Linear Algebra I (MAT223)
Fall 2024 Lecture Notes

Last Updated: November 25, 2024

Contents

Preface

"Mathematics is the art of reducing any problem to linear algebra."
- William Stein

Linear algebra is the study of vectors, vector spaces ("flat spaces") like lines and planes, and linear transformations like rotations and scalings.

Vectors originated in the study of physics and the 3D world, but through the mathematical practice of abstraction, we now use vectors in non-spatial realms, like music, computer graphics, and the study of physical forces. Transformations are functions that move vectors around, and in this class we will focus on linear transformations. Why? Because we have a complete theory of linear functions. And, although mankind has tried to understand the non-linear phenomena of the universe, we haven't gotten very far. Despite 200 years of effort, the non-linear equations governing fluid flow haven't been solved! Because of this, our approach to answering general questions about the universe is often to convert the problems into linear ones—ones that we can actually understand, and ones which we will study in this course.

This course will cover the fundamentals of linear algebra. We will ground our study in $\mathbb{R}^n$ (n-dimensional Euclidean space), using spatial intuition to guide us. However, we will also pay close attention to the mathematical definitions we encounter along the way. These carefully constructed definitions—the result of hundreds of years of human endeavor—will allow us to solve problems where our intuition fails (for example, how can you find the angle between two 17-dimensional vectors?). In the next course, MAT224, the idea of vectors themselves will be decoupled from Euclidean space and linear algebra will become even more broadly applicable.

These lecture notes will follow the 2024 edition of "Linear Algebra, MAT223 Workbook" by Prof J. Siefken, which is available for purchase as a hard copy in the bookstore, or for free as a pdf (posted on Quercus). These lecture notes will be updated and posted at the end of each week, **after** we discuss the content in class. These notes are written to give a terse summary of material covered in class to use as a reference as you study and learn.

Please keep in mind that these notes are only to be used for MAT223 at the University of Toronto, and are not to be distributed. If (and when) you find typos/mistakes, please feel free to let us know by filling out this google form.

List of Notation.

The following list will be updated as new notation appears in the notes.

$\mathbb{R}$	the set of real numbers
$\mathbb{C}$	the set of complex numbers
$\mathbb{Q}$	the set of rational numbers
$\mathbb{Z}$	the set of integers
$\mathbb{R}^n$	n-dimensional Euclidean space
$\in$	is an element of
$\forall$	for all
$\exists$	there exists
$[\vec{x}]_{\mathcal{B}}$	the coordinates of a vector $\vec{x}$ with respect to the basis $\mathcal{B}$
$\dim(V)$	the dimension of a vector space V
T_A	the linear transformation corresponding to the matrix A, given by $T_A(\vec{x}) = A\vec{x}$.
$\ker F$	the kernel of a function F
$\operatorname{im} F$	the image of a function F
$\operatorname{Nul}(A)$	the null space of a matrix A
$\operatorname{Col}(A)$	the column space of a matrix A

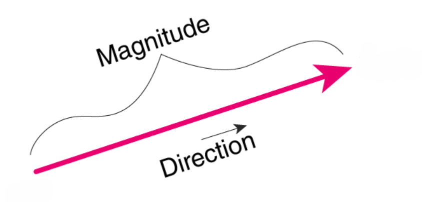
$\text{Row}(A)$	the row space of a matrix A
$A^\top$	the matrix transpose
$[F]_\mathcal{B}$	the defining matrix of a transformation F with respect to the basis $\mathcal{B}$
$\det(A)$	the determinant of A
E_λ	the λ-Eigenspace
χ_A	the characteristic polynomial of A
$\vec{x} \cdot \vec{y}$	the dot product
$\\|\vec{x}\\|$	the norm
$d(\vec{x}, \vec{y})$	the distance between vectors $\vec{x}$ and $\vec{y}$
$V^\perp$	the orthogonal complement of a vector space V
$\vec{x}_V$	the orthogonal projection of a vector $\vec{x}$ onto a vector space V
$\text{proj}_{\vec{v}}\vec{u}$	the orthogonal projection of $\vec{u}$ onto the vector space $V = \text{Span}(\vec{v})$.

Vectors

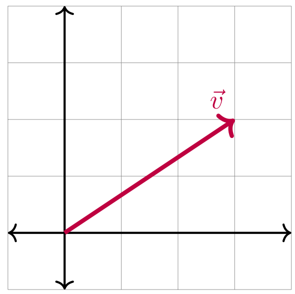
1.1. Vectors in $\mathbb{R}^2$

If someone asks you the question "Where are you?" typically you'd give a response like "I'm near the library, two blocks south of the main entrance". That is, a common way to describe location is to give the displacement (e.g. "two blocks south") from some distinguished point (e.g. "the main entrance of the library"). Furthermore, displacement is described by distance (e.g. "2 blocks") and direction (e.g. "south"). Vectors, our first topic of study in this course, are the mathematical object we use to represent displacement. In this section, we focus only on vectors in $\mathbb{R}^2$ (2-dimensional Euclidean space), and later on generalize the discussion to $\mathbb{R}^n$.

Definition 1.1. A VECTOR is a mathematical object, typically drawn as an arrow, with a magnitude (i.e. length or distance) and direction.



There are a few ways that we can indicate the magnitude and direction of a vector. One common way to do this is to keep track of the displacement along the axes of a standard coordinate grid. For example, consider the vector in $\mathbb{R}^2$ below drawn on a standard coordinate grid



Observe that the displacement this vector represents can be understood by displacement in the x-direction by 3 units and displacement in the y-direction by 2 units. We typically denote this by listing the displacement in each direction, and then wrapping that list up in brackets, as seen below

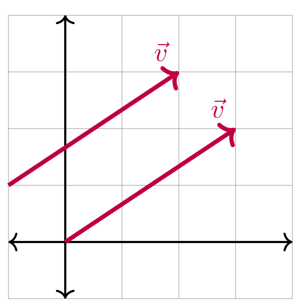
$$\vec{v} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

This is called the STANDARD COORDINATE REPRESENTATION of $\vec{v}$. In general, we have the following.

