $W^\perp$ (pronounced $W$ perp). $W^\perp$ is a subspace of $\mathbb{R}^n$.

Interestingly enough, the row space's orthogonal complement is actually the null space. (Not the column space, because remember the row space and null space are both subspaces of $\mathbb{R}^n$, while the column space is a subspace of $\mathbb{R}^m$). Furthermore, the column space's orthogonal complement is actually just the transpose of the null space.

$$(\text{Row}\, A)^\perp = \text{Nul}\, A \quad \text{and} \quad (\text{Col}\, A)^\perp = \text{Nul}\, A^T$$

This means that if we have any vector $\vec{u}$ in $\mathbb{R}^n$, $\vec{w} \in W$, and $\vec{v} \in W^\perp$, then we can write $\vec{u}$ as a sum of $\vec{w}$ and $\vec{v}$: $\vec{u} = \vec{w} + \vec{v}$.

## Orthogonal sets

An **orthogonal set** is simply a set of vectors that are all orthogonal to each other. To be specific, a set of vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ constitutes an orthogonal set if for any two vectors in the set $\vec{v}_i$ and $\vec{v}_j$ their dot product is zero ($\vec{v}_i \cdot \vec{v}_j = 0$).

The simplest example of an orthogonal set is the standard basis $\mathcal{E}_3 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ (in this case in $\mathbb{R}^3$). It's evident from our above definition that these basis vectors are all orthogonal to each other.

Now, it turns out that the above orthogonal set being a basis is no coincidence. That's because all orthogonal sets are linearly independent and forms a basis for the subspace the orthogonal set spans.

In addition, we can actually backtrack and find each weight of a certain linear combination of an orthogonal set. If we have an orthogonal basis $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ that lives in the subspace $W$, then for each vector in the subspace $W$ (which we'll call $\vec{y}$), the weights of the linear combination are given by:

$$c_j = \frac{\vec{y} \cdot \vec{u}_j}{\vec{u}_j \cdot \vec{u}_j}$$

What we're essentially doing is compressing the vector $\vec{y}$ onto the same direction as the basis vector $\vec{u}_j$ and extracting the scalar multiple $c_j$ that is required to equate that component of $\vec{y}$ to $\vec{u}_j$.

This has extremely important ramifications. In the next next section, we'll discuss the projection operator.

## Orthonormal sets

If the orthogonal set is comprised of unit vectors, this is called an **orthonormal set**, and a **orthonormal matrix** is a square invertible matrix with orthonormal columns.

An $m \times n$ matrix $U$ is orthonormal if and only if $U^T U = I_n$. Furthermore, for an $n \times n$ square matrix $S$ that is an orthonormal matrix, we know that $S^{-1}S = I_n$, but because it is also orthonormal, $S^T S = I_n$ is also true. Therefore, for orthonormal matrices (which are square by definition), $S^{-1} = S^T$. We will use these properties later.

Properties of orthonormal sets:

- $||U\vec{x}|| = ||\vec{x}||$
- $(U\vec{x}) \cdot (U\vec{y}) = \vec{x} \cdot \vec{y}$
- $(U\vec{x}) \cdot (U\vec{y}) = 0 \iff \vec{x} \cdot \vec{y} = 0$

## Orthogonal projections onto lines

Imagine you have vector $\vec{b}$ flat on the $x$-axis and $\vec{a}$ pointing up at an angle of $60°$ from $\vec{b}$. Then the "shadow" that $\vec{a}$ casts upon the direction of $\vec{b}$ is called the **orthogonal projection** (we usually just call it a projection, though).

The projection is found by the following formula:

$$\text{proj}_{\vec{b}}\, \vec{a} = \left( \frac{\vec{a} \cdot \vec{b}}{\vec{b} \cdot \vec{b}} \right) \vec{b}$$

$\frac{\vec{a} \cdot \vec{b}}{\vec{b} \cdot \vec{b}}$ determines the magnitude, whereas the exact direction of the projection is kept from $\vec{b}$.

