Math 308: A First Course in Linear Algebra

Sam Roven

Contents

Chapter 1

Systems of Linear Equations

1.1 Solving Linear Equations

We have all seen a linear system of equations at some point in gradeschool, and we first learned how to attack these systems using the methods of substitution and elimination. We begin with a refreshing example of a linear system with **three** equations and **three** unknowns.

Example 1.1.1. $$\begin{cases} x_1 + x_2 - x_3 = 7 \\ 2x_1 + 3x_3 = 5 \\ -5x_2 = -10 \end{cases}$$

We define a solution of this system as an ordered triple of real numbers, $\underbrace{(x_1, x_2, x_3)}_{\text{also called a 3-tuple}}$, which simultaneously satisfies **all** equations.
One solution of this system is $(4, 2, -1)$ because

$$4 + 2 - (-1) = 7$$

$$2(4) + 3(-1) = 5$$

and

$$-5(2) = -10$$

It is also worth noting that this solution is the same thing as an ordinary point in 3-dimensional Euclidean space, and we are immediately able to talk about geometry (much more to come).

Next, we must get our hands around the vocabulary of linear systems, the first of which is distinguishing one type of variable from another.

Definition 1.1.2. A variable that appears as the first (left-most) term of at least one equation is a **leading variable**. In the above example, x_1 and x_2 are leading variables.

Definition 1.1.3. If a linear system has no solutions, then it is **inconsistent**. If a linear system has at least one solution, then it is **consistent**.

Example 1.1.4. $$\begin{cases} 2x_1 - 3x_2 + x_3 = 8 \\ 2x_2 = 5 \end{cases}$$

This is an example of a linear system with infinitely many solutions. In fact, for any real number t, the tuple

$$(\frac{31}{4} - \frac{1}{2}t, \frac{5}{2}, t)$$

represents a solution and we can verify directly that it is a solution by plugging in $\frac{31}{4} - \frac{1}{2}t$ for x_1, $\frac{5}{2}$ for x_2, and t for x_3 and checking that all the t's cancel to give equality.

Here, t is called a **free variable** or **free parameter**.

Now that we have some language to work with, we will need to investigate the possible forms a system can have. In particular, there are two.

1. Triangular form: An example of a linear system in triangular form is

$$\begin{cases} 4x_1 - 2x_2 + 3x_3 + \ \ x_4 = 17 \\ \qquad\ \ x_2 - 2x_3 - \ \ x_4 = 0 \\ \qquad\qquad\ \ 5x_3 + 2x_4 = 20 \\ \qquad\qquad\qquad\quad 3x_4 = 15 \end{cases}$$

   We can solve a system like this using **back substitution** (using the last equation first). In doing this we see that

$$3x_4 = 15 \implies x_4 = 5$$

   We then apply this to the third equation and get

$$5x_3 + 2x_4 = 5x_3 + 2(5) = 20 \implies 5x_3 = 10 \implies x_3 = 2$$

   Applying the same procedure to the second and first equation we find that $x_2 = 9$ and $x_1 = 6$ (you should verify this for yourself!). The final solution is then given by

$$(x_1, x_2, x_3, x_4) = (6, 9, 2, 5)$$

   In general, triangular forms have three main properties:

   - There are the same number of equations as variables.
   - Every variable is the leading variable of exactly one equation.
   - A triangular system has **exactly one solution**. We refer to this as a unique solution.

2. Echelon Form

   This is the more general form that a linear system can have and we can characterize it according to two (or three) main properties:

   - Every variable is the leading variable of at most one equation.
   - The system is organized in a descending stair-step pattern.

     If a linear system satisfies both of these properties then we say the system is in **echelon form**. The last porperty of a system in echelon form is
   - There are either no solutions, exactly one solution, or infinitely many solutions.

To build off of the last point, we can actually say something more general.

Theorem 1.1.5. *Any system of linear equations has either*

- *No solutions (this is known as an* ***inconsistent*** *linear system).*
- *Exactly one solution.*
  or
- *Infinitely many solutions.*

The latter two cases define what we call a **consistent** linear system.

Example 1.1.6. The following linear system **is** in echelon form.

$$\begin{cases} 3x_1 \qquad - x_3 = 7 \\ \qquad x_2 \qquad = 10 \end{cases}$$

Example 1.1.7. The following linear system **is not** in echelon form because the linear equations do not form a stair-step pattern.

$$\begin{cases} 3x_1 + x_2 - \quad x_3 = 7 \\ \qquad\qquad\quad x_3 = 5 \\ \qquad x_2 + 9x_3 = 11 \end{cases}$$

Example 1.1.8. The following linear system **is not** in echelon form because x_1 is the leading variable of more than one equation.

$$\begin{cases} 3x_1 \qquad\quad - x_3 = 7 \\ x_1 + x_2 \qquad = 9 \end{cases}$$

Definition 1.1.9. For a system in echelon form, any variable that does not appear as a leading variable is called a **free variable**, hence all variables in a system are either leading or free.

Here are some nice facts to remember about systems in echelon form.

1. If an echelon system has no free variables, it must be triangular and therefore has exactly one solution.
2. If an echelon system has at least one free variable, then it has infinitely many solutions.

Now that we have much of the needed vocabulary, lets end the section with a fully worked example.

Example 1.1.10. Consider the linear system

$$\begin{cases} 2x_1 - x_2 + 5x_3 - x_4 = -30 \\ \qquad\qquad\quad x_3 + x_4 = -6 \end{cases}$$

This is a system in echelon form with x_1, x_3 as leading variables and x_2, x_4 as free variables. We solve the system in two steps.

Step 1: Denote free variables. Let $x_2 = t_1$ and $x_4 = t_2$ and remember these can be any real number!

Step 2: Plug the free variables into the system and solve for leading variables. Starting with the second equation we have

$$x_3 + t_2 = -6 \implies x_3 = -t_2 - 6$$

Plugging this into the first equation we have

$$2x_1 - t_1 + 5(-t_2 - 6) - t_2 = -30 \implies 2x_1 - t_1 - 5t_2 - t_2 = 0 \implies 2x_1 - t_1 - 6t_2 = 0$$

Using this to solve for x_1 we get

$$2x_1 = t_1 + 6t_2 \implies x_1 = \frac{1}{2}t_1 + 3t_2$$

The (infinitely many) solutions of this linear system have the form

$$(x_1, x_2, x_3, x_4) = (\frac{1}{2}t_1 + 3t_2, t_1, -t_2 - 6, t_2)$$

with t_1, t_2 as free variables.

1.2 Linear Systems and Matrices

In this section we dive deeper into the procedures for solving linear systems and in the process, encounter matrices for the first time. These procedures will transform any linear system into one in echelon form and produce a new linear system with the exact same solution set.

Definition 1.2.1. Two linear systems are **equivalent** if they have the same solution set. The notion of being equivalent is denoted with the symbol "$\sim$".

The way in which we get from an arbitrary linear system to an echelon one is by applying **elementary row operations**. These consist of three possible "moves" that transform a system into an equivalent one:

1. Interchange two equations.
2. Replace one equation with a non-zero multiple of itself.
3. Add one equation to a multiple of another.

Example 1.2.2. $\begin{cases} -4x_1 + 5x_2 = 20 \\ x_1 - 2x_2 = 14 \end{cases}$

$\sim \begin{cases} x_1 - 2x_2 = 14 \\ -4x_1 + 5x_2 = 20 \end{cases}$ (interchange equations)

$\sim \begin{cases} 4x_1 - 8x_2 = 56 \\ -4x_1 + 5x_2 = 20 \end{cases}$ (multiply equation 1 by 4)

$\sim \begin{cases} 4x_1 - 8x_2 = 56 \\ \quad\; - 3x_2 = 76 \end{cases}$ (add equation 1 to equation 2)

Notice the last (equivalent) system is in echelon form!

This example illustrates the general procedure, but the main tool that we use to streamline the procedure is that of augmented matrices. When we solve a linear system, we are only working with the coefficients of the linear equations, so we place the coefficients in an array called an **augmented matrix**.

Example 1.2.3. The linear system

$$\begin{cases} x_1 - 2x_2 + 3x_3 = 9 \\ -x_1 \qquad\;\; + 3x_3 = -4 \\ 2x_1 - 5x_2 + 5x_3 = 17 \end{cases}$$

has associated augmented matrix given by

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 9 \\ -1 & 0 & 3 & -4 \\ 2 & -5 & 5 & 17 \end{array}\right]$$

We can now translate vocabulary from linear systems into that for matrices. Once we do this we will never look back. Similar to that of linear systems, there are two special types of matrices.

1. Echelon Form.

   - Every leading term (the first nonzero number in a row) is in a column to the **left** of the leading term of the row below it.
   - Any zero rows (rows of all zeroes) are at the bottom.

   In general, we call any leading term of a non-zero row a **pivot**.

   Example 1.2.4.

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

   is a matrix in echelon form with pivots being the entries 3 and 1.

   The matrices

   $$\begin{bmatrix} 0 & 1 & 0 & 3 \\ 4 & 5 & 6 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

   are all not in echelon form. Can you see why?

2. Reduced Echelon Form.

   - It is in echelon form.
   - All pivot positions contain a 1.
   - All other entries in a <u>pivot column</u>(a column that contains a pivot) are 0.

   Example 1.2.5. $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is a matrix that is in reduced echelon form.

   $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 1 \end{bmatrix}$ is neither in echelon nor reduced echelon form.

   $\begin{bmatrix} 2 & 0 & 1 & 3 \\ 0 & -1 & 1 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ is in echelon form but **not** in reduced echelon form.

   $\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 3 & 4 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ is also in echelon form but **not** in reduced echelon form.

When working through a solution to a linear system, we can easily follow our own steps by adopting the following notation for row operations.

1. Interchange row i and row j is denoted

$$R_i \leftrightarrow R_j$$

2. Replacing row i with a non-zero multiple (c) times row j is denoted

$$cR_i \to R_i$$

3. Adding a non-zero multiple of row i to row j and applying the change to row j is denoted

$$cR_i + R_j \to R_j$$

In practice, we will use these row operations to transform augmented matrices into systems that are in echelon or reduced echelon form, at which point we will be able to solve them by back substitution.

This whole process will be most easily learned via examples so let's jump right in with a continuation of Example 1.2.3.

Example 1.2.6.

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 9 \\ -1 & 0 & 3 & -4 \\ 2 & -5 & 5 & 17 \end{array}\right]$$

$$(R_1 + R_2 \to R_2) \implies \left[\begin{array}{ccc|c} 1 & -2 & 3 & 9 \\ 0 & -2 & 6 & 5 \\ 2 & -5 & 5 & 17 \end{array}\right]$$

$$(-2R_1 + R_3 \to R_3) \implies \left[\begin{array}{ccc|c} 1 & -2 & 3 & 9 \\ 0 & -2 & 6 & 5 \\ 0 & -1 & -1 & -1 \end{array}\right]$$

$$(-\frac{1}{2}R_2 \to R_2) \implies \left[\begin{array}{ccc|c} 1 & -2 & 3 & 9 \\ 0 & 1 & -3 & -5/2 \\ 0 & -1 & -1 & -1 \end{array}\right]$$

$$(R_2 + R_3 \to R_3) \implies \underbrace{\left[\begin{array}{ccc|c} 1 & -2 & 3 & 9 \\ 0 & -2 & 6 & 5 \\ 0 & 0 & -4 & -7/2 \end{array}\right]}_{\text{echelon form!}}$$

This matrix represents the (triangular) linear system

$$\begin{cases} x_1 - 2x_2 + 3x_3 = 9 \\ \qquad x_2 + 6x_3 = -5/2 \\ \qquad\quad - 4x_3 = -7/2 \end{cases}$$

hence we can use back substitution to obtain the (unique) solution

$$(x_1, x_2, x_3) = (-113/8, -41/4, 7/8)$$

Definition 1.2.7. The process of using row operations (like above) to transform a matrix into echelon form is called **Gaussian Elimination**.

We can take this one step further, if we prefer, by reducing the given matrix to **reduced** echelon form. This is known as **Gauss-Jordan Elimination**.

Example 1.2.8. Use Gauss-Jordan elimination to solve the linear system

$$\begin{cases} x_1 \qquad - 3x_3 = -2 \\ 3x_1 + \ x_2 - 2x_3 = 5 \\ 2x_1 + 2x_2 + \ x_3 = 4 \end{cases}$$

We begin with the augmented matrix for this linear system and write a string of equivalent matrices, ending with the reduced echelon form. We leave the row operations to be determined by the reader.

$$\left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 3 & 1 & -2 & 5 \\ 2 & 2 & 1 & 4 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 0 & 1 & 7 & 11 \\ 0 & 2 & 7 & 8 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 0 & 1 & 7 & 11 \\ 0 & 0 & -7 & -14 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 0 & 1 & 7 & 11 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

We note here that at this point we could stop and use back substitution, we have performed Gaussian elimination and have arrived at the echelon form. Continuing onward we have

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

Translating back to the linear system, we have the (unique) solution

$$(x_1, x_2, x_3) = (4, -3, 2)$$

Before ending the chapter, we note that there is a methodical way to clear out entries of augmented matrices, starting in the upper left corner, moving down column 1, then to the entry in the second column and second row, then down the entire second column, etc. Having a methodical approach to row reductions will reduce errors and make row reductions much easier with a little practice.

Chapter 2

Euclidean Space

We now translate from the algebraic nature of linear systems to their underlying geometry. We begin with a quick refresher on vectors and Euclidean space, then spend the majority of the chapter introducing the all important notions of span and linear independence.

2.1 Vectors

Vectors are the fundamental object of linear algebra and we will use them frequently.

Definition 2.1.1. A **vector** is an ordered list of real numbers that can be expressed in two ways:

- Column vector

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

- Row vector

$$\mathbf{u} = (u_1, u_2, \ldots, u_n)$$

  We will use column vectors most of the time, but it is good to know that both notations can mean the same thing.

Just like with real numbers, we can perform arithmetic with vectors. Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ with c a real number (also known as a scalar).

We can multiply vectors by scalars as follows

$$c\mathbf{u} = \begin{bmatrix} cu_1 \\ \vdots \\ cu_n \end{bmatrix}$$

We can also add two vectors, as long as they have the same number of coordinates.

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

Lastly, $\mathbf{u} = \mathbf{v}$ if and only if $u_1 = v_1, u_2 = v_2, \ldots, u_n = v_n$.

Definition 2.1.2. The set of all vectors with n entries (components), together with the above operations of scalar multiplication and vector addition, form what is known as n-dimensional Euclidean space. We denote this space by $\mathbb{R}^n$. For the vectors $\mathbf{u}$ and $\mathbf{v}$ defined above, we use the symbol "$\in$" to denote that the vector **lives in** $\mathbb{R}^n$. Similarly, since the scalar c is a real number, it **lives in** the set of real numbers, which we denote by writing $c \in \mathbb{R}$. We will use this notation **frequently** from now on.

In $\mathbb{R}^2$ and $\mathbb{R}^3$ we usually represent vectors with"arrows". The previous three vector properties can also be expressed geometrically.

- Two vectors are equal if and only if they have the same **length** and point in the same **direction**.
- Given a vector $\mathbf{u}$, the vector $c\mathbf{u}$ (for $c \neq 0$, and $c \in \mathbb{R}$) is parallel to $\mathbf{u}$, with length equal to $|c|$ times the length of $\mathbf{u}$. Multiplying a vector by a negative scalar switches the direction that it points in.
- Given $\mathbf{u}, \mathbf{v} \in \mathbb{R}$, the vector $\mathbf{u}+\mathbf{v}$ can be found by using the usual parallelogram law (or tip-to-tail rule) from calculus 3.

Now that we have the fundamentals refreshed, we can move onto one of the central topics of the course.

Definition 2.1.3. If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \in \mathbb{R}^n$ and $c_1, c_2, \ldots, c_m \in \mathbb{R}$ then the vector

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_m\mathbf{u}_m$$

is called a **linear combination** of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$.

Example 2.1.4. Given the vectors $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$, three different linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$ are

$$\mathbf{u}_1 + \mathbf{u}_2 = \begin{bmatrix} 6 \\ -1 \end{bmatrix}, \mathbf{u}_1 - \mathbf{u}_2 = \begin{bmatrix} -4 \\ 5 \end{bmatrix}, 2\mathbf{u}_1 + 30\mathbf{u}_2 = \begin{bmatrix} 152 \\ -86 \end{bmatrix}$$

A very important idea tied to linear combinations is finding when a given vector is a linear combination of a fixed set of vectors.

Example 2.1.5. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} .\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$ and determine if $\mathbf{b} = \begin{bmatrix} 19 \\ 7 \\ -9 \end{bmatrix}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

When approaching a question like this one, starting the problem is often the hardest part. How in the world can we figure this out? We figure it out by assuming it is true and following our nose until we arrive at two possible outcomes. Either we find a solution and we are done or the system is inconsistent and we see that there is no such linear combination. The starting point of this problem is **the most important thing we will learn this far**.

If $\mathbf{b}$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ then **there exist scalars** $c_1, c_2, c_3 \in \mathbb{R}$ such that

$$c_1 \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 19 \\ 7 \\ -9 \end{bmatrix}$$

This is how we always approach these problems. We find values for the c_i or realize that they cannot exist. The way in which we find the c_i is by unpacking what it means for two vectors to be equal. Using vector addition on the left hand side of the equation we get that

$$\begin{bmatrix} c_1 + 2c_2 + 5c_3 \\ c_2 + 2c_3 \\ -2c_1 + c_2 - c_3 \end{bmatrix} = \begin{bmatrix} 19 \\ 7 \\ -9 \end{bmatrix}$$

which translates to the linear system $\begin{cases} c_1 + 2c_2 + 5c_3 = 19 \\ \qquad c_2 + 2c_3 = 7 \\ -2c_1 + c_2 - c_3 = -9 \end{cases}$

We solve this linear system by solving the corresponding augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 2 & 5 & 19 \\ 0 & 1 & 2 & 7 \\ -2 & 1 & -1 & -9 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 4 \end{array}\right]$$

Note that all row reducing from now on will not be explicitly worked out. You are an expert row reducer and you can work it out yourself!

Gauss-Jordan elimination here tells us that

$$(c_1, c_2, c_3) = (2, -1, 4)$$

hence

$$\mathbf{b} = 2\mathbf{v}_1 - \mathbf{v}_2 + 4\mathbf{v}_3$$

and we are done!