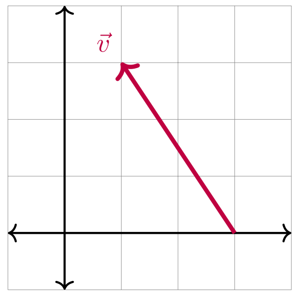
Definition 1.2. Let $\vec{v}$ be the vector which represents displacement in the x direction by v_1 units and displacement in the y direction by v_2 units. Then, the STANDARD COORDINATE REPRESENTATION of $\vec{v}$ is given by $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$.

Note that different resources use slightly different notation for vectors. Throughout this course, we'll generally stick with vectors written in a column. Parentheses and brackets can be used interchangeably, but in these lecture notes we'll tend to use parentheses.

1.1.1. Unrooted Vectors. Note that in the example above, we drew our vector starting from the origin $(0, 0)$. However, it's possible to shift a vector around the plane without changing its direction and magnitude. So, at least in this course, our vectors will be "unrooted", meaning they can "start anywhere." Recall that our vector $\vec{v}$ is the vector which represents displacement in the x direction by 3 and displacement in the y direction by 2, which can be represented as an arrow starting at $(0, 0)$, or at $(-1, 1)$, or anywhere else we'd like



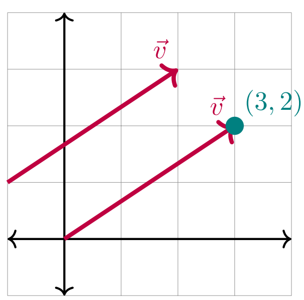
Example 1.3. Find the standard coordinate representation of the vector $\vec{v}$ drawn below.



Answer: $\boxed{\vec{v} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}}$

1.1.2. Vectors as Points, Points as Vectors. In this course, we'll be studying some "algebraic properties" of sets of vectors. One convenient way to do this is to correspond vectors to points in Euclidean space, which will allow us to think of collections of vectors as geometric objects in Euclidean space. The correspondence works as follows: Each vector $\begin{pmatrix} x \\ y \end{pmatrix}$ will correspond to the point (x, y) in $\mathbb{R}^2$.

Graphically, we can visualize this as our vectors corresponding to the point the tip sits at when rooted at the origin. In the example below, we see that the vector $\vec{v}$ corresponds to the point $(3, 2)$, which is simpler to see visually when we root $\vec{v}$ at the origin $(0, 0)$.

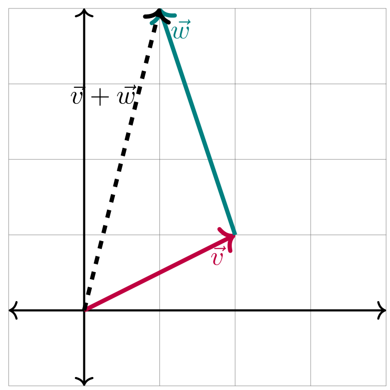


1.1.3. Vector Arithmetic. Suppose that someone gave you the following directions: first walk two blocks west and one block north, and then walk one block east and three blocks north. Assuming we're in a city whose streets are on a grid, this would be the same as walking one block west and four blocks north. Since it doesn't matter what order we go north/south or east/west, we can figure this out by adding up the total north/south displacement and the total east/west displacement to find the total displacement. This is how we'll define vector addition.

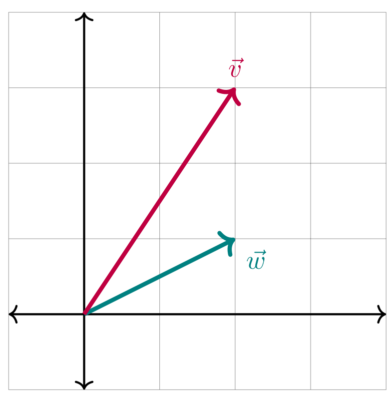
Definition 1.4. Let $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ and $\vec{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$ be vectors in $\mathbb{R}^2$ and let α be a scalar (that is, α is a real number). We define the following

$$\vec{v} + \vec{w} := \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix}, \text{ and } \alpha\vec{v} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \end{pmatrix}.$$

Geometrically, we can visualize vector addition by “stacking” vectors head to tail, as in the example below



Example 1.5. Let $\vec{v}$ and $\vec{w}$ be the vectors drawn below. Find the standard coordinate representation of the vector $\vec{v} - 2\vec{w}$.



Answer: $\boxed{\begin{pmatrix} -2 \\ 1 \end{pmatrix}}$

1.1.4. Linear Combinations and the Standard Basis. In the magic carpet ride problems, we investigated what vectors can be attained as "linear combinations" of others. This will be one of our main topics in the course.

Definition 1.6. A LINEAR COMBINATION of vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ is a vector of the form

$$\vec{w} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n$$

where the $\alpha_1, \alpha_2, \dots, \alpha_n$ are scalars called the COEFFICIENTS of the linear combination.

There are two special vectors that will show up for us throughout the course.

Definition 1.7. The STANDARD BASIS VECTORS for $\mathbb{R}^2$ are the vectors

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

These vectors are important for a couple of reasons:

(1) Every vector $\vec{v}$ in $\mathbb{R}^2$ is a linear combination of $\vec{e}_1$ and $\vec{e}_2$, and
(2) If we only allow ourselves to use one of these vectors, we cannot reach every vector in $\mathbb{R}^2$.

The two conditions above will be formalized later on to define bases more generally. Observe that if we take any vector $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ in $\mathbb{R}^2$ observe that we can write

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2.$$

So, the coordinates of a vector $\vec{v}$ are precisely the coefficients of $\vec{v}$ when written as a linear combination of the standard basis.

1.2. Higher Dimensional Vectors

We can generalize our definitions from the previous section to "n-dimensional Euclidean space". Note that this generalization doesn't have any tangible geometric meaning when $n \geq 4$ (at least for those of us that exist in 3-dimensional space). However, many applications of linear algebra use this generalization to imagine a geometry in higher dimensions. So, if you can translate your object of study (which may have no geometric meaning at all) to one in linear algebra, you can often use geometric tools to solve it. This is a powerful and really cool technique, don't let that be lost on you.

Definition 1.8. An n-DIMENSIONAL VECTOR is a list of n real numbers in a specified order, which we'll write in the form

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

We call the real numbers v_i the STANDARD COORDINATES of the vector $\vec{v}$, and we denote the set of all n-dimensional vectors by $\mathbb{R}^n$.

We can define our operations on n-dimensional vectors just as in $\mathbb{R}^2$.