However, in linear algebra, saying that the projection is simply "in the direction of $\vec{b}$" won't cut it. So we have to use an even more specific definition. Instead, we define projections to be onto subspaces and we'll say that the projection is along the subspace that spans $\vec{b}$. (In this subsection, we will stick with a one-dimensional subspace, but in the next section we will explain orthogonal projections onto n-dimensional subspaces in $\mathbb{R}^n$.)

Let's redefine. Given that $L = \text{Span}\{\vec{b}\}$, an orthogonal projection is given by:

$$\text{proj}_{L}\, \vec{a} = \left( \frac{\vec{a} \cdot \vec{b}}{\vec{b} \cdot \vec{b}} \right) \vec{b}$$

Side note: In physics, you can also find parallel and normal forces by using projections.

## 6.3 Orthogonal projections onto subspaces

Earlier, we established that if you have any vector $\vec{u}$ in $\mathbb{R}^n$, $\vec{w} \in W$, and $\vec{v} \in W^\perp$, then we can write $\vec{u}$ as a sum of $\vec{w}$ and $\vec{v}$: $\vec{u} = \vec{w} + \vec{v}$. The projection of $\vec{u}$ onto the subspace $W$ is really just $\vec{w}$. The same thing applies to the projection of $\vec{u}$ onto $W^\perp$: it's $\vec{v}$.

$$\text{proj }_W \vec{u} = \vec{w} \text{ and } \text{proj }_{W^\perp} \vec{u} = \vec{v}$$

The projection operator is basically extracting the $\vec{w}$ component of the vector $\vec{u}$.

From this, we get the **Orthogonal Decomposition Theorem**: let $W \subseteq \mathbb{R}^n$ be a subspace. Then, you can write any vector $\vec{u}$ in the unique form

$$\vec{u} = \vec{w} + \vec{v}$$

where $\vec{w} \in W$ and $\vec{v} \in W^\perp$. $\vec{w}$ is then the orthogonal projection of $\vec{u}$ onto $W$ as we stated above.

In fact, if $\{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_p\}$ is any orthogonal basis of $W$, then you can write $\vec{u}$ in terms of this basis's projection onto $W$ as the sum of the projections of $\vec{w}$ onto $\vec{b}_i$:

$$\vec{w} = \frac{\vec{u} \cdot \vec{b}_1}{\vec{b}_1 \cdot \vec{b}_1}\vec{b}_1 + \cdots + \frac{\vec{u} \cdot \vec{b}_p}{\vec{b}_p \cdot \vec{b}_p}\vec{b}_p = \text{proj }_{\text{Span}\{\vec{b}_1\}} \vec{u} + \cdots + \text{proj }_{\text{Span}\{\vec{b}_p\}} \vec{u}$$

For a linear transformation, you can use the following formula:

$$\text{proj }_W \vec{u} = A(A^TA)^{-1}A^T\vec{x}$$

## 6.4 Gram–Schmidt process

The Gram–Schmidt process is a simple algorithm to orthogonalize any basis for any nonzero subspace in $\mathbb{R}^n$.

If we have any basis $\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_p\}$ for a nonzero subspace $W$ of $\mathbb{R}^n$, then $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_p\}$ is an orthogonal basis for $W$, where they are defined as follows:

$$\begin{array}{rcl}
\vec{v}_1 & = & \vec{x}_1 \\
\vec{v}_2 & = & \vec{x}_2 - \text{proj }_{\vec{x}_2} \vec{v}_1 \\
\vec{v}_3 & = & \vec{x}_3 - \text{proj }_{\vec{x}_3} \vec{v}_1 - \text{proj }_{\vec{x}_3} \vec{v}_2 \\
 & \vdots & \\
\vec{v}_p & = & \vec{x}_p - \text{proj }_{\vec{x}_p} \vec{v}_1 - \text{proj }_{\vec{x}_p} \vec{v}_2 - \cdots - \text{proj }_{\vec{x}_p} \vec{v}_{p-1}
\end{array}$$

This also means that $\text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\} = \text{Span}\{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k\}$ for $1 \le k \le p$. (In other words, $\text{Span}\{\vec{v}_1\} = \text{Span}\{\vec{x}_1\}$, as well as $\text{Span}\{\vec{v}_1, \vec{v}_2\} = \text{Span}\{\vec{x}_1, \vec{x}_2\}$, etc.)