This example illustrated the best case scenario, that is, we wonder if a fixed vector is a linear combination of some others, and we directly find the coefficients that give us the desired linear combination. If such a linear combination does not exist, we unravel a different conclusion.

Example 2.1.6. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix}$. Is $\mathbf{b}$ a linear combination of $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$?

Just like we did in the previous example, we set up the corresponding linear system as if there did exist such a linear combination. We then proceed by attempting to solve the linear system. In this case we get

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

The equivalent matrix we have found represents an **inconsistent** linear system, therefore $\mathbf{b}$ is **not** a linear combination of $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$

Before ending this section we make one last use of vector notation by expressing solution sets in terms of linear combinations.

Example 2.1.7. Suppose we have the following linear system $\begin{cases} 4x_1 - 2x_2 + x_3 - x_4 = -5 \\ \qquad\qquad x_3 + x_4 = 1 \end{cases}$

This linear system will have infinitely many solutions because there are two free variables. We can express all such solutions in a compact way.

We first label the free variables, namely, $x_2 = t_1$ and $x_4 = t_2$. Then, using the second equation we get that

$$x_3 = 1 - t_2$$

Plugging all of this back into the first equation we see that

$$x_1 = \frac{-6 + 2t_1 + 2t_2}{4} = -\frac{3}{2} + \frac{1}{2}t_1 + \frac{1}{2}t_2$$

We then express this general solution in vector form by grouping together free variables, that is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} + \frac{1}{2}t_1 + \frac{1}{2}t_2 \\ t_1 \\ 1 - t_2 \\ t_2 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} \\ 0 \\ 1 \\ 0 \end{bmatrix} + t_1 \begin{bmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{bmatrix} + t_2 \begin{bmatrix} \frac{1}{2} \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

The expression of a solution set in terms of a linear combination of vectors is known as the **general solution in vector form**.

2.2 Span

In this section we dig deeper into the question "Can we express one vector as a linear combination of others?" Geometrically, this is the same as asking if we can travel to a point in space, by moving along fixed directions. For example, suppose we were a little dot in $\mathbb{R}^2$, located at the origin, and we wanted to find a path to the point (a, b) **but** we could only move along a line with slope 1 or slope zero, i.e. we can only move parallel to the line $y = x$ or horizontally. By translating this into the language of linear algebra, we are asking if the point (a, b) can be expressed as a linear combination of $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

Can we get to the point $(5, 3)$? Yes!

$$3\mathbf{u}_1 + 2\mathbf{u}_2 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

hence the vector $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$ is a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$.

Example 2.2.1. We can extend the question to an arbitrary point in $\mathbb{R}^2$. That is, can we express any vector $\begin{bmatrix} a \\ b \end{bmatrix}$ (for $a, b \in \mathbb{R}$) as a linear combination of $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$? If we could, then there would exist scalars $x_1, x_2 \in \mathbb{R}$ auch that

$$\begin{bmatrix} a \\ b \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Finding such values of the x_i is equivalent to solving the linear system with augmented matrix

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

By performing Gauss-Jordan elimination we see that

$$\left[\begin{array}{cc|c} 1 & 1 & a \\ 1 & 0 & b \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 1 & a \\ 0 & -1 & b-a \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 1 & a \\ 0 & 1 & a-b \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 0 & b \\ 0 & 1 & a-b \end{array}\right]$$

hence $x_1 = b$ and $x_2 = a - b$. In other words

$$\begin{bmatrix} a \\ b \end{bmatrix} = b \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (a-b) \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and we can write **any** vector in $\mathbb{R}^2$ as a linear combination of $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. This example lends itself to the central object of this section.

Definition 2.2.2. Suppose $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{R}^n$. The **span** of the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ denoted $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$, is the set of all linear combinations of $\mathbf{u}_1, \ldots, \mathbf{u}_m$. In other words, $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ consists of all vectors of the form

$$\mathbf{v} = x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \cdots + x_m\mathbf{u}_m$$

for some scalars $x_1, x_2, \ldots, x_n \in \mathbb{R}$.

If $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} = \mathbb{R}^n$ we say that $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ **spans** $\mathbb{R}^n$.

Note that in the above example we showed that $\text{Span}\left\{ \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix} \right\} = \mathbb{R}^2$ so $\left\{ \begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix} \right\}$ spans $\mathbb{R}^2$. Now lets look at some more examples.

Example 2.2.3. Let $\mathbf{u}_1 = \begin{bmatrix}1\\2\\0\end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix}2\\-1\\1\end{bmatrix}$ and $\mathbf{u}_3 = \begin{bmatrix}4\\0\\1\end{bmatrix}$. Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ spans $\mathbb{R}^3$.

Let $\mathbf{v} = \begin{bmatrix}a\\b\\c\end{bmatrix}$ denote an arbitrary vector in $\mathbb{R}^3$. We need to show that there always exist scalars $x_1, x_2, x_3 \in \mathbb{R}$ such that

$$x_1 \begin{bmatrix}1\\2\\0\end{bmatrix} + x_2 \begin{bmatrix}2\\-1\\1\end{bmatrix} + x_3 \begin{bmatrix}4\\0\\1\end{bmatrix} = \begin{bmatrix}a\\b\\c\end{bmatrix}$$

i.e. $\mathbf{v}$ is a linear combination $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$.

In trying to solve the system corresponding to the vector equation above we see that

$$\left[\begin{array}{ccc|c}1 & 2 & 4 & a\\2 & -1 & 0 & b\\0 & 1 & 1 & c\end{array}\right] \sim \left[\begin{array}{ccc|c}1 & 2 & 4 & a\\0 & -5 & -8 & b-2a\\0 & 1 & 1 & c\end{array}\right] \sim \left[\begin{array}{ccc|c}1 & 2 & 4 & a\\0 & 1 & 1 & c\\0 & -5 & -8 & b-2a\end{array}\right]$$

$$\sim \left[\begin{array}{ccc|c}1 & 2 & 4 & a\\0 & 1 & 1 & c\\0 & 0 & -3 & b-2a+5c\end{array}\right] \sim \left[\begin{array}{ccc|c}1 & 2 & 4 & a\\0 & 1 & 1 & c\\0 & 0 & 1 & \frac{b-2a+5c}{-3}\end{array}\right]$$

From here we can use back subtitution and solce for x_1, x_2, and x_3 which means that $\begin{bmatrix}a\\b\\c\end{bmatrix} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ for **every** vector $\begin{bmatrix}a\\b\\c\end{bmatrix} \in \mathbb{R}^3$. This precisely means that $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \mathbb{R}^3$.

Example 2.2.4. Let $\mathbf{u}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix}2\\4\\-3\end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix}2\\2\\5\end{bmatrix}$. Is $\mathbf{v} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$?

We try to solve the vector equation $x_1\mathbf{u}_1 + x_2\mathbf{u}_2 = \mathbf{v}$ by looking at the augmented matrix $\left[\begin{array}{cc|c}\mathbf{u}_1 & \mathbf{u}_2 & \mathbf{v}\end{array}\right]$.

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

The third line of the last equivalent matrix translates to the equation $0 = 3$ hence the linear system is inconsistent! This means there are **no** scalars $x_1, x_2 \in \mathbb{R}$ such that $x_1\mathbf{u}_1 + x_2\mathbf{u}_2 = \mathbf{v}$, hence $\mathbf{v} \notin \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

To recap, we have seen three vectors that spanned $\mathbb{R}^3$ and two vectors that did not span $\mathbb{R}^3$. It turns out that no two vectors in $\mathbb{R}^3$ will ever be able to span $\mathbb{R}^3$, we will actually need at least 3. Will any three vectors span $\mathbb{R}^3$ or do we need to choose them more carefully? The next example tells us that we must choose them more carefully.

Example 2.2.5. Let $\mathbf{u}_1 = \begin{bmatrix}1\\1\\1\end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix}2\\4\\-3\end{bmatrix}$ and $\mathbf{u}_3 = \begin{bmatrix}9\\13\\-1\end{bmatrix}$. Is $\mathbf{v} = \begin{bmatrix}4\\-2\\3\end{bmatrix} \in \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$?

Performing row operations reveals that

$$\left[\begin{array}{ccc|c}1 & 2 & 9 & 4\\1 & 4 & 13 & -2\\1 & -3 & -1 & 3\end{array}\right] \sim \left[\begin{array}{ccc|c}1 & 2 & 9 & 4\\0 & 1 & 2 & -3\\0 & 0 & 0 & -16\end{array}\right]$$

This means that $v = \begin{bmatrix}4\\-2\\3\end{bmatrix} \notin \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ hence any random set of three vectors will not always span $\mathbb{R}^3$. We can drill down the needed specifications a bit more in the following proposition.

Proposition 2.2.6. *Suppose $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^n$.*

- *If $m < n$, then $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ does not span $\mathbb{R}^n$.*
- *If $m \geq n$, then $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$* ***may*** *or* ***may not*** *span $\mathbb{R}^n$ (we have seen that both cases are possible when $m = n$.*

This proposition prompts further investigation on how two spans are related. We will begin this investigation by proving another proposition, and before we do, we lay out some foundational ideas surrounding proofs.

2.2.1 Some modern math techniques

We begin by recalling the definition of a span of a set of vectors. Given vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ the span of these vectors, written as $span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ is the set of all linear combinations of the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.

Tying this into what we mentioned above, we can see that the span of a set of vectors is a set! What does it mean for something to be an element of this set? For this (and all other sets we encounter), being an element of a given set means the element in question satisfies the definiton of what it means to be in that set. Stated in the context of span, a vector $\mathbf{v}$ is in the span of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, written as

$$\mathbf{v} \in span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$$

if $\mathbf{v}$ is a linear combination of the $\mathbf{u}_i$ for $i = 1, 2, \ldots, n$. Digging a little further, we can apply the definition and write

$$\text{If } \mathbf{v} \in span(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n), \text{ then there exist } c_1, \ldots, c_n \text{ such that}$$

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

In proving things about spans, we will constantly come back to this definition, and in general. you should remember that being an element of a set generally involves looking at the definition of what it means to be in that set. This is a very common starting point for many proofs. It should help you get your mind moving and prevent you from getting too stuck.

Some remarks on proofs

Although I don't plan to discuss proofs very much in this course, there are several basic techniques that you will be required to know. They are:
1) Knowing how to show that two sets are equal (in particular we will apply this to spans)

2) The implications of what an "if and only if" statement means.

1) By definition, two sets, A and B, are equal if any element of A is also an element of B, and similarly, every element of B is an element of A. If only one of these conditions holds, say every element of A is an element of B, but not every element of B is an element of A, then we say A is a *subset* of B and write $A \subset B$. Since a span of a set of vectors is a set, we will be interested in showing that two spans are equal.

The key idea is to take an arbitrary element of one set, and show it belongs to the other, then repeat the process in the other direction. Using the notation above we can write out this process in a series of steps.

i) Pick an arbitrary element $a \in A$, and show that $a \in B$. This means that $A \subset B$.
ii) Pick an arbitrary element $b \in B$ and show that $b \in A$. This shows that $B \subset A$.

To summarize, we have that $A = B$ if and only if $A \subset B$ *and* $B \subset A$. Now we explain a short bit about if and only if statements, then illustrate the above proof method with an example.

2) For if and only if statements there is not much to know. The one take away is that you have 4 useful statements that come out of it. If P and Q are two facts, say P is the fact that all cats are black and Q is the fact that all dogs are brown, then P if and only if Q (also written as P $\Leftrightarrow$ Q, or P iff Q) means that all cats are black if and only if all dogs are brown. The 4 statements that we can get out of this come from breaking down the statement into parts.

If we have that P if and only if Q, then this means that
i) If P is true then Q is true (also written as P $\implies$ Q).
ii) If Q is true then P is true (also written as Q $\implies$ P).
iii) If P is false, then Q is false.
iv) If Q is false, then P is false.

Note that the last two statements are the negation of the first two (if this confuses you then just ignore it).

One last thing worth mentioning is what it means if we have a series of statements A,B,C and there is a theorem saying
The following are equivalent:
i) A
ii) B
iii) C

What does this mean? Well the statement "the following are equivalent" means that the statements that follow can all be stated with if and only iff statements between them. The above example then reads as A if and only if B if and only if C. We can pick apart these however we please, i.e. since A if and only if B, then in particular, B implies A.

Taking an if and only if statement in the context of linear algebra, we can see how the four statements can give us different results. Recall the following theorem:

Proposition 2.2.7. *Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ be vectors in $\mathbb{R}^n$ (we could also write $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n \in \mathbb{R}^n$). Then the following statements are equivalent:*
i) $\boldsymbol{b}$ is in span$\{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n\}$
ii) The vector equation $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$ has at least one solution.

Unpacking all of this we have that:

$$\mathbf{b} \in span\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\} \Leftrightarrow x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n \text{ has at least one solution}$$

From this we get the four statements
i) If $\mathbf{b} \in span\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ then $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ has at least one solution.

ii)If $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ has at least one solution, then $\mathbf{b} \in span\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$.

iii)If $\mathbf{b} \notin span\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ (i.e. if the vector $\mathbf{b}$ is NOT in the span, then $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ has NO solutions.

iv) If $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ has at NO solutions, then $\mathbf{b} \notin span\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$.

Now that we're a bit more familiar with if and only if statements, let's finish off with a concrete example of a proof that the spans of two different sets of vectors are equal. Remember that spans of vectors are still sets! This means that showing equality of spanning sets is done in the same way that we show equality of sets.

<u>Example:</u>

Prove that

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

To avoid writing the above vectors as much we let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

We begin by showing that

$$span\{\mathbf{v}_1, \mathbf{v}_2\} \subset span\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

Let $\mathbf{x} \in span\{\mathbf{v}_1, \mathbf{v}_2\}$. This means that there exist scalars a_1, a_2 such that

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 = a_1 \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

This is the "unraveling the definition part".

Now it will be super useful to remember that when we say linear combination, we can include 0 as a scalar! This will prove to be a handy trick and in this context means that

$$\mathbf{x} = a_1 \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + 0\mathbf{v}_3$$

So we just wrote $\mathbf{x}$ as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$! Thus, $x \in span\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ hence we have shown that $span\{\mathbf{v}_1, \mathbf{v}_2\} \subset span\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

Now what remains to show is the other direction, namely that $span\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \subset span\{\mathbf{v}_1, \mathbf{v}_2\}$ and we apply the same procedure. Letting $\mathbf{x} \in span\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ this means that there exist scalars b_1, b_2, b_3 such that

$$\mathbf{x} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 = b_1 \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + b_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Now, we need to show that $\mathbf{x} \in span\{\mathbf{v}_1, \mathbf{v}_2\}$ so how can we do this? Well, showing that $\mathbf{v}_3$ is a linear combination of the other two will allow us to write $\mathbf{x}$ as a linear combo ONLY in $\mathbf{v}_1$ and $\mathbf{v}_2$. So lets try and do that. (You may see this method and think, "how in the world was I supposed to think of that?!", but while seeing it now may seem foreign, you will be doing this trick several times and it will seem less crazy each time).

We want to write $\mathbf{v}_3$ as a linear combo of $\mathbf{v}_1$ and $\mathbf{v}_2$, so lets take a look at what that linear combination would look like. It would give us some scalars a_1, a_2 such that

$$\begin{bmatrix}1\\0\\1\end{bmatrix} = a_1 \begin{bmatrix}1\\-2\\3\end{bmatrix} + a_2 \begin{bmatrix}0\\1\\-1\end{bmatrix}$$

so all we need to do is FIND the scalars. We do this by looking at the components of each vector and deducing what the scalars MUST be in order for the above equation to hold. Let's zoom in on the first components. For the above equality to hold, this must give us the equation

$$1 = a_1 \cdot 1 + a_2 \cdot 0 = a_1$$

hence we need to hace $a_1 = 1$. Now lets look at the second components, assuming we've found a_1 this reduces to the equation

$$0 = 1 \cdot -2 + a_2$$

hence $a_2 = 2$. We can look at the third component and verify that indeed $a_1 = 1, a_2 = 2$ give us

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

So we have the desired linear combo. Plugging this into the original linear combination that we started with, we see that

$$\mathbf{x} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3 = b_1 \begin{bmatrix}1\\-2\\3\end{bmatrix} + b_2 \begin{bmatrix}0\\1\\-1\end{bmatrix} + b_3 \begin{bmatrix}1\\0\\1\end{bmatrix}$$

$$= b_1 \begin{bmatrix}1\\-2\\3\end{bmatrix} + b_2 \begin{bmatrix}0\\1\\-1\end{bmatrix} + b_3 \left(\begin{bmatrix}1\\-2\\3\end{bmatrix} + 2 \begin{bmatrix}0\\1\\-1\end{bmatrix} \right) = (b_1 + b_3) \begin{bmatrix}1\\-2\\3\end{bmatrix} + (b_2 + 2) \begin{bmatrix}0\\1\\-1\end{bmatrix}$$

which we can now see is a linear combination of $\mathbf{v}_1, \mathbf{v}_2$! Thus $\mathbf{x} \in span\{\mathbf{v}_1, \mathbf{v}_2\}$ which now implies that

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

We now return to the main investigation concerning how two spanning sets can be related.

Proposition 2.2.8. *If* $\boldsymbol{u} \in \text{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ *then* $\text{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m, \boldsymbol{u}\} = \text{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$.

Proof. Recall what was mentioned about if-then statements and showing two sets are equal. Our goal will be to show that the two sets $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\}$ and $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ are equal. Our hypothesis is that $\mathbf{u} \in \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and we must use this somewhere along the way.