Definition 1.9. Let $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ and $\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$ be vectors in $\mathbb{R}^n$ and let α be a scalar. We define the following

$$\vec{v} + \vec{w} := \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix}, \text{ and } \alpha\vec{v} = \begin{pmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{pmatrix}.$$

Now that we've defined vector arithmetic, linear combinations of vectors are defined identically to Definition 1.6.

Definition 1.10. The STANDARD BASIS VECTORS in $\mathbb{R}^n$ are the vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ in $\mathbb{R}^n$ given by

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \vec{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

That is, $\vec{e}_i$ is the n-dimensional vector with 1 in the ith coordinate and 0s everywhere else.

Just as in $\mathbb{R}^2$, the standard coordinates of any vector $\vec{v}$ in $\mathbb{R}^n$ are precisely the coefficients when $\vec{v}$ is written as a linear combination of the standard basis vectors. More precisely,

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

precisely when $\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + \cdots + v_n\vec{e}_n$.

Example 1.11. Let $\vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^4$ given by $\vec{v} = \vec{e}_1 - \vec{e}_3$ and $\vec{w} = \vec{e}_1 - \vec{e}_2 + 3\vec{e}_4$. Find the standard coordinate representation of the vector

$$\vec{u} = 2\vec{v} + \vec{w}.$$

Answer: $\boxed{\vec{u} = \begin{pmatrix} 3 \\ -1 \\ -2 \\ 3 \end{pmatrix}}$.

Chapter 2

Systems of Linear Equations

2.1. Introduction to Systems of Linear Equations

Recall, to solve the first magic carpet ride problem, we needed to find real numbers x, y so that

$$x \begin{pmatrix} 3 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 107 \\ 64 \end{pmatrix}.$$

Observe that this is the same problem as finding real numbers x and y that satisfy both of the equations below

$$3x + y = 107$$

$$x + 2y = 64.$$

This is an important correspondence we'll exploit this semester: finding coefficients of a linear combination precisely correspond to the solutions of a system of linear equations. Let's turn our attention now to finding a systematic method to solve any system of linear equations.

Definition 2.1. A LINEAR EQUATION in variables $x_1, x_2, \ldots, x_n$ is an equation of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, where $a_1, \ldots, a_n$ and b are real numbers. In this class, we will often work with systems of equations in four or less variables. For notational convenience, we will often use the letters x, y, z, w and to indicate our variables, rather than x_1, x_2, x_3, x_4.

Example 2.2. Which of the following equations are linear?

(1) $2x + y + z = 3$

(2) $x^2 - y = 3$

(3) $\pi^2(x + y) = \dfrac{z}{2}$

(4) $e^x + \sqrt{y} = z^3$

Answer: $\boxed{\text{(1) and (3)}}$

Definition 2.3. A SYSTEM OF LINEAR EQUATIONS is a collection of one or more linear equations in the same variables. A tuple $(s_1, \dots, s_n) \in \mathbb{R}^n$ is a SOLUTION to a system of linear equations if $(s_1, \dots, s_n)$ is a solution to every linear equation in the system.

Example 2.4. Observe that $(1, 0, 2, -1)$ is a solution to the system of linear equations

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

Example 2.5. The system of linear equations

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

has exactly one solution, given by $\boxed{\text{x=2, y=0, z=1}}$.

Example 2.6. The system of linear equations

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

has $\boxed{\text{no solutions}}$.

Example 2.7. The system of linear equations

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

has $\boxed{\text{infinitely many solutions}}$.

Our goal in this course will be to come up with an algorithm to solve systems of linear equations. In the examples above, you explored some methods of solving systems of linear equations, perhaps using ad-hoc methods or guess and check. Very likely, though, most of you performed some operations to obtain an "equivalent system" that was simpler to solve.

Definition 2.8. We say that two systems of equations are *equivalent* if they have the same set of solutions.

Example 2.9. The only solution to the system of equations

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

is equal to $(x, y) = (1, 1)$. Similarly, the system of equations

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

has the one solution $(x, y) = (1, 1)$. So, these systems are equivalent.

We'll return to finding solutions to systems in the next section by exploiting this idea of equivalent systems, but let's first pause to consider a slightly different problem.

In class, we looked at three systems of linear equations. The first system had exactly one solution, the second had infinitely many solutions, and the third had no solutions. Let's use geometry to investigate the number of solutions a system of linear equations can have.

2.2. Intersections of Lines and Planes

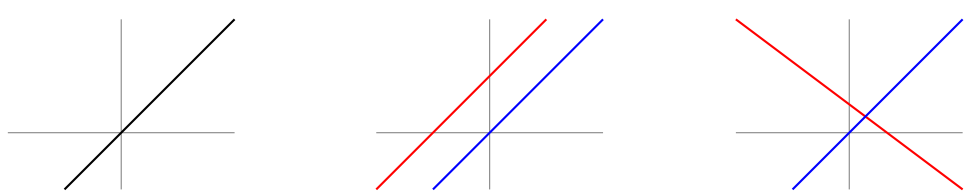
Note that a linear equation in two variables defines a line in $\mathbb{R}^2$, and a linear equation in three variables

$$ax + by + cz = d$$

defines a plane in $\mathbb{R}^3$. This leads to the following observation.

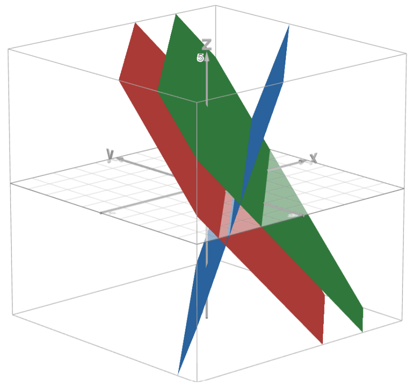
Observation 2.10. Solutions to systems of linear equations in two variables corresponds to intersection points of lines in $\mathbb{R}^2$. Similarly, solutions to systems of linear equations in three variables corresponds to intersection points of planes in $\mathbb{R}^3$.

Let's look first at the two dimensional case. Note that two lines in $\mathbb{R}^2$ can either be the same line, distinct and parallel, or distinct and not parallel

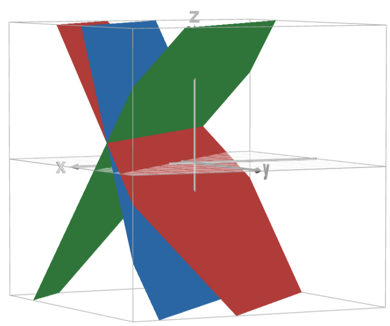


We see geometrically that these cases correspond to systems of linear equations which have infinitely many solutions, no solutions, or exactly one solution.