(In this definition, we use the "primitive" definition of projections, defined as $\text{proj }_{\vec{b}} \vec{a} = \frac{\vec{a} \cdot \vec{b}}{\vec{b} \cdot \vec{b}}\vec{b}$)

To obtain an orthonormal basis from Gram–Schmidt, simply apply the unit vector formula to each basis vector in order to orthonormalize (as opposed to simply orthogonalizing) the basis.

### QR decomposition

$$A = QR$$

QR decomposition is basically converting a matrix $A$ into the equivalent matrix product $QR$ using the Gram–Schmidt process, where $Q$ is an orthonormal matrix (a matrix whose columns are an orthonormal basis) for $\operatorname{Col} A$ and $R$ is an upper triangular invertible matrix with positive entries on the diagonal.

Since $Q$ is orthonormal, you can find $R$ by applying the property that $Q^T Q = I_n$ (see "Orthonormal sets" for a refresher), and so $A = QR \implies Q^T A = Q^T QR = R$. Therefore, to determine $R$, simply do:

$$R = Q^T A$$

## 6.5 Symmetric matrices

A symmetric matrix is one that is the same when flipped across its diagonal. In other words, if a matrix $A = A^T$, then $A$ is symmetric.

For a diagonalized matrix $A = PDP^{-1}$, we can say that if $P$ is comprised of an orthonormal basis, then $P^{-1}$ can be replaced with $P^T$ for computational ease.

$$A = PDP^T \iff P \text{ is an orthogonal matrix}$$

- If $A$ is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.
- An $n \times n$ matrix $A$ is orthogonally diagonalizable if and only if $A$ is a symmetric matrix.

### The Spectral Theorem

Remember the spectrum is the set of all the eigenvalues of a matrix $A$. An $n \times n$ symmetric matrix $A$ has the following properties: (Lay)

- $A$ has $n$ real eigenvalues, including multiplicities.
- The dimension of $\lambda$'s eigenspace is equal to its multiplicity in the characteristic equation.
- The eigenspaces are orthogonal to each other (mutually orthogonal).
- $A$ is orthogonally diagonalizable.

### Spectral decomposition

We can write $A = PDP^T$ as $A = \lambda_1 \vec{u}_1 \vec{u}_1^T + \cdots + \lambda_n \vec{u}_n \vec{u}_n^T$. This is called the **spectral decomposition** of $A$. Every matrix $\vec{u}_i \vec{u}_i^T$ is a projection matrix because the vector $(\vec{u}_i \vec{u}_i^T)\vec{x}$ is the orthogonal projection of $\vec{x}$ onto the subspace spanned by $\vec{u}_i$. (Lay)

# Chapter 7

# Miscellaneous topics

## 7.1 Cross products

Cross products are the other way to multiply vectors. They are denoted as $\vec{u} \times \vec{v}$ and results in a vector perpendicular to both $\vec{u}$ and $\vec{v}$, which is why they are also sometimes called *vector products*.

The magnitude of a cross product is found by taking the area of the parallelogram whose sides are formed by the vectors.

For two vectors in $\mathbb{R}^2$, simply create a $2 \times 2$ matrix whose first column is $\vec{u}$ and the second column is $\vec{v}$. This matrix represents the linear transformation from the standard basis $\mathcal{E}$ to an arbitrary basis $\mathcal{B}$ comprised of $\vec{u}$ and $\vec{v}$. In essence, this matrix is the change-of-coordinates matrix from $\mathcal{E}$ to $\mathcal{B}$, or $\underset{\mathcal{B}\leftarrow\mathcal{E}}{P}$. Note that the area of the parallelogram whose sides are formed by the standard basis vectors $\vec{e}_1$ and $\vec{e}_2$ is 1. So the scaled area of this transformed parallelogram is greater than 1.