Let's first assume that $\mathbf{u} \in \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$. This means that there exist scalars $x_1, x_2, \ldots, x_m \in \mathbb{R}$ such that

$$\mathbf{u} = x_1\mathbf{u}_1 + \cdots + x_m\mathbf{u}_m$$

Since we want to ultimately show that $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\} = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ we pick an arbitrary $\mathbf{v} \in \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\}$ and show that it is also in $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$. This will show that $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\} \subseteq$

$\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$. For this $\mathbf{v}$ that we have chosen, there must exist some other scalars $y_0, y_1, \ldots, y_m \in \mathbb{R}$ such that

$$\mathbf{v} = y_0\mathbf{u} + y_1\mathbf{u}_1 + \cdots + y_m\mathbf{u}_m$$

Now, we use our asssumption that $\mathbf{u} \in \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and substitute $\mathbf{u} = x_1\mathbf{u}_1 + \cdots + x_m\mathbf{u}_m$ into the equation for $\mathbf{v}$. This tells us that

$$\mathbf{v} = y_0\mathbf{u}+y_1\mathbf{u}_1+\cdots+y_m\mathbf{u}_m = y_0(x_1\mathbf{u}_1+\cdots+x_m\mathbf{u}_m)+y_1\mathbf{u}_1+\cdots+y_m\mathbf{u}_m = (y_0x_1+y_1)\mathbf{u}_1+\cdots+(y_0x_m+y_m)\mathbf{u}_m$$

which is an element of $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$! This means that $\mathbf{v} \in \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and

$$\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\} \subseteq \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$$

This shows the first part. It remains to show that $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subseteq \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\}$ so we pick a vector $\mathbf{w} \in \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and conclude that it is also in $\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\}$. The assumption on our vector $\mathbf{w}$ implies that there exist scalars $z_1, z_2, \ldots, z_m$ such that

$$\mathbf{w} = z_1\mathbf{u}_1 + \cdots + z_m\mathbf{u}_m$$

Now, observe that 0 is a scalar that we can always use when constructing linear combinations, hence we can write $\mathbf{w}$ as a linear combination of $\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}$ by writing

$$\mathbf{w} = 0\mathbf{u} + z_1\mathbf{u}_1 + \cdots + z_m\mathbf{u}_m$$

hence $\mathbf{w} \in \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\}$ and we can conclude that

$$\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subseteq \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\}$$

This now means that

$$\text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{u}\} = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$$

which completes the proof.

$\square$

Before ending this section, we exhibit one more bit of compact (and very useful!) notation, namely that of representing a linear system via matrix notation.

Let A be a matrix with columns $\mathbf{a}_1 = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{a}_3 = \begin{bmatrix} -4 \\ 1 \\ -5 \end{bmatrix}$. We can write the matrix A as

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} = \begin{bmatrix} 3 & 2 & -4 \\ 0 & 1 & 1 \\ 1 & 0 & -5 \end{bmatrix}$$

Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, then we can define the following product.

Definition 2.2.9. The **product** $A\mathbf{x}$ is given by

$$A\mathbf{x} = \begin{bmatrix} 3 & 2 & -4 \\ 0 & 1 & 1 \\ 1 & 0 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -4 \\ 1 \\ -5 \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3$$

This allows us to succinctly write out linear systems in terms of matrices as follows.

Example 2.2.10. The linear system $\begin{cases} 3x_1 + 2x_2 - 4x_3 = 1 \\ \quad\quad\; x_2 + \; x_3 = 0 \\ x_1 \quad\quad\; - 5x_3 = 2 \end{cases}$

has augmented matrix

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

Using the product definition above we can see that for a vector $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ the above system can be written as $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

We now end the section with (what we will soon see is) a very useful theorem. We note that any time one says "the following are equivalent", it means that "if and only if" statements should be placed between every item in the list. That is to say, if one sentence in the list if true, all others are true, and likewise, if one sentence is false then all others are false.

Theorem 2.2.11. *Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m, \boldsymbol{b} \in \mathbb{R}^n$. The following statements are equivalent:*

1. $\boldsymbol{b} \in \operatorname{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$.
2. *The vector equation $x_1\boldsymbol{u}_1 + \cdots + x_m\boldsymbol{u}_m = \boldsymbol{b}$ has at least one solution.*
3. *The linear system with augmented matrix $\left[\begin{array}{cccc|c} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m & \boldsymbol{b} \end{array}\right]$ is consistent.*
4. *The equation $A\boldsymbol{x} = \boldsymbol{b}$ with $A = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \end{bmatrix}$ has at least one solution for every choice of $\boldsymbol{b} \in \operatorname{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$.*

2.3 Linear Independence

The topic of linear independence will be precisely what we need to understand when a set of vectors spans some euclidean space. In order to wrap our heads around it, we need one new definition.

Definition 2.3.1. A linear system if **homogeneous** if it has the form

$$x_1\mathbf{a}_1 + x_2 + \mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$

In other words, every linear equation is set equal to zero (the $\mathbf{0}$ denotes the zero vector, all of whose entries are 0).

The beauty of a homogeneous linear system is that it is **always** consistent since we can always find the solution $x_1 = x_2 = \cdots = x_n = 0$. We call this solution the **trivial** solution, any other solutions are referred to as **non-trivial** solutions. It is this notion that allows us to define linear independence.

Definition 2.3.2. Suppose $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m \in \mathbb{R}^n$. If the textbfonly solution to the linear system

$$x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \cdots + x_m\mathbf{u}_m = \mathbf{0}$$

is the trivial solution, then we say $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is a **linearly independent** set of vectors, or that the vectors are linearly independent. If the vector equation above has non-trivial solutions then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is a **linearly dependent** set of vectors. Recalling that a linear system either has $0, 1$, or infinitely many solutions, we can say that a set of vectors is linearly dependent if the associated homogeneous linear system involving those vectors has at least one free variable. If this confuses you then feel free to ignore it.

Example 2.3.3. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} \in \mathbb{R}^3$. Is $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ a linearly independent set of vectors?

Considering the homogeneous linear system

$$x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + x_3\mathbf{u}_3 = \mathbf{0}$$

we can row reduce the corresponding augmented matrix

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

Using back substitution we see that

$$x_3 = 0 \implies x_2 = 0 \implies x_1 = 0$$

hence the only solution is the trivial one. This means that

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

are linearly independent.

Example 2.3.4. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} \in \mathbb{R}^3$. Are $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ linearly independent?

Considering the augmented matrix for the homogeneous linear system associated to the three vectors above we have

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

This equivalent linear system has x_3 as a free variable, and from this we obtain the **non-trivial** solution

$$x_3 = t, x_2 = -t, x_2 = -4t$$

where $t \in \mathbb{R}$. The existence of a non-trivial solution implies that $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}$, and $\begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}$ are linearly dependent.

Now that we have a little bit of a feel for linear independence, let's dig into some important propositions that we may want to use in the future.

Proposition 2.3.5. *If $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m \in \mathbb{R}^n$ then $\{\boldsymbol{0}, \boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$ is always linearly dependent.*

Proof. Given $\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathbb{R}^n$ the equation $x_0\mathbf{0} + x_1\mathbf{u}_1 + \dots + x_m\mathbf{u}_m = \mathbf{0}$ always has the nontrivial solution $x_0 = 1, x_1 = 0, \dots, x_m = 0$. ☐

We can actually say much more about when certain vectors are linearly dependent.

Proposition 2.3.6. *If $\boldsymbol{u}_1, \dots, \boldsymbol{u}_m \in \mathbb{R}^n$ and $m > n$ then $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$ is linearly dependent.*

Proof. We begin by observing that the vector equation

$$x_1\mathbf{u}_1 + \cdots + x_m\mathbf{u}_m = \mathbf{0}$$

always has at least one solution (the trivial one). This means the if we set up the usual augmented matrix and row reduce to a matrix B in echelon form, i.e.

$$\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m & \big| & \mathbf{0}\end{bmatrix} \sim B$$

then the matrix B does not have any rows of the form $\begin{bmatrix}0 & 0 & \cdots & 0 & \big| & c\end{bmatrix}$, where $c \neq 0$. Now, observing that the number of components of each vector is n (this is what it means to say that $\mathbf{u}_i \in \mathbb{R}^n$ and that $m > n$, we can conclude that there are more vectors than there are components of each vector. This means that the corresponding augmented matrix has more columns than rows, hence there **must** be at least one free variable, hence infinitely many (non-trivial) solutions, which completes the proof. ☐

This Proposition will be a very important one moving forward so we will want to keep it in our toolbox. Next, we get after a bigger question. How are the ideas of span and linear independence related? The answer as we will soon see, is quite nice, especially when phrased in terms of pivots. Recall that a pivot position in a matrix is a coefficient that sits in front of what would be a leading variable, in the corresponding linear system. We now give three relationships between these two ideas, and prove the third statement in detail.

Proposition 2.3.7. *Let* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^n$ *and suppose* $A = \begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m\end{bmatrix} \sim B$ *where* B *is a matrix in echelon form.*

1. $\operatorname{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\} = \mathbb{R}^n$ *exactly when* B *has a pivot in every* ***row****.*

2. $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ *is linearly independent exactly when* B *has a pivot in every* ***column****.*

This proposition is a personal favorite of many. It essentially gives an algorithm for determining when a given set of vectors span $\mathbb{R}^n$ and/or are linearly indepdnent. The question of spanning is a question about pivots of rows and the question of independence is a question about pivots of columns. All one needs to do before checking rows and/or columns, is put the given vectors as the columns of a matrix and row reduce to echelon form.

The last relationship is the following theorem.

Theorem 2.3.8. *Let* $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ *be a set of vector in* $\mathbb{R}^n$*. This set is linearly dependent if and only if one of the vectors in the set is in the span of the others.*

Proof. As with any "if and only if" proof, we must show both directions of the statement, We begin by assuming that the given vectors are linearly dependent, then deduce that one of the vectors is in the span of the others. This is the forward direction of the proof and is indicated with "→". After proving this direction, we tackle the reverse direction, denoted by "←", where we assume that one of the vectors is in the span of the others, and conclude linear dependence.

→

Suppose $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is linearly dependent. Then the vector equation $x_1\mathbf{u}_1 + \cdots + x_m\mathbf{u}_m = \mathbf{0}$ has a non-trivial solution, which we call $(x_1, \ldots, x_m)$. Note that this solution being non-trivial means that **at least one** of the x_i is non-zero (so we can divide by it!). Without loss of generality, lets assume that $x_1 \neq 0$. Using the vector equation above, we can then solve for $\mathbf{u}_1$

$$x_1\mathbf{u}_1 + \cdots + x_m\mathbf{u}_m = \mathbf{0} \implies x_1\mathbf{u}_1 = -(x_2\mathbf{u}_2 + \cdots + \mathbf{x}_m\mathbf{u}_m) \implies \mathbf{u}_1 = \frac{-(x_2\mathbf{u}_2 + \cdots + \mathbf{x}_m\mathbf{u}_m)}{x_1}$$

hence $\mathbf{u}_1 \in \operatorname{Span}\{\mathbf{u}_2, \ldots, \mathbf{u}_m\}$.

$\leftarrow$

Now assume that one of the vectors (say $\mathbf{u}_1$) is a linear combination of the others. Then there exist scalars $c_2, \ldots, c_m$ such that

$$\mathbf{u}_1 = c_2\mathbf{u}_2 + \cdots + c_m\mathbf{u}_m \quad \text{hence} \quad \mathbf{u}_1 - c_2\mathbf{u}_2 - \cdots - c_m\mathbf{u}_m = \mathbf{0}$$

so we have a non-trivial solution to the equation $x_1\mathbf{u}_1 + \cdots + x_m\mathbf{u}_m = \mathbf{0}$, which is exactly what it means for the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ to be linearly dependent. $\square$

Example 2.3.9. One can show that the set $\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 10 \\ 9 \end{bmatrix}, \begin{bmatrix} -4 \\ 17 \end{bmatrix}\}$ is linearly dependent (this is a good exercise), hence the above theorem implies that one of them is a linear combination of the others. In fact, we have

$$\begin{bmatrix} -4 \\ 17 \end{bmatrix} = \frac{-206}{19}\begin{bmatrix} 1 \\ -1 \end{bmatrix} + \frac{13}{19}\begin{bmatrix} 10 \\ 9 \end{bmatrix}$$

Warning: This does not mean that **every** vector is a linear combination of the others. An easy example of this is $\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$. Can you see which vectors are not in the span of the others?

We end the chapter with what is arguably the most important theorem of linear algebra which we refer to it as the big theorem. It is given as a list of equivalent statements and we will add to the list throughout the course. The key thing to note about the big theorem is that its statements are only true if we have **n** vectors in $\mathbb{R}^n$. In most of the statements of propositions we have m vectors in $\mathbb{R}^n$ and we do not assume that m and n are the same. This is something you should always be aware of if you try to use the big theorem to solve a problem.

Theorem 2.3.10. *Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$. The following statements are equivalent:*

1. *S spans $\mathbb{R}^n$.*
2. *S is linearly independent.*
3. *The system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b} \in \mathbb{R}^n$.*

We end the chapter with an example question that would be impossible to solve without the big theorem.

Example 2.3.11. Let $A = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ for $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix}$, and show that $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^3$.

By the big theorem, $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^3$ if and only if the three vectors are linearly independent. This was shown in Example 2.3.3, hence the question is true by the big theorem.

Chapter 3

Linear Transformations

Up to this point, we have done a tremendous amount of algebra with vectors and matrices, but we have not examined the geometry underlying linear systems. As we will soon see, the notion of a linear transformation allows us to translate our algebraic notions into geometric ones. Often times in practice, we aim to answer hard geometric questions and the methods we use involve translating the geometry into an algebraic question involving matrices, then using the matrices to answer the question, and translating the answer back to the underlying geometric picture.

3.1 The Basics of Linear Maps

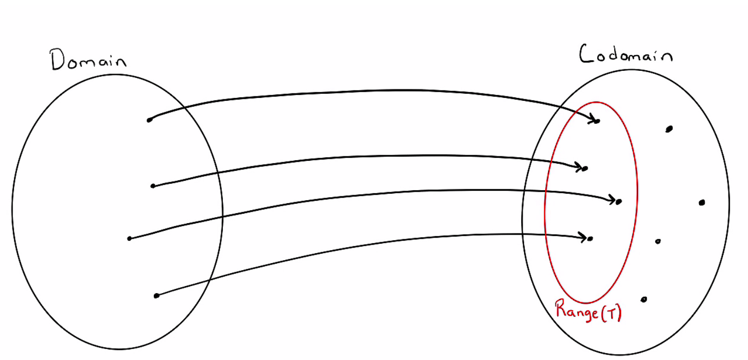
We begin by outlining the basic vocabulary of linear transformations, otherwise known as linear maps. A priori, a linear map is just a function that takes vectors as input and outputs vectors (of possibly different size than the input). The notation

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

reads as "T is a function from $\mathbb{R}^m$ to $\mathbb{R}^n$".

- The set $\mathbb{R}^m$ is the **domain** of T (and T must be defined for every element of $\mathbb{R}^m$).
- The set $\mathbb{R}^n$ is the **codomain** of T. It is the set where all the output vectors live.
- The subset of $\mathbb{R}^n$ consisting of all output vectors, that is, all vectors of the form $\mathbf{w} = T(\mathbf{x})$ for some $\mathbf{x} \in \mathbb{R}^m$ is known as the **Range of** T, denoted $\text{Range}(T)$. It is also often called the image of T.

The following picture can serve as a visual summary of these definitions



Before defining what a linear map is, let's look at an example of a vector valued function.

Example 3.1.1. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1x_2 \\ x_2 - x_3 \end{bmatrix}$$

The domain of this map is $\mathbb{R}^3$ and the codomain is $\mathbb{R}^2$. The vector $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is in Range(T) because $T\left(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Definition 3.1.2. A function $T\colon \mathbb{R}^m \to \mathbb{R}^n$ is a **linear transformation** or **linear map** if, for every $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ and every scalar $r \in \mathbb{R}$, we have:

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$.
- $T(r\mathbf{u}) = rT(\mathbf{u})$

Some people like to combine the two conditions of linearity by saying that T is a linear transformation if

$$T(r\mathbf{u} + s\mathbf{v}) = rT(\mathbf{u}) + sT(\mathbf{v})$$

for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ and all scalars $r, s \in \mathbb{R}$.

Example 3.1.3. Let's show that the map $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -x_2 \\ x_1 + x_2 \\ 4x_1 \end{bmatrix}$$

is a linear map.

Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ be arbitrary vectors in the domain ($\mathbb{R}^2$). Then $\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$ hence

$$T(\mathbf{u} + \mathbf{v}) = \begin{bmatrix} -(u_1 + v_1) \\ (u_1 + v_1) + (u_2 + v_2) \\ 4(u_1 + v_1) \end{bmatrix} = \begin{bmatrix} -u_2 \\ u_1 + u_2 \\ 4u_1 \end{bmatrix} + \begin{bmatrix} -v_2 \\ v_1 + v_2 \\ 4v_1 \end{bmatrix} = T(\mathbf{u}) + T(\mathbf{v})$$

Moreover, if $r \in \mathbb{R}$ then $r\mathbf{u} = \begin{bmatrix} ru_1 \\ ru_2 \end{bmatrix}$ and

$$T(r\mathbf{u}) = \begin{bmatrix} -ru_2 \\ ru_1 + ru_2 \\ 4ru_1 \end{bmatrix} = r\begin{bmatrix} -u_2 \\ u_1 + u_2 \\ 4u_1 \end{bmatrix} = rT(\mathbf{u})$$

hence T is indeed a linear transformation.

Example 3.1.4. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^2$ be the map defined earlier by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1x_2 \\ x_2 - x_3 \end{bmatrix}$$

This is **not** a linear map since, for example, if $r = 2$ then

$$T(2\mathbf{x}) = \begin{bmatrix} 4x_1x_2 \\ 2(x_2 - x_3) \end{bmatrix} \neq 2T(\mathbf{x}) = \begin{bmatrix} 2x_1x_2 \\ 2(x_2 - x_3) \end{bmatrix}$$

One way to see why this is not a linear map is that the first coordinate of an arbitrary output vector is a quadratic function in the input variables. In general, linear maps have coordinate functions that are linear.

One of the most amazing things about linear maps is that they are intimately tied to matrices.

Definition 3.1.5. A matrix with n rows and m columns has **dimensions** $n \times m$ and is referred to as an $n \times m$ matrix. An $n \times n$ matrix is often called a **square** matrix.

Now, by recalling Definition 2.2.9 (the product $A\mathbf{x}$) we can see the connection with matrices and linear maps. If A is an $n \times m$ matrix and $\mathbf{x} \in \mathbb{R}^m$ then the product $A\mathbf{x}$ is always a vector in $\mathbb{R}^m$ (you should verify this for yourself). In other words, an $n \times m$ matrix, when multiplied by a vector in $\mathbf{x} \in \mathbb{R}^m$, takes $\mathbf{x}$ to a vector $A\mathbf{x}$, in $\mathbb{R}^n$.