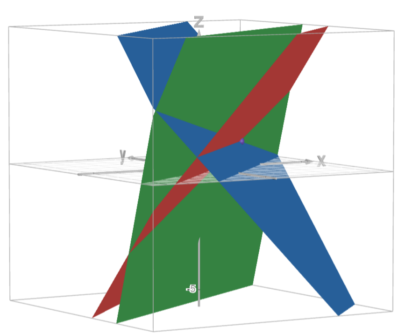
We can argue similarly for systems of linear equations in three variables. If all three of the corresponding planes are **distinct** then we have the following cases: at least two planes are parallel, in which case there are no intersection points, as in the example below



or no two of our planes are parallel, in which case our planes can intersect at a line



or at a point



Again we see geometrically that these cases correspond to systems of linear equations which have no solutions, infinitely many solutions, or exactly one solution.

Now that we've experimented both numerically and geometrically, I propose we make the following conjecture.

Conjecture 2.11. Any system of linear equations either has a unique solution, no solutions, or infinitely many solutions.

To determine the validity of this conjecture we'll return to our task of finding an algorithm to solve any system of linear equations. To do this, we first introduce a few notational tools that will simplify our problem.

2.3. The Matrix Representation of a Linear System

Consider the system of linear equations

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

Since our variables must all be the same in any system of linear equations, it's a bit redundant to write them every time. Instead, we could just write down the important pieces of this system into an array

$$\begin{array}{cccc} 1 & -1 & -1 & 1 \\ 2 & -3 & -1 & 3 \\ -1 & 1 & -1 & -3 \end{array}$$

Observe that the first three columns of this array correspond to the coefficients of our independent variables x, y and z for each of our three equations, and the last column corresponds to the constant on the right-hand side of our linear equations. This array is called a *matrix*, and can be used as a bookkeeping device to help us solve systems of linear equations. Let's give some formal definitions.

Definition 2.12. A MATRIX is any rectangular array of quantities or expressions. The quantities or expressions in a matrix are called its ENTRIES. If a matrix has n rows and m columns, then we call this an $n \times m$ matrix.

In this class, our matrices will typically contain real number entries (or sometimes variables working as placeholders for real number entries). Matrices are often written using soft brackets, such as in the 2×2 matrix below

$$\begin{pmatrix} 1 & 2 \\ 0 & -3 \end{pmatrix}$$

or by using hard brackets, such as in the 3×2 matrix below

$$\begin{bmatrix} 0 & \pi \\ 3 & 2 \\ 1 & 7 \end{bmatrix}.$$

We'll be using soft brackets throughout these notes, but either one is perfectly fine.

Definition 2.13. Consider a general system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_x + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

The MATRIX OF COEFFICIENTS corresponding to this system is the $m \times n$ matrix given by
$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$
The AUGMENTED MATRIX of this system is the $m \times (n+1)$ matrix given by
$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right).$$
Note that the augmented matrix accounts for the constants on the right-hand side of our equations, while the coefficient matrix does not. Also note that sometimes people do not draw the bar pictured above in an augmented matrix, this is just for clarity in notation.

Example 2.14. Consider the system of linear equations we saw in the previous section

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

The matrix of coefficients of this system is given by
$$\begin{pmatrix} 1 & -1 & -1 \\ 2 & -3 & -1 \\ -1 & 1 & -1 \end{pmatrix}$$
and the augmented matrix of this system is given by
$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 2 & -3 & -1 & 3 \\ -1 & 1 & -1 & -3 \end{array}\right)$$

Remark 2.15. Note that every matrix has an associated system of linear equations. For example, let
$$A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$
Then, A is the augmented matrix of the system
$$\begin{aligned} x + 2y &= 0 \\ 3x - y &= 1 \\ x + y &= 1 \end{aligned}$$
This correspondence between systems of linear equations and matrices is an important idea we'll revisit many times throughout the course.

2.4. Elementary Row Operations

So far, we've been solving systems of linear equation using ad-hoc methods. This is fine for small examples, but it doesn't scale well. For example, what if our system of linear equations had 100 variables? Ad hoc methods would be a disaster.

Let's instead develop an algorithm that guarantees we can always find a solution to any system of linear equations. To do this, we'll restrict ourselves to three "elementary operations". Observe that, given any system of linear equations, performing any of the following will yield an equivalent system of linear equations:

(E1) Interchange two equations;

(E2) Replace an equation by a nonzero multiple of itself;

(E3) Replace one equation by the sum of that equation and a scalar multiple of another equation;

Example 2.16. Consider the system of linear equations

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

Recall, when we solved this system before, our first step was to replace the third equation with the sum of the first and third equations. This is an application of (E3) above. Applying this operation gives the equivalent system

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

In the next section, we'll show that these three operations are all we need to solve any system. To do this, we'll make use of our new bookkeeping device: matrices.

Example 2.17. Observe that the system of linear equations from the previous example is represented by the augmented matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 2 & -3 & -1 & 3 \\ -1 & 1 & -1 & -3 \end{array}\right)$$

Furthermore, note that performing the elementary operation described above is the same as performing the following operation on the **rows** of the matrix above: replace row three with the sum of row three and row one. This gives a new matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 2 & -3 & -1 & 3 \\ 0 & 0 & -2 & -2 \end{array}\right)$$

which represents our equivalent system.

More generally, our three elementary operations (E1), (E2), and (E3) can be viewed as operations on the rows of the corresponding augmented matrix. We define the following.

Definition 2.18. The ELEMENTARY ROW OPERATIONS are defined as follows:

(ER1) Interchange two rows;

(ER2) Replace a row by a nonzero scalar multiple of itself.

(ER3) Replace one row by the sum of that row and a scalar multiple of another row;

We say that two matrices are ROW EQUIVALENT if one can be obtained from the other by a sequence of elementary row operations.

Noting again that applying an elementary row operation is the same as applying an operation to a system of linear equations, which yields an equivalent system of linear equations, we observe the following:

Theorem 2.19. *If two matrices are row equivalent, then the system of linear equations they represent have the same solution sets.*

We'll see in a couple of sections that the three operations above will be enough to solve any system of linear equations. To do that, let's first think about which systems of linear equations are "simplest" to solve.