For two vectors in $\mathbb{R}^3$, you can use the following procedure to find the cross product using components of the vector. Assuming you are working with the standard basis $\mathcal{E}$, you take the determinant of a matrix with the top row as $\hat{\imath}, \hat{\jmath}, \hat{k}$. The result is then a vector. We can also make the first *column* the standard basis vectors and then the next two columns as $\vec{u}$ and $\vec{v}$, since $\det A = \det A^T$.

$$\vec{u} \times \vec{v} = \begin{vmatrix} \hat{\imath} & \hat{\jmath} & \hat{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \begin{vmatrix} \hat{\imath} & u_1 & v_1 \\ \hat{\jmath} & u_2 & v_2 \\ \hat{k} & u_3 & v_3 \end{vmatrix}$$

Note that the cross products results in a vector. The determinant gives you a scalar output, which is the magnitude of the cross product vector. However, its direction is unclear unless you use the **right-hand rule**.

Here's how you use the right-hand rule for $\vec{u} \times \vec{v}$. Start at the first operand ($\vec{u}$) with your fingers pointed straight. Then curl your fingers towards the second operand ($\vec{v}$). If your thumb points upwards, then the result of the cross product is also upwards. Similarly, if your thumb points downwards, the result is also downwards. This only works for 2D vectors being crossed though. For 3D vectors, you must use the determinant method.

If you have the magnitudes of your two vectors and the angle between them, you can simply use the following formula:

$$|\vec{u} \times \vec{v}| = ||\vec{u}||||\vec{v}|| \sin \theta$$

Cross products are **anticommunicative**, meaning that $\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$. In other words, order matters!!

There are different ways to calculate the determinant, so if you do not know how to calculate it, please research it.

Additionally, in $\mathbb{R}^2$, the magnitude of a cross product is equivalent to the area of the parallelogram made if the two operand vectors were placed tail-to-tail and formed a parallelogram. For $\mathbb{R}^3$, the magnitude of a cross product would be equivalent to a *parallelepiped*, which is the 3D version of a parallelogram.

This study guide recommends you watch the *Essence of Linear Algebra* video series for certain topics in linear algebra in order to get the utmost understanding. Below is the recommended video for this section:

Essence of Linear Algebra – Chapter 8: Cross products:
`https://www.youtube.com/watch?v=eu6i7WJeinw`

# Reference material

## Full Invertible Matrix Theorem

Let $A$ be an $n \times n$ square matrix. If any of the following statements pertaining to $A$ are true, then the rest of the statements must also be true.

1. $A$ is an invertible matrix.
2. $A$ is row equivalent to $I_n$, the $n \times n$ identity matrix.
3. $A$ has $n$ pivot positions. (i.e. there is a pivot in every row and every column).
4. The homogeneous matrix equation $A\vec{x} = \vec{0}$ has only the trivial solution.
5. The columns of $A$ form a linearly independent set.
6. The linear transformation $x \mapsto Ax$ is one-to-one.
7. The equation $A\vec{x} = \vec{b}$ has exactly one solution $\forall \vec{b} \in \mathbb{R}^n$. ($\forall$ means "for all".)
8. The columns of $A$ span $\mathbb{R}^n$.
9. The linear transformation $x \mapsto Ax$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.
10. There is an $n \times n$ matrix $C$ so that $CA = I_n$.
11. There is an $n \times n$ matrix $D$ so that $AD = I_n$. $C = D$ always.
12. $A^T$ is an invertible matrix.
13. The columns of $A$ form a basis for $\mathbb{R}^n$.
14. $\mathrm{Col}\, A = \mathbb{R}^n$.
15. $\mathrm{rank}\, A = n$.
16. $\mathrm{Nul}\, A = \{\vec{0}\}$.
17. $\mathrm{nullity} \ = 0$.
18. $\mathrm{Row}\, A = \mathbb{R}^n$.
19. The number 0 is NOT an eigenvalue of $A$. $(0 \notin \lambda)$
20. The determinant of $A$ is NOT zero. $(\det A \neq 0)$
21. The orthogonal complement of the column space is $\{\vec{0}\}$.
22. The orthogonal complement of the null space is $\mathbb{R}^n$.
23. The matrix $A$ has $n$ nonzero singular values.