Theorem 3.1.6. *Let A be an $n \times m$ matrix and define $T\colon \mathbb{R}^m \times \mathbb{R}^n$ via*

$$T(\boldsymbol{x}) = A\boldsymbol{x}$$

then T is a linear transformation.

The above theorem is powerful and can be used to easily show that a given map is linear, without verifying the two properties of the original definition. That is, to show that a function $T\colon \mathbb{R}^m \to \mathbb{R}^n$ is a linear map, it suffices to find a matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

Example 3.1.7. Consider the linear map from Example 3.1.3, which we now know is indeed linear. Using Definition 2.2.9 we have

$$T(\mathbf{x}) = \begin{bmatrix} 0x_1 - x_2 \\ x_1 + x_2 \\ 4x_1 + 0x_2 \end{bmatrix} = x_1 \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

so $T(\mathbf{x}) = A\mathbf{x}$ for $A = \begin{bmatrix} 0 & -1 \\ 1 & 1 \\ 4 & 0 \end{bmatrix}$ and by the above theorem, T is a linear map.

Continuing with this example, let $\mathbf{w} = \begin{bmatrix} 10 \\ 5 \\ 2 \end{bmatrix}$. Is $\mathbf{w} \in \text{Range}(T)$? That is, does there exist a vector $\mathbf{x} \in \mathbb{R}^2$ such that $T(\mathbf{x}) = \mathbf{w}$. Since $T(\mathbf{x}) = A\mathbf{x}$, the existence of such a vector $\mathbf{x}$ would imply that $A\mathbf{x} = \mathbf{w}$ so to find the vector $\mathbf{x}$ we need to solve the system $A\mathbf{x} = \mathbf{w}$ which has augmented matrix

$$\left[\begin{array}{cc|c} 0 & -1 & 10 \\ 1 & 1 & 5 \\ 4 & 0 & 2 \end{array}\right] \sim \left[\begin{array}{cc|c} 1 & 1 & 5 \\ 0 & 1 & -10 \\ 0 & 0 & -58 \end{array}\right]$$

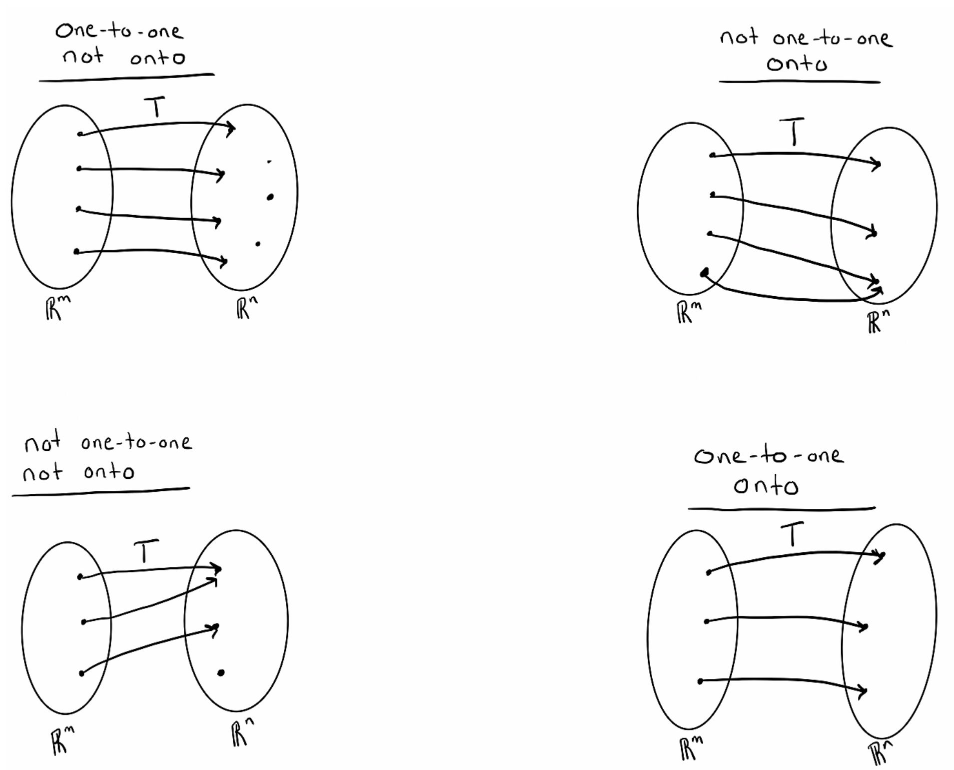
hence $\mathbf{w} \notin \text{Range}(T)$ because there does not exist a vector $\mathbf{x}$ with $T(\mathbf{x}) = \mathbf{w}$. This is an example of a linear map that is **not onto**, which leads us to our next set of definitions.

Definition 3.1.8. Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation.

1. T is **one-to-one** if for each $w \in \mathbb{R}^n$, there is <u>at most</u> one vector $\mathbf{x} \in \mathbb{R}^m$ such that $T(\mathbf{x}) = \mathbf{w}$. In other words, every domain vector $\mathbf{x}$ goes to exactly one vector in the codomain. It is not possible for one-to-one maps to send two different vectors to the same one. This would be "two-to-one".

2. T is **onto** if for every $\mathbf{w} \in \mathbb{R}^n$, there is <u>at least</u> one vector $\mathbf{x} \in \mathbb{R}^m$ such that $T(\mathbf{x}) = \mathbf{w}$. In other words, T is onto if every vector in the codomain is mapped to by some vector in the domain.

Less formally, T is one-to-one if nothing in the codomain gets "hit" more than once, and T is onto if everything in the codomain gets "hit".



All possibilities involving these definitions are most easily understood through these helpful pictures. Now let's get familiar with these concepts through examples.

Example 3.1.9. Let $A = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$ with $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{x}) = A\mathbf{x}$. We have

$$T\left(\begin{bmatrix} 1 \\ -2 \end{bmatrix}\right) = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 5 \\ -10 \end{bmatrix} \quad \text{and} \quad T\left(\begin{bmatrix} -3 \\ -4 \end{bmatrix}\right) = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} -3 \\ -4 \end{bmatrix} = \begin{bmatrix} 5 \\ -10 \end{bmatrix}$$

which means that T is not one-to-one. Moreover, (exercise) T is not onto since there is no $\mathbf{x} \in \mathbb{R}^2$ with $T(\mathbf{x}) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

Example 3.1.10. Let $A = \begin{bmatrix} 2 & 0 \\ 1 & -1 \\ 0 & 0 \end{bmatrix}$ with $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by $T(\mathbf{x}) = A\mathbf{x}$. Is T onto?

If it were, then for any vector $\mathbf{w} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3$ we could always find a vector $\mathbf{x} \in \mathbb{R}^2$ such that $T(\mathbf{x}) = A\mathbf{x} = \mathbf{w}$. Solving the associated linear system in the usual way we get that

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

which corresponds to a linear system whose third equation is $0 = c$. Now, if $\mathbf{w}$ was a vector with non-zero third coordinate, then $\mathbf{w} \notin \text{Range}(T)$ by what we have stated above. For example, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \notin \text{Range}(T)$ whereas $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \in \text{Range}(T)$.

Example 3.1.11. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ with $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(\mathbf{x}) = A\mathbf{x}$. In coordinates, we have $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{bmatrix} 2x_1 \\ 4x_2 \end{bmatrix}$. If $\mathbf{w} = \begin{bmatrix} a \\ b \end{bmatrix}$ then $T(\mathbf{x}) = \mathbf{w}$ has **exactly** one solution, namely $\mathbf{x} = \begin{bmatrix} a/2 \\ b/4 \end{bmatrix}$. There are no other vectors that get mapped to $\mathbf{w} = \begin{bmatrix} a \\ b \end{bmatrix}$. This means that every vector gets "hit" **and** there is exactly one $\mathbf{x}$ such that $T(\mathbf{x}) = \mathbf{w}$ for any $\mathbf{w}$, hence T is both one-to-one and onto.

There is an alternative definition of one-to-one that some may find useful.

Definition 3.1.12. T is one-to-one if

$$T(\mathbf{u}) = T(\mathbf{v}) \quad \text{implies} \quad \mathbf{u} = \mathbf{v}$$

Continuing along with the notion of a one-to-one map, we have one essential property of a linear map, that closely ties into being one-to-one.

If $T\colon \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$.

Theorem 3.1.13. *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear map. T is one-to-one if and only if the equation $T(\boldsymbol{x}) = \boldsymbol{0}$ has only the trivial solution $\boldsymbol{x} = \boldsymbol{0}$. That is, if $T(\boldsymbol{x}) = \boldsymbol{0}$, we must have $\boldsymbol{x} = \boldsymbol{0}$.*

Proof. If $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution, then by the alternative definition of one-to-one, we can conclude that if $T(\mathbf{u}) = T(\mathbf{v})$ then $T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{0}$. Moreover, since T is linear, we have $T(\mathbf{u}) - T(\mathbf{v}) = T(\mathbf{u} - \mathbf{v})$ hence $T(\mathbf{u} - \mathbf{v}) = \mathbf{0}$. THe trivial solution here implies that $\mathbf{u} = v = \mathbf{0}$, thus we must have $\mathbf{u} = \mathbf{v}$, implying that T is indeed one-to-one. □

In practice, when checking if a linear map is one-to-one, this theorem is the easiest method to use. In general, the following theorem outlines some other useful methods of checking when linear maps are one-to-one or onto.

Proposition 3.1.14. *Let A be an $n \times m$ matrix and define $T\colon \mathbb{R}^m \to \mathbb{R}^n$ via $T(\boldsymbol{x}) = A\boldsymbol{x}$. Then*

1. *T is one-to-one if and only if the columns of A are linearly independent.*
2. *If $m > n$, then T is never one-to-one.*
3. *T is onto if and only if the columns of A span $\mathbb{R}^n$ (the codomain).*
4. *If $m < n$, then T is never onto.*

In practice, you should always try to use statements 2 and 4 from the proposition, they are super useful!

Next, let's illustrate the last proposition with an example.

Example 3.1.15. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 3 & -3 \end{bmatrix}$ with $T\colon \mathbb{R}^2 \to \mathbb{R}^3$ given by $T(\mathbf{x}) = A\mathbf{x}$. Since $m < n$ we know immediately that T is not onto, but it could be one-to-one. To find out if it is, we look at the equation

$T(\mathbf{x}) = A\mathbf{x} = \mathbf{0}$. This is a linear system which reduces via

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

hence only the trivial solution exists and we must have $\mathbf{x} = \mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. We can then conclude, by the theorem we just proved, that T is one-to-one.

Now we have seen quite a few linear maps in action. Every one that we have seen was given by some matrix so it is natural to ask if **all** linear maps are given by matrices. The emphatic answer is yes!

Theorem 3.1.16. *If $T \colon \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation, then there exists an $n \times m$ matrix A such that $T(\boldsymbol{x}) = A\boldsymbol{x}$.*

This means that we can **always** find the matrix associated to a linear map (and we should!). Working with linear maps is always easier when working with their associated matrices and because of this, we move interchangeably between linear maps and matrices from here on out. When you think of a matrix you shoudl always be thinking about what it does as a linear transformation.

The beauty of this thoerem extends further. In fact, given any linear map, we can **always** find its assoctaed matrix fairly easily.

Example 3.1.17. Let $T \colon \mathbb{R}^3 \to \mathbb{R}^5$ be given by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) = \begin{bmatrix} x_3 - x_1 \\ x_2 + x_3 \\ 4x_1 + 3x_2 \\ x_1 - 5x_3 \\ 9x_2 \end{bmatrix}$$

Lets find the matrix for T.

Let

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

We call these the standard basis vectors of the domain. If the domain is $\mathbb{R}^m$ then there are m of these vectors. The general rule is that the formula for A is

$$A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & T(\mathbf{e}_3) \end{bmatrix}$$

so by using the coordinate definition of T we have that $T(\mathbf{e}_1) = \begin{bmatrix} -1 \\ 0 \\ 4 \\ 1 \\ 0 \end{bmatrix}, T(\mathbf{e}_2) = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 0 \\ 9 \end{bmatrix}, T(\mathbf{e}_3) = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -5 \\ 0 \end{bmatrix}$ hence

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

This formula is our new best friend is is **extremely useful**. Let's summarize how this works in general.

Proposition 3.1.18. *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Then $T(\boldsymbol{x}) = A\boldsymbol{x}$ with A an $n \times m$ matrix given by*

$$A = \begin{bmatrix} T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) & \cdots & T(\boldsymbol{e}_m) \end{bmatrix}$$

Last but not least, we use results from this section to add to the big theorem.

Theorem 3.1.19. *Let $S = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ with associated linear transformation given by $T\colon \mathbb{R}^n \to \mathbb{R}^n$. The following statements are equivalent:*

1. *S spans $\mathbb{R}^n$.*
2. *S is linearly independent.*
3. *The system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b} \in \mathbb{R}^n$.*
4. *T is onto.*
5. *T is one-to-one.*

An interesting consequence of this, in stark contrast to Proposition 3.1.14, is that a linear map from $\mathbb{R}^n$ to itself is either one-to-one **and** onto, or neither.

3.2 Matrix Algebra

In continuing with our geometric theme, the tools of matrix algebra provide the algebraic notions that we will use to answer geometric questions concerning multiple linear transformations.

The first notion we need is matrix addition. This is done component-wise, in a way that is similar to vectors.

Example 3.2.1. Let $A = \begin{bmatrix} 4 & 0 & -1 \\ 2 & 2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 9 & 10 & 6 \\ -1 & 0 & -1 \end{bmatrix}$. We define $A + B$ to be the 2×3 matrix

$$A + B = \begin{bmatrix} 13 & 10 & 5 \\ 1 & 2 & 4 \end{bmatrix}$$

Note that we obtained this matrix just by adding matching coordinates of each matrix. For a given scalar $r \in \mathbb{R}$ we define rA to be

$$rA = \begin{bmatrix} 4r & 0 & -r \\ 2r & 2r & 5r \end{bmatrix}$$

In general, there are just several things to note about matrix addition.

1. One can only add matrices of the same size. That is, if C is a 2×3 matrix and D is a 3×4 matrix then $C + D$ is undefined.
2. We denote the zero matrix with n rows and m columns by 0_{nm}, or simply write 0 when the context is clear. The zero matrix satisfies the property that $0 + A = A$ for any matrix A where the addition is defined.
3. Matrix addition is commutative, that is, $A + B = B + A$.

We now move onto the slightly more complicated (but also more important) notion of matrix multiplication. This can be thought of as a generalization of multiplying a matrix by a vector.

Definition 3.2.2. If A is an $n \times k$ matrix and B is a $k \times m$ matrix, written column-wise as $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_m \end{bmatrix}$ then the **product** AB is the $n \times m$ matrix given by

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_m \end{bmatrix}$$

where each column $A\mathbf{b}_i$ is computed using Definition 2.2.9.

Example 3.2.3. Let $A = \begin{bmatrix} 4 & 0 & -1 \\ 2 & 2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} -2 & 1 & 2 & 0 \\ 6 & 0 & -3 & -1 \\ 7 & -1 & 4 & 1 \end{bmatrix}$. Then

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

$$= A\begin{bmatrix} -2 \\ 6 \\ 7 \end{bmatrix} + A\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + A\begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} + A\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -15 & 5 & 4 & -1 \\ 43 & -3 & 18 & 3 \end{bmatrix}$$

This can be taken to be the original definition of matrix multiplication, but in practice, there is a much easier way of computing it.

Given an $n \times k$ matrix A and a $k \times m$ matrix B, the product AB is the $n \times m$ matrix, whose ij-entry (the entry in row i column j) is the dot product of the i^{th} row of A with the j^{th} column of B. It would be a great exercise to run back through the example above using this method. In doing so, you should also see why the product is not defined when the number of columns of A does not equal the number of rows of B.

Warning: In general, the order in which one multiplies matrices matters. With the example above, even though AB is defined, BA is not. Always exercise care with the order in which you multiply matrices.

It is now a good time to introduce some special types of matrices that we will encounter more frequently as well as some useful ideas that come from our new perspective of matrices. We begin with two definitions, then lay out some special classes of matrices.

Definition 3.2.4. The **transpose** of an $m \times n$ matrix A, denoted $A^\top$, is the $m \times n$ matrix obtained by interchanging the rows and columns of A. For example, if $A = \begin{bmatrix} 3 & 0 & 1 \\ 4 & 1 & -2 \end{bmatrix}$ then $A^\top = \begin{bmatrix} 3 & 4 \\ 0 & 1 \\ 1 & -1 \end{bmatrix}$. The main properties of transposing matrices is that the transpose of a product is the product of transposes (with order swapped), that is

$$(AB)^\top = B^\top A^\top$$

Definition 3.2.5. Given a square matrix A, we define the $k^{\mathbf{th}}$ power of A to be the matrix A^k. That is, the matrix obtained by multiplying A by itself k times. For example, $A^2 = AA$ and $A^3 = A(A^2) = AAA$.

- **The $n \times n$ identity Matrix, I_n**: Considering (again) the standard basis vectors for $\mathbb{R}^n$, $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$, the $n \times n$ identity matrix is given by

$$I_n = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}$$

I_n is the unique matrix for which $AI_n = I_nA = A$ for any $n \times n$ matrix A.

To see some small examples, we have

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- **Diagonal Matrices**: If the only non-zero entries of a square matrix A lie on thh main diagonal, then we call A a **diagonal matrix**. For example,

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

and we can check that

$$A^2 = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 36 \end{bmatrix} \quad \text{and} \quad A^3 = \begin{bmatrix} 8 & 0 & 0 & 0 \\ 0 & 64 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 216 \end{bmatrix}$$

In general, if

$$A = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix} \quad \text{then} \quad A^k = \begin{bmatrix} a_{11}^k & 0 & \dots & 0 \\ 0 & a_{22}^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn}^k \end{bmatrix}$$

- **Triangular Matrix**: If A is a square matrix with zeroes in each entry below the main diagonal, then A is an **upper triangular** matrix. We can similarly define a lower triangular matrix to have zeroes below the main diagonal. If A is either upper or lower triangular, then we say that A is a **triangular matrix**. For example, given the two matrices

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 4 & 4 & 4 \end{bmatrix}$$

we have that A is upper triangular and B is lower triangular.