2.5. Echelon Forms of a Matrix

Consider the following system of linear equations

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

Since we only have one variable in the last equation, we can solve $z = 1$. Now, since we only have two variables in the second equation, we can back substitute to find that $y - 1 = 1 \Rightarrow y = 2$. Finally, since we've solved for two of our three variables, we can substitute these into the first equation to find that

$$x + 2 + 1 = 1 \Rightarrow x = -2.$$

The method above is called *back substitution.* Observe that we were able to carry this method through because each of the equations in our system had (at least) one less variables than the equation above it.

Note that the system of equations from above has augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 2 & 2 \end{array}\right).$$

The fact that each of the equations in our system had at least one less variables than the equation above it caused the corresponding augmented matrix to form an "inverted staircase" pattern. Let's give some definitions to make sense of this pattern more carefully.

Definition 2.20. The PIVOT (also called the LEADING ENTRY) of a row in a matrix is the leftmost nonzero entry.

Example 2.21. The matrix

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

has pivot 1 in the first row, 2 in the second row, and 3 in the third row.

Definition 2.22. A matrix is said to be in ROW ECHELON FORM if

(1) all rows consisting only of zeros are at the bottom, and
(2) the pivot of each row in the matrix is in a column to the right of the pivot of the row above it.

Example 2.23. Consider the matrices

$$A = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 2 & 0 & 1 & 0 \\ 3 & 4 & -1 & 1/2 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; C = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{pmatrix},$$

$$D = \begin{pmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; E = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; F = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

Observe that matrices B, C, D and F are in row echelon form, while the matrices A and E are not.

Definition 2.24. A matrix is said to be in REDUCED ROW ECHELON FORM if the matrix is in echelon form and

(1) the pivot in each nonzero row is 1, and
(2) each pivot is the only nonzero entry in its column.

Example 2.25. The matrices B and F are in reduced row echelon form.

Let's look at a few examples to see how we can solve systems of linear equations represented by a matrix in reduced row echelon form.

Example 2.26. Find all solutions to the system of linear equations with corresponding augmented matrix

$$\left(\begin{array}{ccc|c} 2 & 0 & 0 & 4 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1/2 & 2 \end{array}\right).$$

Answer: $\boxed{(2, -1, 4)}$

Remark 2.27. Note that augmented matrices are the same as matrices with some extra decoration: the bar on the right-hand side reminds us that the last column is our column of coefficients. The matrix above is precisely the same matrix as

$$\begin{pmatrix} 2 & 0 & 0 & 4 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1/2 & 2 \end{pmatrix}$$

written without the extra decoration. Both matrices have three rows and four columns.

Example 2.28. Find all solutions to the system of linear equations with corresponding augmented matrix

$$\begin{pmatrix} 0 & 2 & -1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Answer: $\boxed{\text{No Solutions}}$.

Definition 2.29. A system of linear equations is called CONSISTENT if it has at least one solution. Otherwise, the system is called INCONSISTENT.

Example 2.30. Find all solutions to the system of linear equations with corresponding augmented matrix

$$\left(\begin{array}{ccc|c} 2 & 2 & 1 & 2 \\ 0 & 0 & 3 & 6 \end{array}\right).$$

Observe that this matrix represents the following system of linear equations

$$2x + 2y + z = 2$$

$$3z = 6.$$

We solve $z = 2$ from the second equation, and substitute this into the first equation to get

$$2x + 2y + 2 = 2 \Rightarrow x + y = 0.$$

What this means is that any tuple $(x, y, 2)$ satisfying $x + y = 0$ will work as a solution! For example,

$$(1, -1, 2), (0, 0, 2), \text{ and } (-1/5, 1/5, 2)$$

are all solutions to our system.

We can describe *all solutions* to our system as follows. Using our equation $x+y = 0$ we can solve for x in terms of y to get $x = -y$. Since y can be any real number, we'll just leave it "free" and describe all solutions in parametric form as

$$\boxed{(-y, y, 2) \text{ for any real number } y.}$$

for any real number y. Observe that our variable which ended up being free corresponded to the column in our matrix without a pivot.

The examples above demonstrate a key observation.

Observation 2.31. Any system of linear equations whose matrix is in (reduced) row echelon form can be solved using back substitution.

2.6. Gauss-Jordan Elimination

Observation 2.31 tells us that one way to solve any system of linear equations is to find an equivalent system whose augmented matrix is in reduced row echelon form. In this section, we show that this can in fact always be done.

Example 2.32. Find a matrix in reduced row echelon form which is row equivalent to the matrix below

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

Here's one method:

$$\begin{aligned}
\begin{pmatrix} 1 & 1 & 2 & 1 \\ 1 & 3 & 6 & 1 \\ 1 & 2 & 4 & 1 \end{pmatrix} &\sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 1 & 3 & 6 & 1 \\ 0 & 1 & 2 & 0 \end{pmatrix}, \ R_3 - R_1 \\
&\sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 2 & 4 & 0 \\ 0 & 1 & 2 & 0 \end{pmatrix}, \ R_2 - R_1 \\
&\sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \end{pmatrix}, \ R_2 - 2R_3 \\
&\sim \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ R_2 \leftrightarrow R_3 \\
&\sim \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \ R_1 - R_2
\end{aligned}$$

On your next homework assignment, you'll practice examples similar to the one above. With a bit more practice, you should convince yourself of the following important result.

Theorem 2.33 (Gauss-Jordan)**.** *Every matrix A is row equivalent to a unique matrix in reduced row echelon form X.*

We include the proof below for completeness. Interested students are encouraged to read through this argument, but we will not expect you to reproduce this argument.

Proof. We'll prove the existence of such a row equivalent matrix using "Gauss-Jordan elimination". This method is attributed to the two German mathematicians Carl Friedrich Gauss and Wilhelm Jordan due to their work in the 1800s, but was previously documented in a Chinese text dating back to around 150 BC. We'll save the uniqueness step of this argument for a later section once we've developed a bit more machinery. Suppose that A is an $n \times m$ matrix.

Step 1: Move all rows of zeros to the bottom. Suppose that A has a row of zeros in Row i where $i \neq n$. Letting j be the largest index so that Row j is not a row of zeros, interchange Rows i and j. Repeating this process for all rows gives a row equivalent matrix A_1 so that all rows consisting entirely of zeros are at the bottom of A_1.

Step 2: Move pivots. For all $i < j$ so that the pivot in R_j is to the left of the pivot in R_i in A_1, interchange R_i with R_j. This gives a row equivalent matrix A_2 so that all pivots are either to the right or below of the pivots in columns above it. Furthermore, A_2 still has all rows of zeros at the bottom of the matrix.

Next, for all i, j so that the pivot in R_i is in the same column as the pivot in R_j in A_2, perform the row operation $R_j - a_j/b_i R_i$, where a_j is the pivot in R_j and b_i is the pivot in R_i. This gives a row equivalent matrix A_3 in row echelon form.

Step 3: Make all pivots equal to 1. For all rows R_i containing a pivot a_i in A_3, replace R_i with $(1/a_i)R_i$. This gives a row equivalent matrix A_4 that is still in row echelon form, and where every pivot is equal to 1.

Step 4: Remove nonzero nonpivot entries from pivot columns. Finally, letting a_i be the pivot in row i, if there exists a nonzero entry b_i in row j and column i (the same column as a_i) replace R_j with $R_j - \frac{a_i}{b_i} R_i$. If necessary, repeat previous steps until the matrix is in reduced row echelon form. □

The uniqueness of the matrix X in the previous theorem lets us define the following.

Definition 2.34. Let A be a matrix. If A is row equivalent to a matrix X in reduced row echelon form, we call X THE REDUCED ROW ECHELON FORM OF A.

Theorem 2.33 will allow us to resolve our guiding questions from this Chapter. Let's introduce a few more definitions.

Definition 2.35. Let A be a matrix with reduced row echelon form equal to X.

(1) We say the ith column of A is a PIVOT COLUMN if X has a pivot in column i.
(2) We say that x_i is a BASIC VARIABLE of the system of linear equations represented by A if the ith column of A is a pivot column.
(3) We say that x_i is a FREE VARIABLE of the system of linear equations represented by A if the ith column of A is not a pivot column.

Example 2.36. Recall that the matrix

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

has reduced row echelon form

$$X = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

So, the pivot columns of A are columns 1 and 2. Since A represents a system of linear equations in three variables x_1, x_2, x_3, we get that x_1, x_2 are basic variables, while x_3 is free.

We have the following **very important result** (probably the most important result of the semester!), which we hope you gathered from our in-class activity:

Theorem 2.37. *Suppose that a system of linear equations has augmented matrix A. Then,*

(1) the system is inconsistent if and only if the last column of A is a pivot column;

(2) the system of equations has exactly one solution if and only if every column ***except for the last column*** *of A is a pivot column; and*

(3) the system has infinitely many solutions if and only if the last column is not a pivot column (i.e. the system is consistent) and some other column is also not a pivot column (i.e. the system has a free variable).

Proof. We'll prove one direction of each of these conditional statements as an example of how you would go about proving this in general. Note that we will not expect you to recreate this proof, it is included in these notes for interested students.

Observe that if A has a pivot in the last column, then A contains a row of the form

$$\left(\begin{array}{ccc|c} 0 & \cdots & 0 & d \end{array}\right)$$

for a nonzero real number d. This corresponds to the equation $0 = d$, a contradiction, and so this system has no solutions.

Next, suppose that A has a pivot in every column. Observe that A can have at most n pivots (the number of rows) and so when $n < m$ there cannot be a pivot in every column. So, either $n = m$ or $n > m$. If $n = m$ then A is of the form

$$A = \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & b_1 \\ 0 & 1 & \cdots & 0 & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & b_n \end{array}\right),$$

in which case the corresponding system has solution $(x_1, \ldots, x_n) = (b_1, \ldots, b_n)$. Define the $k \times k$ IDENTITY MATRIX I_k to be the matrix with 1's on the diagonal and 0's everywhere else. That is,

$$I_k = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

If $n > m$ observe that A is of the form

$$A = \left(\begin{array}{c|c} I_m & * \\ \hline 0 & 0 \end{array}\right)$$

which can be seen to correspond to a system with a unique solution similar to above.

Finally, suppose that A is consistent and contains a column with no pivot. Then, either the x_n variable is free, or there must be two rows in A of the form

$$\left(\begin{array}{cccccccc|c} 0 & \cdots & \boxed{1} & \cdots & \boldsymbol{*} & \cdots & * & \cdots & b_i \\ 0 & \cdots & 0 & \cdots & \mathbf{0} & \cdots & \boxed{1} & \cdots & b_{i+1} \end{array}\right).$$

Supposing the jth column has no pivot, we we see that our variable x_j can be taken to be free, which gives a system with infinitely many solutions, as needed. $\square$

Note that Conjecture 2.11 follows as a Corollary to Theorem 2.37, since A must either have a pivot in the last column or not. Furthermore, the proofs of Theorems 2.33 and 2.37 together give a complete algorithm to solve any system of linear equations.

Chapter 3

Linear Combinations and Spans

3.1. Solving Vector Equations

Recall, to solve the first magic carpet ride problem, we needed to find a solution to the vector equation

$$x\begin{pmatrix}3\\1\end{pmatrix}+y\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}107\\64\end{pmatrix}$$

which is equivalent to finding a solution to the following system of linear equations

$$3x+y=107$$

$$x+2y=64.$$

Observe that this system has corresponding augmented matrix

$$\left(\begin{array}{cc|c}3&1&107\\1&2&64\end{array}\right)$$

which has reduced row echelon form

$$\begin{pmatrix}1&0&30\\0&1&17\end{pmatrix}.$$

So, using what we learned in the previous chapter, we obtain solution

$$x=30, y=17.$$

The following result generalizes the method outlined above.