Expanding on what we said for diagonal matrices, we have the (great) fact.

Proposition 3.2.6. *If A is a triangular matrix, then A^k is triangular.*

Note that what was said about powers of diagonal matrices follows from this Proposition because all diagonal matrices are triangular.

We now outline some things that we need to be very careful about, when it comes to matrices and products of them. These are things that you will want to always keep in mind when computing matrix products.

1. In general, if AB is defined, the product BA is not defined. This is always the case if A or B are not square matrices and can be seen in the previous example.
2. The commutative property does not hold for matrix multiplication. For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 8 \\ 15 & -10 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$$

3. Unlike with numbers, it is possible to multiply two non-zero matrices together and obtain the zero matrix. For example,
$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$
The takeaway from this is that if $AB = 0$ we cannot conclude that either $A = 0$ or $B = 0$.

4. Its possible that $AC = BC$ but $A \neq B$ and $C \neq 0$. For example,
$$\begin{bmatrix} 2 & 3 \\ 6 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 4 & 4 \end{bmatrix}$$
The takeaway from this is that if $AC = BC$ and $C \neq 0$ then we cannot conclude that $A = B$.

We now end the section with what is arguably the most important aspect of matrix multiplication.

Proposition 3.2.7. *Let $T_1 \colon \mathbb{R}^m \to \mathbb{R}^k$ be given by $T_1(\boldsymbol{x}) = A_1\boldsymbol{x}$ and let $T_2 \colon \mathbb{R}^k \to \mathbb{R}^n$ be given by $T_2(\boldsymbol{x}) = A_2\boldsymbol{x}$. The matrix associated to the composition $T_2 \circ T_1(\boldsymbol{x}) = T_2(T_1(\boldsymbol{x}))$ is A_2A_1, that is, matrix multiplication corresponds to composition of associated linear maps.*

Proof. We can quickly verify that
$$T_2 \circ T_1(\mathbf{x}) = T_2(T_1(\mathbf{x})) = T_2(A_1\mathbf{x}) = A_2A_1\mathbf{x}$$
□

Example 3.2.8. Let $T_1, T_2 \colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T_1(\mathbf{x}) = A_1\mathbf{x}$ and $T_2(\mathbf{x}) = A_2\mathbf{x}$ with
$$A_1 = \begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} 2 & 1 \\ 3 & -2 \end{bmatrix}$$
If $\mathbf{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ then we can compute $T_1(T_2(\mathbf{x}))$. First observe that
$$T_2(\mathbf{x}) = \begin{bmatrix} 2 & 1 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 7 \end{bmatrix}$$
Now we have
$$T_1(T_2(\mathbf{x})) = \begin{bmatrix} 1 & 2 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 7 \end{bmatrix} = \begin{bmatrix} 15 \\ 25 \end{bmatrix}$$
We can also verify that
$$A_1A_2 = \begin{bmatrix} 8 & -7 \\ 6 & -19 \end{bmatrix}$$
which means that $T_1(T_2(\mathbf{x})) = \begin{bmatrix} 8 & -7 \\ 6 & -19 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. A direct computation indeed yields the desired result.

As a last note, we emphasize that the order of matrix multiplication is an essential component of computing a composition of linear maps correctly. In practice, always make sure that the order in which you multiply is correct.

3.3 Inverses

We now come to the last topic of this chapter, the all important idea of an inverse. We will see that the notion of an inverse will correspond to the same notion of an inverse function. They will also make solving certain linear systems much easier.

Definition 3.3.1. If A is an $n \times n$ matrix and there exists another $n \times n$ matrix, A^{-1} (pronounced A inverse), satisfying

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

then A is **invertible** and we say A^{-1} is the inverse of A.

Example 3.3.2. Let $A = \begin{bmatrix} 1 & -1 \\ 3 & 2 \end{bmatrix}$. We can see that A is invertible and $A^{-1} = \begin{bmatrix} 2/5 & 1/5 \\ -3/5 & 1/5 \end{bmatrix}$ since an easy computation shows that $A^{-1}A = AA^{-1} = I_2$.

Going along the lines of the example, we actually have a nice closed formula for the inverse of a 2×2 matrix (larger matrices do not have such nice formulas). Given $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, if $ad - bc \neq 0$ then A is invertible and

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

Our main task in this section will be to compute inverses for $n \times n$ matrices where $n > 2$. The process is as follows:

Suppose we are given the $n \times n$ matrix $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$. To find A^{-1} (if it exists) we

1. Augment A with the $n \times n$ identity matrix $I_n = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}$ to get

$$\left[\begin{array}{cccc|cccc} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{array}\right]$$

2. Reduce the left hand side (the matrix A) to reduced echelon form and apply the same row operations to I_n.

3. If this algorithm can be completed, the right hand side of the augmented matrix will be A^{-1}. That is

$$\left[A \mid I_n\right] \sim \left[I_n \mid A^{-1}\right]$$

Example 3.3.3. Find A^{-1} if $A = \begin{bmatrix} -1 & 4 & 1 \\ 1 & 0 & 1 \\ 2 & 0 & 1 \end{bmatrix}$. By applying all the necessary row operations we get

$$\left[\begin{array}{ccc|ccc} -1 & 4 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 2 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & -4 & -1 & -1 & 0 & 0 \\ 0 & 4 & 2 & 1 & 1 & 0 \\ 0 & 8 & 3 & 2 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & -4 & -1 & -1 & 0 & 0 \\ 0 & 1 & 1/2 & 1/4 & 1/4 & 0 \\ 0 & 8 & 3 & 2 & 0 & 1 \end{array}\right]$$

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

hence $A^{-1} = \begin{bmatrix} 0 & -1 & 1 \\ 1/4 & -3/4 & 1/2 \\ 0 & 2 & -1 \end{bmatrix}$.

Definition 3.3.4. An $n \times n$ matrix A is **non-singular** if it has an inverse, otherwise we say it is **singular**. It is also important to note that if A^{-1} exists, it is unique.

Inverses also relate nicely to linear transformations.

Definition 3.3.5. If $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation that is one-to-one and onto then T is **invertible**. Its inverse is the function $T^{-1}\colon \mathbb{R}^n \to \mathbb{R}^n$ with the property that for each $\mathbf{x} \in \mathbb{R}^n$ we have $T^{-1}(T(\mathbf{x})) = \mathbf{x}$. In fact, if T is given by $T(\mathbf{x}) = A\mathbf{x}$, then if T is invertible, we always have $T^{-1}(\mathbf{x}) = A^{-1}\mathbf{x}$.

Example 3.3.6. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} 4x_1 + 3x_2 \\ -6x_1 + 5x_2 \end{bmatrix}$ so that $T(\mathbf{x}) = A\mathbf{x}$ with $A = \begin{bmatrix} 4 & 3 \\ -6 & 5 \end{bmatrix}$. Using the formula for the inverse of a 2×2 matrix, we have that

$$A^{-1} = \frac{1}{38}\begin{bmatrix} 5 & -3 \\ 6 & 4 \end{bmatrix}$$

We can then verify that

$$T^{-1}\Big(T\Big(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big)\Big) = T^{-1}\Big(\begin{bmatrix} 4 & 3 \\ -6 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \Big(\frac{1}{38}\begin{bmatrix} 5 & -3 \\ 6 & 4 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ -6 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\Big) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Now that we have a bit of a handle of inverses of matrices, we can return to the algebraic mishaps of last section, and see that invertibility was indeed the solution we needed to make sense of when matrix multiplication behaves like regular multiplication of numbers.

Proposition 3.3.7. *Suppose A and B are non-singular $n \times n$ matrices and C and D are $n \times m$ matrices. Then*

1. *A^{-1} is invertible with inverse $(A^{-1})^{-1} = A$.*
2. *AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$ This is known as the shoes and socks lemma. If B represents putting on your socks and A represents putting on your shoes, then undoing this process translates to **first** taking off your shoes (A^{-1}), then taking off your socks (B^{-1}).*
3. *If $AC = AD$, then $C = D$. We can obtain this logically by taking the first equation and multiplying **on the left** by A^{-1} on both sides.*
4. *If $AC = O_{nm}$ then $C = 0_{nm}$. This can similarly be obtained by multiplying both sides by A^{-1} on the left.*

It is essential to note here that invertibility of A is precisely what gives us the ability to draw all the conclusions we have made. Without invertibility of A, we cannot deduce any of the four statements.

We can now add some more results to the big theorem (which some refer to as the invertible matrix theorem).

Theorem 3.3.8. *Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ with associated linear transformation given by $T\colon \mathbb{R}^n \to \mathbb{R}^n$. The following statements are equivalent:*

1. *S spans $\mathbb{R}^n$.*
2. *S is linearly independent.*
3. *The system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b} \in \mathbb{R}^n$.*
4. *T is onto.*
5. *T is one-to-one.*
6. *A is invertible.*

We now end the section with one illustration of why we love invertible matrices.

Example 3.3.9. Consider the linear system

$$\begin{cases} 4x_1 + 3x_2 = 5 \\ -2x_1 - x_2 = 7 \end{cases}$$

This system is the same as $A\mathbf{x} = \mathbf{b}$ for $A = \begin{bmatrix} 4 & 3 \\ -2 & -1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$. Moreover, A is invertible with inverse given by $A^{-1} = \begin{bmatrix} -1/2 & -3/2 \\ 1 & 2 \end{bmatrix}$. Note that we found this matrix by using the formua for 2×2 matrices. Looking at the matrix equation $A\mathbf{x} = \mathbf{b}$, we can see that isolating $\mathbf{x}$ is equivalent to multiplying both sides by A^{-1} **on the left**, hence

$$A^{-1}A\mathbf{x} = A^{-1}\mathbf{b} \implies \mathbf{x} = A^{-1}\mathbf{b}$$

and

$$A^{-1}\mathbf{b} = \begin{bmatrix} -1/2 & -3/2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} -13 \\ 19 \end{bmatrix} = \mathbf{x}$$

which uniquely solves the system.

Chapter 4

Basis and Subspaces

We now enter the second half of the topics list for this course, the first of which is subspaces. The language of subspaces gives us precise notions that allow one to describe things like planes and lines in $\mathbb{R}^3$ in greater generality. Once we have the basics of subspaces, we will define the all important notion of a basis, which will also lead us to the definition of dimension. We then take an in depth look at some of the most important subspaces related to a matrix, namely the column space and null space. We then finish the chapter with a description of change of basis, a central theme in all of linear algebra.

4.1 Subspaces

Definition 4.1.1. A subset S of $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ is vectors in S satisfy the three following conditions:

1. $\mathbf{0} \in S$.
2. If $\mathbf{u}, \mathbf{v} \in S$, then $\mathbf{u} + \mathbf{v} \in S$. This is known as closure under addition.
3. If $r \in \mathbb{R}$ and $\mathbf{u} \in S$ then $r\mathbf{u} \in S$. This is known as closure under scaling.

It is worth noting that the first condition introduces the necessary condition that no two parallel subspaces can ever exist.

Example 4.1.2. Let $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^n$ and $S = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Is S a subspace of $\mathbb{R}^n$?

We must verify each of the three conditions for S to be a subspace.

1. $\mathbf{0} \in S$ because $\mathbf{0} = 0\mathbf{u}_1 + 0\mathbf{u}_2$.
2. Elements of S are closed under addition. Let $\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$ and $\mathbf{v} = d_1\mathbf{u}_2 + d_2\mathbf{u}_2$. Clearly both of these are arbitrary elements of S. By adding them together we see that

$$\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{u}_1 + (c_2 + d_2)\mathbf{u}_2 \in S$$

   which shows that the second condition is met.
3. If $r \in \mathbb{R}$ and $\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \in S$, then

$$r\mathbf{u} = rc_1\mathbf{u}_1 + rc_2\mathbf{u}_2 \in S$$

   which shows that the third condition is met. We can now conclude that S is a subspace.

In general, the span of any set of vectors in $\mathbb{R}^n$ is always a subspace of $\mathbb{R}^n$. This can easily be seen by reworking the above example with arbitrarily many vectors. This is such a fundamental fact that we state it as a theorem, which may be freely used from here on out.

Theorem 4.1.3. *If $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m \in \mathbb{R}^n$, then* $\text{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ *is a subspace of* $\mathbb{R}^n$*.*

Example 4.1.4. Let S be the set of solutions of the linear system

$$\begin{cases} 5x_2 + 5x_2 = 10 \\ \;\; x_2 + \;\; x_2 = 5 \end{cases}$$

Is S a subspace of $\mathbb{R}^2$?

NO! The easiest way to see this is by verifying that the first subspace condition is broken, That is, $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \notin S$ because it is not a solution to the *non-homogeneous* set of equations which define S.

Example 4.1.5. Let S be a subset of vectors in $\mathbb{R}^3$ consisting of the vectors $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ such that $ab = 0$. It turns out that S is not a subspace of $\mathbb{R}^3$ because S is not closed under addition. For example, the vectors $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ are both in S (condition that $ab = 0$ is satisfied) but

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

and $a = b = 1$ so $ab \neq 0$, which means that $\mathbf{u}, \mathbf{v} \in S$ but $\mathbf{u} + \mathbf{v} \notin S$.

We now introduce one of the fundamental subspaces associated to a matrix.

Theorem 4.1.6. *Let A be an $n \times m$ matrix. If S is the set of solutions of the homogeneous linear system $A\boldsymbol{x} = \boldsymbol{0}$, then S is a subspace of $\mathbb{R}^m$.*

Proof. First, we can see that $A\mathbf{0} = \mathbf{0}$ for any matrix A, hence $\mathbf{0} \in S$. Moreover, if $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$ (meaning $\mathbf{u}, \mathbf{v} \in S$), then $\mathbf{u} + \mathbf{v} \in S$ because

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

Finally, we can also see that $r\mathbf{u} \in S$ for any $\mathbf{u} \in S$. By assuming that $A\mathbf{u} = \mathbf{0}$ (since $\mathbf{u} \in S$) we have that for any scalar $r \in \mathbb{R}$

$$A(r\mathbf{u}) = rA\mathbf{u} = r(\mathbf{0}) = \mathbf{0}$$

This shows that S is a subspace. □

Definition 4.1.7. If A is an $n \times m$ matrix, then the set of all solutions to the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ is called **the null space of** A. It is denoted $\text{Null}(A)$ and is a subspace of $\mathbb{R}^m$. In other words

$$\text{Null}(A) = \Big\{\mathbf{x} \in \mathbb{R}^m \colon A\mathbf{x} = \mathbf{0}\Big\}$$

Example 4.1.8. Find the null space of $A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 4 & 3 \end{bmatrix}$.

The procedure for finding the null space of a matrix is always the same. We begin by augmenting with the zero vector and row reducing.

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

Looking at the echelon matrix, we can see that x_3 is a free variable so we set $x_3 = t$. Using back substitution from here get the general solution

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}t \\ -\frac{1}{2}t \\ t \end{bmatrix} = t \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

which means that

$$\text{Null}(A) = \text{Span}\left\{ \begin{bmatrix} -1/2 \\ -1/2 \\ 1 \end{bmatrix} \right\}$$

In general, one may encounter a situation where they have to determine if a given set is a subspace. Here are some helpful tips to carry out this task successfully:

1. Check if $\mathbf{0} \in S$. If not, then S is not a subspace.
2. If you can find **specific** vectors whose span is preciesly equal to S, then you can leverage Theorem 4.1.3 to argue that S is a subspace.
3. Recognize that S can in fact be expressed as the null space of some matrix and leverage Theorem 4.1.6 to show that S is a subspace (this method is powerful if you can get good at using it). As an example, consider the set of vectors of the form $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ such that $a - b = -c$. The condition that $a - b = -c$ is equivalent to $a - b + c = 0$. This set of vectors is then the solution set of the linear system $x_1 - x_2 + x_3 = 0$ which can be expressed as $\text{Null}(A)$ where $A = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}$. Note that in general, if you can algebraically manipulate something to obtain a zero somewhere, you are probably looking at a null space in disguise.
4. If all else fails, show closure under addition and scaling directly. If you encounter a road block in trying to prove this, it may mean that S is not a subspace. If you suspect this is the case, you should then seek out a counterexample. Either two vectors in S whose sum is not in S, or a fixed vector and fixed scalar which break closure under scaling.

We end this section by investigating how this relates to linear maps.

Definition 4.1.9. Let $T \colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. The set of all vectors $x \in \mathbb{R}^m$ such that $T(\mathbf{x}) = \mathbf{0}$ is called **the kernel of** T and is denoted $\ker(T)$.

Theorem 4.1.10. *If* $T \colon \mathbb{R}^m \to \mathbb{R}^n$ *is a linear transformation, then* $\ker(T)$ *is a subspace of* $\mathbb{R}^m$ *and* $\text{Range}(T)$ *is a subspace of* $\mathbb{R}^n$ *(recall that* $\text{Range}(T) = \{\boldsymbol{y} \in \mathbb{R}^n \colon T(\boldsymbol{x}) = \boldsymbol{y} \text{ for some } \in \mathbb{R}^m\}$*).*

Proof. The proof of this is very instructive and will be useful for the remainder of the course.

Since T is a linear transformation, we know that there exists a matrix A such that $T(\mathbf{x}) = A\mathbf{x}$. This means that if $T(\mathbf{x}) = \mathbf{0}$ then $A\mathbf{x} = \mathbf{0}$ hence $\ker(T)$ and $\text{Null}(A)$ are the same! By Theorem 4.1.6 we can conclude that $\ker(T)$ is a subspace. Similarly by recalling the formula of Definition 2.2.9, we can see that

$$\text{Range}(T) = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$$

where $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ (Remember this fact!). It then follows from Theorem 4.1.3 that $\text{Range}(T)$ is a subspace. □

Tracing back to the results of the previous chapter, we now have a nice new fact.

Proposition 4.1.11. *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. T is one-to-one if and only if* $\ker(T) = \{\boldsymbol{0}\}$.

We now end the section by adding to the big theorem.

Theorem 4.1.12. *Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix}\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n\end{bmatrix}$ with associated linear transformation given by $T\colon \mathbb{R}^n \to \mathbb{R}^n$. The following statements are equivalent:*

1. *S spans $\mathbb{R}^n$.*
2. *S is linearly independent.*
3. *The system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b} \in \mathbb{R}^n$.*
4. *T is onto.*
5. *T is one-to-one.*
6. *A is invertible.*
7. $\ker(T) = \{\boldsymbol{0}\}$.