Theorem 3.1. *The vector equation*

$$x_1\vec{v}_1+x_2\vec{v}_2+\cdots+x_n\vec{v}_n=\vec{w}$$

has the same solution set as the linear system represented by the augmented matrix

$$\left(\begin{array}{cccc|c}\vec{v}_1&\vec{v}_2&\cdots&\vec{v}_m&\vec{w}\end{array}\right).$$

Example 3.2. Find all solutions to the following vector equation

$$x_1 \begin{pmatrix} 1 \\ 3 \\ 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 3 \\ 0 \\ 1 \\ 1 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ 2 \\ 5 \\ 3/2 \end{pmatrix} + x_4 \begin{pmatrix} 1/2 \\ 2 \\ 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \end{pmatrix}.$$

Answer: $\boxed{\left(\frac{17}{26}, \frac{139}{26}, \frac{131}{13}, -\frac{120}{13}\right)}$

Observe that, since the vector equation above had a solution, we know that the vector

$$\vec{w} = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \end{pmatrix}$$

is a linear combination of the vectors

$$\begin{pmatrix} 1 \\ 3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 5 \\ 3/2 \end{pmatrix}, \text{ and } \begin{pmatrix} 1/2 \\ 2 \\ 6 \\ 2 \end{pmatrix}.$$

In general, we have the following.

Theorem 3.3. *The vector $\vec{w}$ is a linear combination of vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ if and only if the system of linear equations represented by the augmented matrix*

$$\left(\begin{array}{cccc|c} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m & \vec{w} \end{array}\right)$$

is consistent.

3.2. Spans

In our second magic carpet ride problem, we were investigating which locations in $\mathbb{R}^2$ can be obtained as linear combinations of the vectors

$$\begin{pmatrix} 3 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Using either algebraic or geometric reasoning, we found that all vectors in $\mathbb{R}^2$ are linear combinations of the vectors above. Let's introduce some terminology to formalize questions of this type.

Definition 3.4. The SPAN of vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m$ in $\mathbb{R}^n$ is the set

$$\operatorname{Span}(\vec{v}_1, \dots, \vec{v}_m) = \{c_1\vec{v}_1 + \cdots + c_n\vec{v}_n \mid c_1, c_2, \dots, c_m \in \mathbb{R}\}.$$

That is $\operatorname{Span}(\vec{v}_1, \dots, \vec{v}_m)$ is the set of all linear combinations of vectors $\vec{v}_1, \dots, \vec{v}_m$.

So, using our work from Chapter 1, we can see that

$$\operatorname{Span}\left(\begin{pmatrix} 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) = \mathbb{R}^2.$$

Note that, the method we used to show that our hoverboard and magic carpet vectors span all of $\mathbb{R}^2$ doesn't generalize well to higher dimensions. Let's solve this problem again using a different technique

Example 3.5. Suppose that Gauss is located at $\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$. We're interested in showing that there always exists a solution x, y to the vector equation

$$x\begin{pmatrix} 3 \\ 1 \end{pmatrix} + y\begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Recall that solutions to the vector equation above are given by solutions to the system with augmented matrix

$$\left(\begin{array}{cc|c} 3 & 1 & g_1 \\ 1 & 2 & g_2 \end{array}\right).$$

Now, if we row reduce this matrix, we would get

$$\left(\begin{array}{cc|c} 1 & 0 & g_1' \\ 0 & 1 & g_2' \end{array}\right)$$

for some real numbers g_1', g_2'. Since there is never a pivot in the last column, by Theorem 2.37 this system must always have a solution.

Example 3.6. Let ℓ be any line through the origin. Recall from your reading that we can write ℓ in vector form as

$$\ell = \{t\vec{d} : t \in \mathbb{R}\}$$

for some direction vector $\vec{d}$. But this is precisely the definition of a span. That is, all points on ℓ are given by the set $\mathrm{Span}(\vec{d})$. So, all **lines through the origin are spans**.

Example 3.7. Let $\mathcal{P}$ be a plane passing through the origin. Recall from your reading that we can write

$$\mathcal{P} = \{t\vec{d}_1 + s\vec{d}_2 \mid t, s \in \mathbb{R}\}$$

for direction vectors $\vec{d}_1, \vec{d}_2$. Observe that this is precisely

$$\mathrm{Span}(\vec{d}_1, \vec{d}_2).$$

So, all **planes through the origin are also spans**.

In the next chapter, we'll see that spans generalize "flat spaces" like lines and planes. Furthermore, we'll develop tools to be able to quickly determine what type of "flat space" a span will give us.

3.3. Linear Dependence and Independence

In Core Problem 18, we saw that $\mathrm{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3)$ did not give us all of $\mathbb{R}^3$. One way that we could see this is to notice that $\vec{v}_3$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$, and so $\mathrm{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3) = \mathrm{Span}(\vec{v}_1, \vec{v}_2)$. Since two vectors at most span a plane, it's not possible for this span to reach all of $\mathbb{R}^3$.

The fact that $\mathrm{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3) = \mathrm{Span}(\vec{v}_1, \vec{v}_2)$ tells us that vector $\vec{v}_3$ was "redundant" information. That is, it was not necessary to include the vector $\vec{v}_3$ in our definition of the span, since it already was an element of that set. Finding a "minimal" subset of vectors that generates a span will be an important problem for us in this course. In this section, we develop some machinery to detect when we have redundant information in our definition of a span.

Definition 3.8 (Geometric definition of linear dependence)**.** We say that vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are LINEARLY DEPENDENT if for at least one i we have

$$\vec{v}_i \in \mathrm{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{i-1}, \vec{v}_{i+1}, \ldots, \vec{v}_n).$$

Otherwise, the vectors are called LINEARLY INDEPENDENT.

In Core Problem 19, we were able to detect whether small sets of vectors were linearly dependent or independent. When we have three or more independent vectors in our set, checking this by hand can become computationally challenging. For example, if we wanted to show that the vectors

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

are linearly independent, we could need to setup several vector equations.

In Core Exercises 20 and 21, you'll show that the following algebraic definition of linear independence is identical to the geometric definition above.

Definition 3.9 (Algebraic definition of linear dependence)**.** Vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ are LINEARLY DEPENDENT if there is a non-trivial linear combination of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ that equals the zero vector. Otherwise, the vectors are LINEARLY DEPENDENT.

Example 3.10. Let's use our new definition to show the vectors

$$\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}2\\1\end{pmatrix}$$

are linearly dependent. To see this, we need to consider solutions to the vector equation

$$x\begin{pmatrix}1\\1\end{pmatrix} + y\begin{pmatrix}1\\0\end{pmatrix} + z\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}. \tag{3.1}$$

Observe that the solutions to this equation are identical to the solutions to the system of linear equations with augmented matrix

$$\left(\begin{array}{ccc|c}1 & 1 & 2 & 0\\1 & 0 & 1 & 0\end{array}\right)$$

which has reduced row echelon form

$$\left(\begin{array}{ccc|c}1 & 1 & 2 & 0\\0 & 1 & 1 & 0\end{array}\right).$$

Note that this matrix has a column without a pivot, and so by Theorem 2.37 we know that there are infinitely many solutions to Equation 3.1. Namely, there has to be a solution other than $(0, 0, 0)$, and so these vectors are linearly dependent.