4.2 Basis and Dimension

We saw in the previous section that spans of any number of vectors always forms a subspace. From this fact, we can ask the question, is every subspace the span of some set of vectors? The answer to this is yes! Moreover, we can go one step further and ask wether or not we can find the smallest set of vectors that span a given subspace. It is the notion of a basis that stems from this idea.

Definition 4.2.1. Let S be a subspace of $\mathbb{R}^n$. A **basis** for S is a set of vectors $\mathcal{B}_S = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ that spans S **and** is linearly independent.

Example 4.2.2. Let $S = \text{Span}\left\{ \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-3\\-6\end{bmatrix}, \begin{bmatrix}10\\20\end{bmatrix} \right\}$. We can observe that $\begin{bmatrix}-3\\-6\end{bmatrix} = -3\begin{bmatrix}1\\2\end{bmatrix}$ and $\begin{bmatrix}10\\20\end{bmatrix} = 10\begin{bmatrix}1\\2\end{bmatrix}$, hence, we can see that the spanning vectors for S are linearly dependent. Based on the definition for a basis, this means that the given vectors are not a basis. Moreover, along the lines of Proposition 2.2.8, we have that

$$\text{Span}\left\{ \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-3\\-6\end{bmatrix}, \begin{bmatrix}10\\20\end{bmatrix} \right\} = \text{Span}\left\{ \begin{bmatrix}1\\2\end{bmatrix} \right\}$$

Since $\left\{ \begin{bmatrix}1\\2\end{bmatrix} \right\}$ spans S and is linearly independent, we have that $\mathcal{B}_S = \left\{ \begin{bmatrix}1\\2\end{bmatrix} \right\}$ forms a basis for S.

Example 4.2.3. Let

$$\mathbf{u}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix}0\\1\\1\end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix}1\\1\\1\end{bmatrix}, \text{ and } \mathbf{u}_4 = \begin{bmatrix}1\\2\\2\end{bmatrix}$$

and let $S = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$. We can observe that $\mathbf{u}_3 = \mathbf{u}_1 + \mathbf{u}_2$ and $\mathbf{u}_4 = \mathbf{u}_1 + 2\mathbf{u}_2$ hence $S = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly independent, we can conclude that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is a basis for S.

Example 4.2.4. Consider the zero vector $\mathbf{0} \in \mathbb{R}^n$. The **zero subspace** $S = \{\mathbf{0}\}$ is the only subspace of $\mathbb{R}^n$ that has no basis. It consists of the origin and nothing else.

A task that will arise again and again is that of finding a basis for a given subspace. There are two ways of doing this and we break down each "recipe". Both have their advantages depending on the context in which you want to find a basis. In both cases, assume $S = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$.

1. **Recipe 1**:

   - Form a matrix A whose **<u>ROWS</u>** are the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$.
   - Use row reductions to transform A to an echelon matrix B.
   - The **non-zero** rows of B form a basis for S.

   Example 4.2.5. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \\ -2 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} 3 \\ -3 \\ 5 \\ 4 \end{bmatrix}$ and suppose $S = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

   Then

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

   The non-zero rows of B form a basis for S hence $\mathcal{B}_S = \left\{ \begin{bmatrix} -1 \\ -2 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -4 \\ 2 \end{bmatrix} \right\}$ is a basis for S.

2. **Recipe 2**:

   - Form a matrix A whose **<u>COLUMNS</u>** are the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$.
   - Use row reductions to transform A to an echelon matrix B.
   - The columns of A that correspond to the pivot columns of B form a basis for S.

   Example 4.2.6. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \\ -2 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} 3 \\ -3 \\ 5 \\ 4 \end{bmatrix}$ and suppose $S = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

   Then

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

   The pivot columns of B are columns 1 and 2 hence our basis for S is

   $$\mathcal{B}_S = \left\{ \begin{bmatrix} 1 \\ -2 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} \right\}$$

As a general rule, one should always remember that both recipes always work but recipe 1 tends to give "simpler" basis vectors (with more zeroes) whereas recipe 2 always gives basis vectors that are a subset of the vectors you started with. It is very common to want to reduce a spanning set to a basis (known as **reducing to a basis**), and this makes recipe 2 especially useful in many scenarios.
We can now use the notion of a basis to define dimension. The first fundamental fact that we need is the following.

Proposition 4.2.7. *If S is a subspace of $\mathbb{R}^n$, then every basis of S has the same number of vectors in it.*

Definition 4.2.8. The **dimension** of a subspace S, denoted $\dim(S)$, is the number of vectors in any basis for S. Note that in the previous example, we had $\dim(S) = 2$. In general, we always have $\dim(\{\mathbf{0}\}) = 0$.

Example 4.2.9. If S is a subspace of $\mathbb{R}^3$, what are the possible values of $\dim(S)$?

- S could be the zero subspace, in which case $\dim(S) = 0$.
- S could be a line through the origin in which case it has the form $S = \text{Span}\{\mathbf{u}_1\}$ for $\mathbf{u}_1 \neq \mathbf{0}$ and $\dim(S) = 1$.
- S could be a plane through the origin in which case it has the form $S = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ for $\mathbf{u}_1, \mathbf{u}_2$ linearly independent. In this case we have $\dim(S) = 2$.
- S could be all of $\mathbb{R}^3$, which we could write as

$$\mathbb{R}^3 = \text{Span}\left\{ \begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix} \right\}$$

  We call $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ the **standard basis of** $\mathbb{R}^3$. In this case, $\dim(S) = 3$ and in general, this is the only 3-dimensional subspace of $\mathbb{R}^3$.

This completes our list because any subspace $S = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ where $m > 3$ can never be m dimensional. This follows from the fact that any set of $m > 3$ vectors in $\mathbb{R}^3$ is never linearly independent, hence we can never have a basis containing more than 3 vectors.

Let's illustrate all of these ideas on some more complex examples.

Example 4.2.10. Find a basis for $\mathbb{R}^4$ containing the vectors

$$\mathbf{u}_1 = \begin{bmatrix}1\\0\\2\\0\end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix}2\\2\\1\\1\end{bmatrix}$$

We know that $\mathcal{B}_{\mathbb{R}^4} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ is a basis for $\mathbb{R}^4$ and since $\mathbf{u}_1, \mathbf{u}_2 \in \text{Span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$ we know (by Proposition 2.2.8) that

$$\mathbb{R}^4 = \text{Span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\} = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4\}$$

We can then apply recipe 2, placing $\mathbf{u}_1$ and $\mathbf{u}_2$ as the left-most vectors. Upon row reducing we get that

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

The pivots columns of B are columns $1, 2, 3$, and 4 hence

$$\mathcal{B}_{\mathbb{R}^4} = \left\{ \begin{bmatrix}1\\0\\2\\0\end{bmatrix}, \begin{bmatrix}2\\2\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix} \right\}$$

is a basis for $\mathbb{R}^4$ containing the prescribed vectors.

Next, let's up the difficulty a little bit and find a basis for a new but increasingly familiar subspace.

Example 4.2.11. Let $A = \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -2 & 3 \\ 2 & -1 & 0 & 3 \end{bmatrix}$ and compute $\dim(\mathrm{Null}(A))$.

We first need to find $\mathrm{Null}(A)$ which involves solving the linear system $A\mathbf{x} = \mathbf{0}$. We see that

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

We have 2 free variables so we set $x_3 = t$ and $x_4 = s$. Then, by back substitution, we get $x_2 = 2t - 3s$ and $x_1 = t - 2s$ hence

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = t \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} -3 \\ -3 \\ 0 \\ 1 \end{bmatrix}$$

It then follows immediately that

$$\mathcal{B}_{\mathrm{Null}(A)} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for $\mathrm{Null}(A)$ and we can conclude that $\dim(\mathrm{Null}(A)) = 2$. This procedure for finding a basis **always** works because the free variables will always contribute a 1 to one entry of a basis vector and a 0 to the corresponding entries of all other vectors. The offset 0's and 1's always ensure linear independence of the spanning vectors that we find, hence a basis is obtained automatically.

This number is so important that it has its own name.

Definition 4.2.12. The **nullity** of a matrix A, denoted $\mathrm{nullity}(A)$, is the number $\dim(\mathrm{Null}(A))$.

We will have much more to say about this numerical invariant in the next section, but before ending this section, we add to the big theorem once more.

Theorem 4.2.13. *Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ with associated linear transformation given by $T\colon \mathbb{R}^n \to \mathbb{R}^n$. The following statements are equivalent:*

1. *S spans $\mathbb{R}^n$.*
2. *S is linearly independent.*
3. *The system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b} \in \mathbb{R}^n$.*
4. *T is onto.*
5. *T is one-to-one.*
6. *A is invertible.*
7. $\ker(T) = \{\boldsymbol{0}\}$.
8. *S is a basis for $\mathbb{R}^n$.*

4.3 Row Space, Column Space, and Rank

We now introduce several more fundamental subspaces associated to a matrix. Once we have these additional definitions, we state the all important Rank-Nullity theorem, sometimes known as the fundemantal theorem of linear algebra. This theorem allows us to "decompose" $\mathbb{R}^n$ into disjoint subspaces.

Definition 4.3.1. Let A be an $n \times m$ matrix.

- The **row space** of A is the subspace of $\mathbb{R}^m$ spanned by the row vectors of A. It is denoted $\text{row}(A)$.
- The **column space** of A is the subspace of $\mathbb{R}^n$ spanned by the columns of A. It is denoted $\text{col}(A)$ and is the set of all outputs of the form $A\mathbf{x}$ or alternatively, just the span of the columns of A.

Combining these definition with our "recipes" from the last section we can deduce that given any matrix $A \sim B$ with B in echelon form

- The non-zero rows of B form a basis for $\text{row}(A)$.
- The columns of A corresponding to the pivot columns of B form a basis for $\text{col}(A)$.

Example 4.3.2.

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

Using the recipes, we can see that

$$\mathcal{B}_{\text{row}(A)} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\} \text{ and } \mathcal{B}_{\text{col}(A)} = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} \right\}$$

You may notice in this example that the row space and column space have the same dimension, even though one of them is a subspace of $\mathbb{R}^4$ and the other one is a subspace of $\mathbb{R}^3$. It turns out this phenomenon is always true.

Theorem 4.3.3. *Given any matrix A we have*

$$\dim(\text{Row}(A)) = \dim(\text{Col}(A))$$

Proof. Let B be a matrix in echelon form that is row equivalent to A. Every non-zero row of B contains a pivot and similarly, the pivot in each pivot column must lie in one of these non-zero rows. This means that the number of non-zero rows of B must equal the number of pivot columns of B. The number of pivot rows (resp. columns) is precisely what we use to find bases of these subspaces, hence these numbers always being equal impllies that $\text{row}(A)$ and $\text{col}(A)$ must always have the same dimension. □

This new numerical invariant also has its own name.

Definition 4.3.4. The **rank** of a matrix A, denoted $\text{rank}(A)$, is the dimension of the row, or column, space of A. In the above example we have $\text{rank}(A) = 3$ and we say that the matrix A has rank 3.

We now have everything we need to state what is, without question, the most amazing and useful theorem in this course, known most commonly as the rank-nullity theorem.

Theorem 4.3.5. *If A is an $n \times m$ matrix then*

$$\text{rank}(A) + \text{Nullity}(A) = m$$

Proof. If A is an $n \times m$ matrix, and $A \sim B$, then the number of non-zero rows of B is the rank of A by definition. This is also equal to the number of pivot columns of B. Each non-pivot column will correspond to a free variable and each free variable corresponds to a basis vector for $\text{Null}(A)$ (think about how we did this in Example 4.2.11). Putting this all together we have that
$\text{rank}(A) =$ the number of pivot columns of B
and
$\text{Nullity}(A) =$ the number of non-pivot columns of B.
The total number of columns of B, which is equal to m, is then the sum of the $\text{rank}(A)$ and $\text{Nullity}(A)$. □

The power of this theorem pops up again and again but at this stage, we can already find it useful in doing routine computations. In particular, if you want to find the rank or nullity of a matrix, you only need to find one and you get the other for free. This allows you to apply the "find a basis for the null space" procedure of Example 4.2.11 to find the nullity, or, apply your favorite recipe to find the rank, then the other numerical invariant follows immediately from rank-nullity.

Example 4.3.6. Consider the following matrix, and an equivalent echelon form

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

We saw in the previous example that $\text{rank}(A) = 3$ so we immediately know that this matrix has nullity equal to 1. You should verify this for yourself and in doing so will see, that $\left\{ \begin{bmatrix} 5/3 \\ 1/3 \\ -1/3 \\ 1 \end{bmatrix} \right\}$ is a basis for $\text{Null}(A)$.

We can also relate this theorem to linear transformations. Recall that given a linear transformation $T\colon \mathbb{R}^m \to \mathbb{R}^n$ with associated matrix A, we deduced that the span of the columns of A was equal to the range. If this does not ring a bell, take a look at Definition 2.2.9. With our new terminology, this means that $\text{Col}(A) = \text{Range}(T)$. Moreover, the solution set of $A\mathbf{x} = \mathbf{0}$ consisted of the vectors $\mathbf{x}$ such that $T(\mathbf{x}) = \mathbf{0}$. In other words, we had $\ker(T) = \text{Null}(A)$. This means that $\text{rank}(A) = \dim(\text{Range}(T))$ and $\text{Nullity}(A) = \dim(\ker(T))$. This is a dense paragraph but is worth spending the time to understand every sentence.

Combining these geometric notions with the rank-nullity theorem we can see that

$$m = \dim(\text{Range}(T) + \dim(\ker(T))$$

It is worth noting that the dimension of the row space being equal to $\text{rank}(A)$ and the dimension of the null space being equal to $\text{Nullity}(A)$ says something significant about $\mathbb{R}^m$. Both the row space and null space of A are subspaces of $\mathbb{R}^m$, whose dimensions add up to m. It is rank-nullity that allows us to conclude that $\mathbb{R}^m$ "decomposes" into the row space and the null space of the given matrix. This would not be possible to understand without our notions of linear maps and the rank-nullity theorem.

We now finish the section with one more example, followed by one more addition to the big theorem.

Example 4.3.7. Let $T\colon \mathbb{R}^{11} \to \mathbb{R}^9$ be given by $T(\mathbf{x}) = A\mathbf{x}$ and further assume that T is onto. How many dimensions of $\mathbb{R}^{11}$ are occupied by $\ker(T)$?

Since T is onto, we know that its range is the entire codomain, that is, $\text{Range}(T) = \mathbb{R}^9$. This means that $\dim(\text{Range}(T)) = \dim(\text{Col}(A)) = \text{rank}(A) = 9$. Rank-nullity then implies that

$$11 = 9 + \dim(\ker(T))$$

hence $\dim(\ker(T)) = 2$.

Theorem 4.3.8. *Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ with associated linear transformation given by $T\colon \mathbb{R}^n \to \mathbb{R}^n$. The following statements are equivalent:*

1. *S spans $\mathbb{R}^n$.*
2. *S is linearly independent.*
3. *The system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b} \in \mathbb{R}^n$.*
4. *T is onto.*
5. *T is one-to-one.*
6. *A is invertible.*
7. $\ker(T) = \{\boldsymbol{0}\}$.
8. *S is a basis for $\mathbb{R}^n$.*
9. $\operatorname{rank}(A) = n$.
10. $\operatorname{Nullity}(A) = 0$.

This is quite a bit of information so we breifly summarize the main ideas in the following list:

- $\operatorname{Null}(A) = \{\mathbf{x} \in \mathbb{R}^m \colon A\mathbf{x} = \mathbf{0}\} = \{\mathbf{x} \in \mathbb{R}^m \colon T(\mathbf{x}) = \mathbf{0}\} = \ker(T)$.
- $\operatorname{Col}(A) =$ span of columns of $A = \operatorname{Range}(T)$.
- $\dim(\operatorname{Col}(A)) = \dim(\operatorname{Row}(A)) = \operatorname{rank}(A)$.
- $\dim(\operatorname{Null}(A)) = \operatorname{Nullity}(A)$.
- If $T\colon \mathbb{R}^m \to \mathbb{R}^n$ then $Null(A) \subset \mathbb{R}^m$, $\operatorname{Row}(A) \subset \mathbb{R}^m$, and $\operatorname{Col}(A) \subset \mathbb{R}^n$.

4.4 Change of Basis

We now encounter the all important idea surrounding changing a basis. This can be one of the trickiest concepts to understand, but the hard work will pay off. Reading this section several times over may be helpful in gaining a full understanding and when in doubt, do more examples!

Let's first address notation. Let $\mathbf{x} = \begin{bmatrix} 3 \\ -2 \end{bmatrix} \in \mathbb{R}^2$ be written in the standard basis. The coordinates of $\mathbf{x}$ are expressing its geometric location in the plane. That is, to arrive at the tip of the vector $\mathbf{x}$, you move 3 units to the right of the origin (3 units along $\mathbf{e}_1$) and -2 units down from there (-2 units along $\mathbf{e}_2$). This is because

$$\mathbf{x} = 3\mathbf{e}_1 - 2\mathbf{e}_2$$

The coefficients of $\mathbf{x}$ in this expression involving the standard basis are what determine its coordinates. This is the general idea behind change of basis.

Example 4.4.1. Let $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right\}$ be a (non-standard) basis of $\mathbb{R}^2$, In this basis we can express the same vector $\mathbf{x}$ as

$$\mathbf{x} = 14 \begin{bmatrix} 2 \\ 7 \end{bmatrix} - 25 \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

and we express this notationally as

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 14 \\ -25 \end{bmatrix}$$

With this idea in mind, we can now define this notion in greater generality.

Definition 4.4.2. Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a basis for $\mathbb{R}^n$ and let

$$\mathbf{y} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n$$

then **the coordinate vector of y with respect to the basis $\mathcal{B}$** is

$$[\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Let $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \in \mathbb{R}^{n\times n}$. We call U the **change of basis matrix for the basis $\mathcal{B}$** (note that is has the basis vectors as it's columns). If $\mathbf{y}$ is taken to be a vector written in the standard basis, then

$$U[\mathbf{y}]_{\mathcal{B}} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n = \mathbf{y}$$

Example 4.4.3. Let

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 1 \end{bmatrix} \right\}$$

be a basis for $\mathbb{R}^3$ and let $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$. Find $\mathbf{x}$ with respect to the standard basis for $\mathbb{R}^3$.

Given the basis $\mathcal{B}$, our change of basis matrix is

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

so we can find $\mathbf{x}$ via

$$\mathbf{x} = U[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 0 & 5 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ -1 \\ 8 \end{bmatrix}$$

Note that U took a vector **from** the non-standard basis **to** the standard basis.

A natural question one can ask is, how can we go the other direction? That is, if we are given a vector, written in the standard basis, how can we find its representation in some other non-standard basis?

The key is to look at the equation we get from the change of basis matrix, namely

$$\mathbf{x} = U[\mathbf{x}]_{\mathcal{B}}$$

We can see that U is **always** invertible (by the big theorem) because it's columns form a basis, hence we can take the above equation and multiply both sides by U^{-1} on the left to obtain

$$U^{-1}\mathbf{x} = [\mathbf{x}]_{\mathcal{B}}$$

We summarize in the following proposition.

Proposition 4.4.4. *Let $\boldsymbol{x}$ be expressed in the standard basis with $\mathcal{B} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ a non-standard basis for $\mathbb{R}^n$. If $U = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ is the change of basis matrix for the basis $\mathcal{B}$ then*

$$U[\boldsymbol{x}]_{\mathcal{B}} = \boldsymbol{x} \quad \text{and} \quad [\boldsymbol{x}]_{\mathcal{B}} = U^{-1}\boldsymbol{x}$$

Example 4.4.5. Continuing from example 4.4.1, we have $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 7 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix} \right\}$ and $\mathbf{x} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Going from the standard basis to this one we see that

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 14 \\ -25 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 7 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

In words, the proposition is saying that U takes a vector **from** the non-standard basis to the standard basis, and its inverse does the opposite.

What remains is to find a fluid way to go from one non-standard basis to another. The short solution is to "go through the standard basis" but this requires some explination.

Let $\mathcal{B}_1 = \{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ and $\mathcal{B}_2 = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be non-standard bases for $\mathbb{R}^n$. We aim to find a matrix that takes $[\mathbf{x}]_{\mathcal{B}_1}$ as input, and outputs $[\mathbf{x}]_{\mathcal{B}_2}$. Let $\mathcal{B}_{st}$ denote the standard basis for $\mathbb{R}^n$. We carry out the task in two steps

1. Go from $[\mathbf{x}]_{\mathcal{B}_1}$ to $[\mathbf{x}]_{\mathcal{B}_{st}}$.
2. Go from $[\mathbf{x}]_{\mathcal{B}_{st}}$ to $[\mathbf{x}]_{\mathcal{B}_2}$.

We use matrix multiplication to combine the steps.

Theorem 4.4.6. *Let $\mathcal{B}_1 = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ and $\mathcal{B}_2 = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be non-standard bases for $\mathbb{R}^n$ with change of basis matrices given by $U = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ and $V = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ respectively. Then*

$$[\boldsymbol{x}]_{\mathcal{B}_2} = V^{-1}U[\boldsymbol{x}]_{\mathcal{B}_1}$$

and

$$[\boldsymbol{x}]_{\mathcal{B}_1} = U^{-1}V[\boldsymbol{x}]_{\mathcal{B}_2}$$

Proof. We know that U and V are change of basis matrices, hence by Proposition 4.4.4 we know that

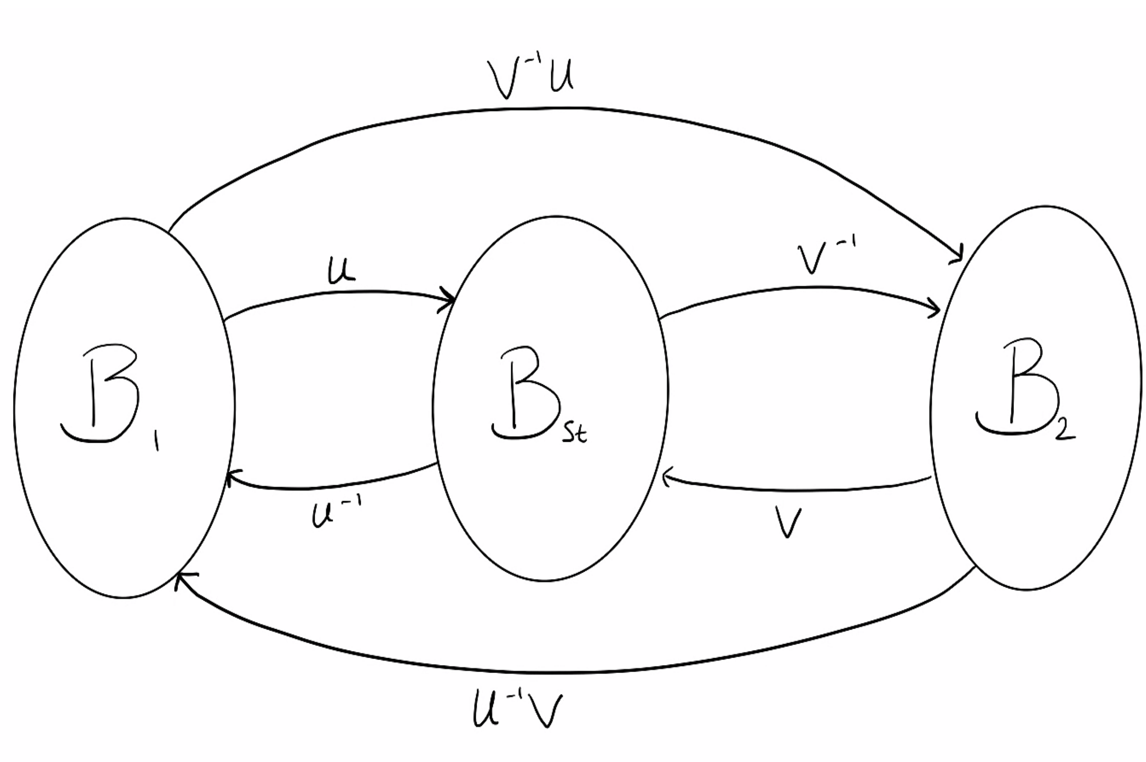
$$\mathbf{x} = U[\mathbf{x}]_{\mathcal{B}_1} \quad \text{and} \quad [\mathbf{x}]_{\mathcal{B}_2} = V^{-1}\mathbf{x}$$

Note here that we are writing $\mathbf{x}$ to mean $[\mathbf{x}]_{\mathcal{B}_{st}}$ (this is standard convention). Combining these two equations we see that

$$[\mathbf{x}]_{\mathcal{B}_2} = V^{-1}\mathbf{x} = V^{-1}(U[\mathbf{x}]_{\mathcal{B}_1}) = V^{-1}U[\mathbf{x}]_{\mathcal{B}_1}$$

This means that the change of basis matrix from $\mathcal{B}_1$ to $\mathcal{B}_2$ is $V^{-1}U$. By taking inverses and using the shoes and socks lemma, we get the second result. □

We end this chapter with an illustration of this entire idea. The whole of change of basis can be summarized in the following picture.



Chapter 5

Determinants

The determinant can be thought of as a useful number that we can associate with a fixed matrix. In particular, viewing it as a function, it takes an $n \times n$ matrix as input and outputs a real number. In this chapter we will begin by discussing ways to compute the determinant of a matrix, and once we have the basics down, we will see how it can be used.

5.1 The Determinant Function

We can compute the determinant of $n \times n$ matrices, for small n, quite easily.

1. $n = 1$: If $A = [a_{11}]$ then $\det(A) = a_{11}$.

2. $n = 2$: If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ then $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

3. $n = 3$: If $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ then

$$\det(A) = a_{11} \det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right) - a_{12} \det\left(\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}\right) + a_{13} \det\left(\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right)$$

The following definition is very formal and can be quite complicated to understand. The best way to get a grasp on it is to do LOTS of examples! Starting with a 3×3 matrix is the best place to begin. For practice (after reading the definition), compute the determinant of $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 4 \\ 5 & 6 & 2 \end{bmatrix}$ and verify that the final answer is 43.

Definition 5.1.1. Let A be an $n \times n$ matrix given by

$$(a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Note that writing a matrix as (a_{ij}) is common compact matrix notation, which denotes that the entry in row i and column j is the real number a_{ij}.
For $n = 2, \ldots, n$, let M_{ij} be that $(n-1) \times (n-1)$ matrix obtained by removing the i^{th} row and j^{th} column

of A. The minor of a_{ij} is the real number $\det(M_{ij})$. The **cofactor** of a_{ij} is $C_{ij} = (-1)^{i+j}\det(M_{ij})$. The **determinant of** A is then the scalar

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

$$= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

where i and j can be any **fixed** values from 1 up to n. Note that the first equation is known as the **cofactor expansion along the i^{th} row**, and the second equation is known as the **cofactor expansion along the j^{th} column**.

The profound fact concerning computation of determinants is the following.

Theorem 5.1.2. *Given any $n \times n$ matrix A, the value of $\det(A)$ obtained by performing a cofactor expansion along the i^{th} row or j^{th} column is always the same.*

The main consequence of this theorem is that, when computing the determinant of a matrix, we can seek out the least labor intensive method possible. In practice, this involves finding the row or column of the given matrix that has the most zeroes, and computing a cofactor expansion along that row or column. This is always the least computationally expensive method.

Example 5.1.3. Compute the determinant of

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

Observe that the third column of A has the most zeroes. Moreover, if we compute the cofactor expansion along the 3rd column, we will only need to compute one minor explicitly (as opposed to as many as 4!).

$$\det(A) = 2\det\left(\begin{bmatrix} 1 & 0 & 3 \\ 5 & 6 & 7 \\ 3 & 1 & 8 \end{bmatrix}\right) = 2\Big(1 \cdot \det\left(\begin{bmatrix} 6 & 7 \\ 1 & 8 \end{bmatrix}\right) + 3 \cdot \det\left(\begin{bmatrix} 5 & 6 \\ 3 & 1 \end{bmatrix}\right)\Big) = 4$$

5.2 Properties of the Determinant

One other way in which we can compute a determinant is to row reduce the given matrix and track how the determinant changes at each step. We do so according to the following proposition.

Proposition 5.2.1. *Suppose B is an $n \times n$ matrix obtained by performing one of the following row operations on A. The determinant of B and A are related as follows:*

1. *Switch two rows of A to get $B \implies \det(B) = -\det(A)$.*
2. *Multiply a row of A by a non-zero constant c to get $B \implies \det(B) = c\det(A)$.*
3. *Add a multiple of one row to another to get $B \implies \det(B) = \det(A)$.*

This Proposition implies the following shortcuts:

- If A has a row or column of zeroes then $\det(A) = 0$.
- If A has a two identical rows then $\det(A) = 0$.

We also add one more useful trick in computing determinants of triangular matrices.

Proposition 5.2.2. *If A is a triangular matrox, then* $\det(A)$ *is the product of the diagonal entries of A.*

Proof. You can do this one yourself! Try drawing an arbitrary 3×3 upper triangular matrix (with entries labeled a_{ij}), then compute the determinant by doing a cofactor expansion along the first column or third row. □

A nice (but sort of obvious) corollary of this is the following.

Corollary 5.2.3. $\det(I_n) = 1$

We now end this chapter with arguably the most useful and important theorems concerning determinants.

Theorem 5.2.4. *Let A be an $n \times n$ matrix. Then A is invertible if and only if* $\det(A) \neq 0$.

Proof. We know that here exists a sequence of row operations taking A to B, where B is in reduced echelon form. This means that every row of B contains a pivot, or the main diagonal has at least one 0 entry. Since B is triangular and $A \sim B$, we know that $\det(A) = c\det(B)$ for some non-zero scalar c. Based on both possibilities for the diagonal entries, we can conclude that if A was invertible, then every column of B is a pivot column so the product of the diagonal entries must be non-zero. If there is a non-pivot column, then there must be a zero entry on the diagonal, hence $\det(B) = 0$. □

The second useful fact concerns the determinant of a product.

Proposition 5.2.5. *If A and B are $n \times n$ matrices, then*

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

The third useful fact, is that when A is invertible, we have a nice explicit form for the determinant of A^{-1}.

Proposition 5.2.6. *If A is invertible, then*

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

Proof. Invertibility of A implies that A^{-1} exists and satisfies the equation $AA^{-1} = I_n$. Taking determinants of both sides and using properties of determinants (which ones?) we conclude that

$$\det(AA^{-1}) = \det(I_n) \implies \det(A)\det(A^{-1}) = 1 \implies \det(A^{-1}) = \frac{1}{\det(A)}$$

□

Before ending the chapter with an addition to the big theorem, we add several interesting notes on how the determinant relates to geometry and area.

Proposition 5.2.7. *Let S denote the unit square in $\mathbb{R}^2$ and let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map with associated matrix A. If $P = T(S)$ denotes the image of the unit square under T, then we have Area$(P) = |\det(A)|$.*

This means that if $\det(A) = 1$, the associated linear transformation preserves area. An example of this is rotation. Building on this, we have a similar result in higher dimensions.

Proposition 5.2.8. *Let D be a region of finite volume in R^n and suppose $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is a linear map with associated matrix A. If $T(D)$ denotes the image of D under T, then Volume$(T(D)) = |\det(A)| \cdot$ Volume(D).*

We now end with an updated (and very powerful!) big theorem.

Theorem 5.2.9. *Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ with associated linear transformation given by $T \colon \mathbb{R}^n \to \mathbb{R}^n$. The following statements are equivalent:*

1. *S spans $\mathbb{R}^n$.*
2. *S is linearly independent.*
3. *The system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b} \in \mathbb{R}^n$.*
4. *T is onto.*
5. *T is one-to-one.*
6. *A is invertible.*
7. $\ker(T) = \{\boldsymbol{0}\}$.
8. *S is a basis for $\mathbb{R}^n$.*
9. $\operatorname{rank}(A) = n$.
10. $\operatorname{Nullity}(A) = 0$.
11. $\det(A) \neq 0$.

Chapter 6

Eigenvalues and Diagonalization

All of our hard work thus far will finally pay off in this chapter. Much of linear algebra past this point is centered around the idea of eigenvalues and eigenvectors and it is certainly something you will want to remember for future classes in any stem field..

6.1 Eigenvalues and Eigenvectors

Let's quickly recall the basics of the geometry of linear transformations. Given a linear map $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $T(\mathbf{x}) = A\mathbf{x}$, for some 2×2 matrix A, we can plug in any vector $\mathbf{x} \in \mathbb{R}^2$ and T will output a new vector $A\mathbf{x}$ with a (potentially different) direction and length. The idea of eigenvalues and eigenvectors investigates when the direction and/or length of the output vector is related to the input vector.

Definition 6.1.1. Let A be an $n \times n$ matrix. If $\mathbf{u}$ is a non-zero vector and $\lambda \in \mathbb{R}$ is a scalar such that $A\mathbf{u} = \lambda\mathbf{u}$, then λ is an **eigenvalue** of A and $\mathbf{u}$ is an **eigenvector** of A associated with eigenvalue λ.

There are a few fundamental facts concerning eigenvectors that will allow us to gain extra structure on the set of all eigenvectors associated to some fixed eigenvalue. The first is that the sum of two eigenvectors associated to the same eigenvalue is another (different!) eigenvector associated to the same eigenvalue (you should verify this for yourself). We also have a related result.

Proposition 6.1.2. *Suppose A is a square matrix and λ is an eigenvalue of A with associated eigenvector $\boldsymbol{u}$, that is, $A\boldsymbol{u} = \lambda\boldsymbol{u}$. Then for any non-zero scalar c, we have that $c\boldsymbol{u}$ is en eigenvector of A associated to λ.*

Proof. If $A\mathbf{u} = \lambda\mathbf{u}$ then A being linear implies that for any $c \in \mathbb{R}$

$$A(c\mathbf{u}) = cA\mathbf{u} = c\lambda\mathbf{u} = \lambda(c\mathbf{u})$$

hence $c\mathbf{u}$ is en eigenvector of A associated to eigenvalue λ. □

Combining the last two facts, we obtain the notion of eigenspaces.

Definition 6.1.3. Let A be an $n \times n$ matrix with eigenvalue λ. The set $\mathcal{S}$ consisting of the zero vector and all eigenvectors of A associated with λ forms a subspace of $\mathbb{R}^n$ known as the **eigenspace associated to eigenvalue** λ, often denoted by E_λ.

Example 6.1.4. Let $A = \begin{bmatrix} 6 & -2 \\ 5 & -1 \end{bmatrix}$. One can check that if $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$ then $A\mathbf{u} = 4\mathbf{u}$ and $A\mathbf{v} = \mathbf{v}$. This means that $\mathbf{u}$ is an eigenvector of A of eigenvalue 4 and $\mathbf{v}$ is an eigenvector of A with eigenvalue 1. It follows (by reasons we will soon see) that the eigenspace of eigenvalue 4 is

$$E_4 = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

and the eigenspace of eigenvalue 1 is

$$E_1 = \text{Span}\left\{ \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right\}$$

What we need moving forward is a streamlined way to find eigenvalues and a basis for each associated eigenspace, when given an arbitrary matrix A. What follows is the reasoning behind how we find eigenvalues.

If we have an eigenvalue/eigenvector pair so that $A\mathbf{u} = \lambda\mathbf{u}$ for some vector $\mathbf{u}$ and scalar λ, then we can obtain the closely related equation

$$A\mathbf{u} - \lambda\mathbf{u} = \mathbf{0}$$

By rewriting $\mathbf{u}$ as $I\mathbf{u}$, where I is the $n \times n$ identity matrix, the above equation can be more compactly written as

$$(A - \lambda I)\mathbf{u} = \mathbf{0}$$

Note that $\lambda I = \begin{bmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{bmatrix}$ and if $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ then

$$\lambda I\mathbf{u} = \begin{bmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} \lambda u_1 \\ \vdots \\ \lambda u_n \end{bmatrix} = \lambda\mathbf{u}$$

so the expression $A - \lambda I$ does indeed make sense. With this equation being understood, we can now classify how one finds eigenvalues of a matrix.

Proposition 6.1.5. *Let A be an $n \times n$ matrix. A scalar $\lambda \in R$ is an eigenvalue of A if and only if* $\det(A - \lambda I) = 0$.

Proof. Summarizing what was said above, we have that λ is an eigenvalue of A if and only if $A\mathbf{u} = \lambda\mathbf{u}$ for some vector $\mathbf{u} \neq \mathbf{0}$ if and only if $A\mathbf{u} - \lambda I\mathbf{u} = \mathbf{0}$ if and only if $(A - \lambda I)\mathbf{u} = \mathbf{0}$. This means that λ is an eigenvalue of A if and only if the homogeneous equation $(A - \lambda I)\mathbf{u} = \mathbf{0}$ has a non-trivial solution, and this is true if and only if $A - \lambda I$ is **not** invertible (by the big theorem). It follows that $A - \lambda I$ is **not** invertible if and only if $\det(A - \lambda I) = 0$ which completes the proof. □

The heart of our method lies in this proof. We will soon see that $\det(A - \lambda I)$ is a polynomial in the variable λ (note that λ is merely a placeholder at first and the values of λ that satisfy $\det(A - \lambda I) = 0$ are the eigenvalues of A). Looking more closely at the polynomial $\det(A - \lambda I) = 0$, we will see that the eigenvalues of A are the roots of this polynomial. The above proposition then takes the task of finding eigenvalues to the task of finding roots of a polynomial. In general we call $\det(A - \lambda I)$ the **characteristic polynomial of** A.

Example 6.1.6. Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$. Find the eigenvalues and a basis for each eigenspace.

We first see that

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

and

$$\det(A - \lambda I) = (1-\lambda)(-\lambda) - 2 = \lambda^2 - \lambda - 2 = (\lambda - 2)(\lambda + 1)$$

We need the solutions to $\det(A - \lambda I) = 0$ and these are the solutions to $(\lambda - 2)(\lambda + 1) = 0$ hence $\lambda = 2$ and $\lambda = -1$ are the eigenvalues of A. It is worth noting that no other scalars are eigenvalues of A, these two are the only ones. To find bases for the eigenspaces, we then only need to find the vectors $\mathbf{u}$ and $\mathbf{v}$ respectively, that satisfy $A\mathbf{u} = 2\mathbf{u}$ and $A\mathbf{v} = -\mathbf{v}$.

If $A\mathbf{x} = \lambda\mathbf{x}$ for some eigenvalue λ, then $\mathbf{x}$ satisfies the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$. In other words, all the eigenvectors with eigenvalue λ are precisely the vectors in $\text{Null}(A - \lambda I)$! This means that the eigenspace for eigenvalue λ is the same thing as $\text{Null}(A - \lambda I)$. That is

$$E_\lambda = \text{Null}(A - \lambda I)$$

We can now find a basis for E_{-1}. We need to find a basis for $Null(A - \lambda I)$ with $\lambda = -1$ so we plug in $\lambda = -1$ to $A - \lambda I$ and we get

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

We see that all vectors in the null space of this matrix are of the form $t\begin{bmatrix} -1 \\ 2 \end{bmatrix}$ for some free variable t, hence the basis for E_{-1} is $\left\{ \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$. We leave the computation of a basis for E_2 to the reader as practice. The answer you should get is

$$\mathcal{B}_{E_2} = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

Next, we outline so shortcuts that can be used in finding eigenvalues of simple types of matrices.

Example 6.1.7. Find the eigenvalues of the triangular matrix

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

By computing the characteristic polynomial of A we see that

$$Det(A - \lambda I) = \det\left(\begin{bmatrix} 1-\lambda & 2 & 0 & 1 \\ 0 & 1-\lambda & 0 & 1 \\ 0 & 0 & 2-\lambda & 0 \\ 0 & 0 & 0 & 2-\lambda \end{bmatrix} \right)$$

Recalling that the determinant of a triangular matrix is the product of the diagonal entries, it follows that

$$\det(A - \lambda I) = (1-\lambda)^2(2-\lambda)^2$$

Looking back at the matrix A, we can see that the eigenvalues of A are exactly the diagonal entries. This is in fact true for eigenvalues of all triangular matrices.

Next, let's find bases for the eigenspaces E_1 and E_2, using some shortcuts along the way.

To compute a basis for E_1, we need to find a basis for $\text{Null}(A - I)$ which is

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

We can see that this matrix has three pivot columns hence $\text{rank}(A - I) = 3$. By rank-nullity this means its null space is 1 dimensional, hence is of the form $\text{Span}\{\mathbf{x}\}$ for some non-zero vector $\mathbf{x} \in \mathbb{R}^4$. Since the first column of $A - I$ is the zero vector, this means that $A - I$ sends $\mathbf{e}_1$ to $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ hence

$$\mathcal{B}_{E_1} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\}$$

As an exercise, we leave the computation of a basis for E_2 to the reader, but to check your work you should get

$$\mathcal{B}_2 = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

No tricks to doing this one, just compute the correct null space in the usual way.

Looking back at the example above, we can see that one of the eigenspaces was one-dimensional while the other one was two dimensional. This is a phenomenon that is subtle and requires a bit more discussion.

The first thing to note is that if A is an $n \times n$ matrix, then the characteristic polynomial $\det(A - \lambda I)$ is always a degree n polynomial. The concept we will need to understand further is that of (algebraic) multiplicity of a root of a polynomial, which we now define.

Definition 6.1.8. Let $P(x)$ denote a polynomial of degree n, in the variable x, and suppose $P(x)$ is the characteristic polynomial of some matrix A. If we can factor this polynomial as

$$P(x) = (x - \alpha)^m Q(x)$$

where $Q(\alpha) \neq 0$, then we say $x = \alpha$ is an eigenvalue of A with **multiplicity m**. In other words, the exponent attached to the linear term of a polynomial is the multiplicity we associate to the root of that polynomial that comes from the given linear term.

This notion of multiplicity is precisely what we need to say more about dimensions of eigenspaces.

Theorem 6.1.9. *Let λ be an eigenvalue of a matrix A and let $m(\lambda)$ denote the multiplicity of the eigenvalue λ. Then we always have*

$$\dim E_\lambda \leq m(\lambda)$$

That is, the dimension of the eigenspace for eigenvalue λ never exceeds the multiplicity of that eigenvalue.

Looking back at the previous example, we can see that both eigenvalues 1 and 2 have multiplicity 2, yet $\dim E_1 = 1$ and $\dim E_2 = 2$. The inequality holds in both cases but we only obtained equality in one. There is lots more that one can say about multiplicities of eigenvalues but we leave it at this for now, and say a bit more in the next section. We now end this section with one more important fact, which will be our last addition to the big theorem.

Proposition 6.1.10. $\lambda = 0$ *is not an eigenvalue of A if and only if* $\det(A) \neq 0$.

Proof. We show that $\lambda = 0$ is an eigenvalue of A if and only if $\det(A) = 0$. We can see that $\lambda = 0$ is an eigenvalue of A if and only if $\det(A - \lambda I) = \det(A - 0I) = \det(A) = 0$, which is all we needed to show. ☐

Since we can only discuss eigenvalues for square matrices, this proposition can extend our list of results coming from the big theorem.

Theorem 6.1.11. *Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ be a set of vectors in $\mathbb{R}^n$ and let $A = \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n \end{bmatrix}$ with associated linear transformation given by $T\colon \mathbb{R}^n \to \mathbb{R}^n$. The following statements are equivalent:*

1. *S spans $\mathbb{R}^n$.*
2. *S is linearly independent.*
3. *The system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every $\boldsymbol{b} \in \mathbb{R}^n$.*
4. *T is onto.*
5. *T is one-to-one.*
6. *A is invertible.*
7. $\ker(T) = \{\boldsymbol{0}\}$.
8. *S is a basis for $\mathbb{R}^n$.*
9. $\operatorname{rank}(A) = n$.
10. $\operatorname{Nullity}(A) = 0$.
11. $\det(A) \neq 0$.
12. *$\lambda = 0$ is not an eigenvalue of A.*

6.2 Diagonalization

Let's jump right in.

Definition 6.2.1. An $n \times n$ matrix is **diagonalizable** if there exists $n \times n$ matrices Λ and X such that

- Λ is diagonal.
- X is invertible.

and

$$A = X\Lambda X^{-1}$$

Note that Λ is the greek capital letter for λ. This is intentional, and we will soon see that the diagonal entries of Λ are precisely the eigenvalues of A.

Example 6.2.2. If $X = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$ and $\Lambda = \begin{bmatrix} 4 & 0 \\ 0 & -3 \end{bmatrix}$, then $X^{-1} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$ and $X\Lambda X^{-1} = \begin{bmatrix} 4 & 0 \\ 14 & -3 \end{bmatrix}$. If $A = \begin{bmatrix} 4 & 0 \\ 14 & -3 \end{bmatrix}$ then we say A is diagonalizable.

This example doesn't help much. In general, we need to find a way to construct the matrices X and Λ that diagonalize A, and in doing so, we will see when a given matrix is not diagonalizable. Before embarking on this adventure, it is worth noting one of the many reasons why diagonalization is useful. In many applied fields, systems can be modeled by matrix multiplication and iterates of our system can be taken via computing powers of a matrix. If the given matrix is diagonalizable, computing powers can be very easy.

Assume that A is diagonalizable so that we can write $A = X\Lambda X^{-1}$, then

$$A^2 = (X\Lambda X^{-1})(X\Lambda X^{-1}) = X\Lambda^2 X^{-1}$$

and

$$A^3 = A^2 A = (X\Lambda^2 X^{-1})(X\Lambda X^{-1}) = X\Lambda^3 X^{-1}$$

Continuing this process we can see that

$$A^k = X\Lambda^k X^{-1}$$

and since Λ is a diagonal matrix, computing powers of it is excessively easy. We now begin the investigation of when A is diagonalizable by stating the main result and digging into the details.

Theorem 6.2.3. *An $n \times n$ matrix A is diagonalizable if and only if A has eigenvectors that form a basis for $\mathbb{R}^n$.*

There are a few important things to point out regarding what we mean when we say eigenvectors.

Proposition 6.2.4. *If λ_1 and λ_2 are eigenvalues of a matrix A and $\lambda_1 \neq \lambda_2$, then if $\boldsymbol{x} \in E_{\lambda_1}$ and $\boldsymbol{y} \in E_{\lambda_2}$ it is always true that $\{\boldsymbol{x}, \boldsymbol{y}\}$ form a linearly independent set. That is, eigenvectors corresponding to different eigenvalues are always linearly independent.*

This means that when given a square matrix A we can

1. Find all the eigenvalues of A (by finding roots of $\det(A = \lambda I) = 0$).
2. Find bases for all eigenspaces.
3. Put all basis vectors from different eigenspaces in a set and see if this set forms a basis for $\mathbb{R}^n$. By way of the above proposition, we know that the eigenvectors will form a basis (called an eigenbasis) if there are n of them.

We now prove Theorem 6.2.3.

Proof. Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be n (linearly independent) eigenvectors for a matrix A, with eigenvalues labeled $\lambda_1, \ldots, \lambda_n$ (note here that we are assuming there are n distinct eigenvalues for simplicity of the proof but this is not always the case), so that $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ forms an (eigen)basis for $\mathbb{R}^n$. Let $X = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix}$ and

$$\Lambda = \begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}$$

be the diagonal matrix with eigenvalues on the diagonal. Since $\mathcal{B}$ is a basis for $\mathbb{R}^n$ we know that X is invertible, by the big theorem. Looking at the matrix multiplication, we see that

$$AX = A\begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix} = \begin{bmatrix}A\mathbf{u}_1 & \ldots & A\mathbf{u}_n\end{bmatrix} = \begin{bmatrix}\lambda_1\mathbf{u}_1 & \ldots & \lambda_n\mathbf{u}_n\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix} = X\Lambda$$

Since $AX = X\Lambda$ we can conclude that $A = X\Lambda X^{-1}$ and A is diagonalizable. □

Example 6.2.5. Let $A = \begin{bmatrix}4 & -2 \\ 4 & -2\end{bmatrix}$ and show that A is diagonalizable by finding matrices X and Λ such that $A = X\Lambda X^{-1}$.

We begin by finding the eigenvalues of A as well as bases for the eigenspaces. By computing the characteristic polynomial for A we see that

$$\det(A - \lambda I) = -(4-\lambda)(2+\lambda) + 8 = -(8 + 2\lambda + \lambda^2) + 8 = \lambda^2 - 2\lambda = \lambda(\lambda - 2) = 0$$

This means that $\lambda = 0, 2$ are the eigenvalues of A. In computing bases for both eigenspaces, we just need to find bases for $\mathrm{Null}(A)$ and $\mathrm{Null}(A - 2I)$ respectively. We get that

$$\mathrm{Null}(A) = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\} \quad \text{and} \quad \mathrm{Null}(A = 2I) = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$$

Following the proof of Theorem 6.2.3, we set $X = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$ and $\Lambda = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ and these are precisely the matrices that diagonalize A. Note the the order in which we place the vectors is very important, if we swapped the order of 0 and 2, on the diagonal of Λ, while leaving the columns of X unchanged, the resulting matrix product would not equal A. You must always have the columns of X correspond, in the same order, with the eigenvalues for those column vectors. Moreover, if you have an eigenvalue of multiplicity k, then there will be exactly k diagonal entries of Λ that are equal to that given eigenvalue.

Example 6.2.6. Construct a 3×3 matrix A with the following eigenvalues and eigenvectors.

$$\lambda_1 = 2, \mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \lambda_2 = -1, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \qquad \lambda_3 = 5, \mathbf{u}_3 = \begin{bmatrix} 4 \\ 4 \\ 2 \end{bmatrix}$$

This can quickly be done by working backwards through the mechanics of the proof of Theorem 6.2.3. Let

$$X = \begin{bmatrix} 0 & -1 & 4 \\ 1 & 0 & 4 \\ 0 & 2 & 2 \end{bmatrix} \qquad \text{and} \qquad \Lambda = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

By computing X^{-1} (which we know exists), the resulting matrix $X\Lambda X^{-1}$ will have the prescribed eigenvalues and eigenvectors.

We now use the full strength of Theorem 6.2.3.

Example 6.2.7. Is $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ diagonalizable?

It suffices to see if there exists a basis of eigenvectors of A. We have

$$\det(A - \lambda I) = \det\left(\begin{bmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{bmatrix} \right) = (1-\lambda)^2$$

hence 1 is the only eigenvalue of A, with multiplicity 2. Since we need a basis of eigenvectors in order to diagonalize A, we must have the dimension of E_1 be equal to 2. If it is not, then there is no way for our eigenvectors to form a basis for $\mathbb{R}^2$, since we will not have enough of them. In computing the eigenspace we see that

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

If we wanted, we could stop right here since the null space of this matrix can never be 2 dimensional, because it has rank 1 (rank-nullity is being used here). If we want to be more explicit, we can directly compute $E_1 = \mathrm{Null}(A - I)$ and find that

$$\mathrm{Null}(A - I) = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

Regardless of the argument we prefer, we can now see that A is not diagonalizable.

Shortcuts like this one can be very helpful in practice. There is one shortcut in particular that can really come in handy.

Proposition 6.2.8. *Let A be an $n \times n$ matrix and assume that $\{\lambda_1, \dots, \lambda_n\}$ are **distinct** eigenvalues (the word distinct here means that there are exactly n eigenvalues, no two of which are equal), then A is always diagonalizable.*

Proof. If A has n distinct eigenvalues, then the characteristic polynomial of A has exactly n distinct roots. That is,

$$\det(A - \lambda I) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n)$$

This means that the multiplicity of each eigenvalue is 1. Now, recalling that $\dim(E_\lambda) \leq m(\lambda)$ for all eigenvalues, we have that $\dim(E_{\lambda_i}) \leq m(\lambda_i) = 1$ for all $i = 1, \dots, n$, hence we must have $\dim(E_{\lambda_i}) = 1$ for all i, because eigenspaces of actual eigenvalues of a matrix are never 0 dimensional (in fact, the only instance when $E_\lambda = \{\mathbf{0}\}$ for some matrix A is when λ is **not** and eigenvalue of A). This means that we get exactly one (linearly independent) eigenvector coming from each eigenspace, of which there are n in total. Putting them all together in one set, we obtain a set of n linearly indepedent vectors in $\mathbb{R}^n$, which (by the theorem) forms our desired eigenbasis. This implies that A is diagonalizable. □

Example 6.2.9. Is $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix}$ diagonalizable?

Recalling our nice little trick for triangular matrices, we can see that the eigenvalues are $1, 2$, and 3 respectively, which are distinct! This means that A is diagonalizable, by the above proposition.

Although this proposition is great, we still need to treat it with care. In particular, not all diagonalizable matrices have distinct eigenvalues.

For an easy example, one should note that the zero matrix is diagonalizable. Moreover, if we consider the identity matrix $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, then we can write I as

$$I = III^{-1}$$

which (in a silly way) satisfies the definition of being diagonalizable. Moreover, we could compute the null space of $I - I$, and see that it admits a basis of eigenvectors (in particular it admits the standard basis).

Before ending this chapter, we provide one last alternative way to check if a matrix is diagonalizable. One can think of this as a generalization of the proposition on distinct eigenvalues.

Proposition 6.2.10. *Suppose A is an $n \times n$ matrix with **only real eigenvalues** (so none of them are complex numbers). A is diagonalizable if and only if the dimension of each eigenspace is equal to the multiplicity of the corresponding eigenvalue.*

In general, one calls the dimension of E_λ the *geometric* multiplicity of λ whereas the usual multiplicity $m(\lambda)$ is known as the *algebraic* multiplicity. This proposition is saying that the geometric multiplicity is always less than or equal to the algebraic multiplicity, and when they are equal, the given matrix is diagonalizable.

Proof. An $n \times n$ matrix A is diagonalizable is and only if it admits n linearly independent eigenvectors (by Theorem 6.2.3). Moreover, each eigenspace has dimension no greater than the multiplicity of the associated eigenvalue, i.e. $\dim(E_\lambda) \leq m(\lambda)$. Since A is an $n \times n$ matrix, the sum of the multiplicities of the eigenvalues must equal n, because $\det(A - \lambda I)$ is a degree n polynomial. Lastly, since the eigenvectors coming from different eigenspaces are always linearly independent, we can conclude that the sum of the dimensions of all eigenspaces equals n. Counting up one basis vector for each dimension, we end up with exactly n eigenvectors, hence a basis for $\mathbb{R}^n$, completing the proof. □