Example 3.11. Let's use our new definition to show that the vectors

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

are linearly independent. We can set this up similar to the previous problem. Since the matrix

$$\left(\begin{array}{ccc|c}1 & 3 & -1 & 0\\1 & 1 & 2 & 0\\0 & 2 & 0 & 0\end{array}\right)$$

has reduced row echelon form

$$\left(\begin{array}{ccc|c}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\end{array}\right)$$

which has a pivot in every column, except for the last one, then by Theorem 2.37 the vector equation

$$x\begin{pmatrix}1\\1\\0\end{pmatrix} + y\begin{pmatrix}3\\1\\2\end{pmatrix} + z\begin{pmatrix}-1\\2\\0\end{pmatrix} = \vec{0}$$

only has **one solution**, namely the trivial solution $x = y = z = 0$. So our vectors are linearly independent, by our algebraic definition of linear independence.

Remark 3.12. Observe that, since the right-hand side of the vector equation

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0} \tag{3.2}$$

is equal to zero, the last column in augmented matrix corresponding to this equation is going to always remain a column of zeros under any of our elementary row operations. So, it's enough to only consider the pivot columns of the matrix

$$\begin{pmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n\end{pmatrix}.$$

The following observations give us a quick method to detect linear dependence and independence.

(1) If the reduced row echelon form of the matrix $\begin{pmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n\end{pmatrix}$ has a column without a pivot, then by Theorem 2.37 we know that the system represented by this matrix has infinitely many solutions. Hence, equation (3.2) has a nontrivial solution, and so our vectors are linearly dependent.

(2) If every column of the matrix $\begin{pmatrix}\vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n\end{pmatrix}$ is a pivot column, then the vector equation (3.2) only has one solution, given by the trivial solution $c_1 = c_2 = \cdots = c_n = 0$. Hence, our vectors are linearly independent.

Example 3.13. Determine which of the following sets of vectors are linearly dependent and which are linearly independent. For those that are linearly dependent, find a nontrivial linear combination of the vectors that is equal to $\vec{0}$.

(1) $\begin{pmatrix}0\\1\\2\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\end{pmatrix}$

(2) $\begin{pmatrix}0\\1\\2\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}-2\\1\\2\\0\end{pmatrix}$

(3) $\begin{pmatrix}0\\1\\2\\1\end{pmatrix}, \begin{pmatrix}-1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\\0\end{pmatrix}$

(4) Any set of four vectors in $\mathbb{R}^3$ (give a geometric argument and an algebraic argument for your answer)

Solution for (1): We have

$$\begin{pmatrix}1 & -1 & 1\\1 & 1 & 1\\2 & 1 & 1\\1 & 0 & 1\end{pmatrix} \sim \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\\0 & 0 & 0\end{pmatrix},$$

and so the vector equation

$$x_1\begin{pmatrix}0\\1\\2\\1\end{pmatrix} + x_2\begin{pmatrix}-1\\1\\1\\0\end{pmatrix} + x_3\begin{pmatrix}1\\1\\1\\1\end{pmatrix} = \vec{0}$$

only has one solution $(0,0,0)$. Hence, using our algebraic definition, these vectors are $\boxed{\text{linearly independent}}$.

Solution for (2): We have

$$\begin{pmatrix}0 & -1 & 1 & -2\\1 & 1 & 1 & 1\\2 & 1 & 1 & 2\\1 & 0 & 1 & 0\end{pmatrix} \sim \begin{pmatrix}1 & 0 & 0 & 1\\0 & 1 & 0 & 1\\0 & 0 & 1 & -1\\0 & 0 & 0 & 0\end{pmatrix},$$

which has a column without a pivot, and so by Theorem 2.37 the vector equation

$$x_1\begin{pmatrix}0\\1\\2\\1\end{pmatrix} + x_2\begin{pmatrix}-1\\1\\1\\0\end{pmatrix} + x_3\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + x_4\begin{pmatrix}-2\\1\\2\\0\end{pmatrix} = \vec{0}$$

has infinitely many solutions. Namely, there must exist a solution other than the trivial one $(0,0,0,0)$. Hence, using our algebraic definition, these vectors are $\boxed{\text{linearly dependent}}$.

Solution for (3): We have

$$\begin{pmatrix} 0 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 2 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and so the vector equation

$$x_1 \begin{pmatrix} 0 \\ 1 \\ 2 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + x_4 \begin{pmatrix} 1 \\ -1 \\ 1 \\ 0 \end{pmatrix} = \vec{0}$$

only has one solution $(0, 0, 0, 0)$. Hence, using our algebraic definition, these vectors are $\boxed{\text{linearly independent}}$

Solution for (4): Suppose that we have four vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ in $\mathbb{R}^3$. Then, the vector equation

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 + x_4\vec{v}_4 = \vec{0}$$

will have the same solution set as the augmented matrix

$$\left(A \mid \vec{0}\right)$$

where A is a 3×4 matrix. Since each column of A can have *at most* one pivot, then the matrix A has no more than three pivots. Since A has four columns, it must have a column without a pivot. Since the system $\left(A \mid \vec{0}\right)$ cannot have a pivot in he last column (since it's a column of zeros), and there's another column without a pivot, then by Theorem 2.37 this system has infinitely many solutions. Hence, as we argued in part (2) above, we know that these vectors are $\boxed{\text{linearly dependent}}$.

3.4. The Matrix-Vector Form of a Linear System

Recall in Theorem 3.1 we saw that every vector equation can be represented by a system of linear equations and vice versa. We introduce one more way to represent systems of linear equations, which will be useful as we expand our perspectives throughout the course.

Definition 3.14. Let A be an $n \times m$ matrix and $\vec{x}$ a vector in $\mathbb{R}^m$. Write

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \text{ and } \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

The matrix-vector product $A\vec{x}$ is the vector in $\mathbb{R}^m$ defined by

$$A\vec{x} := x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix}.$$

Example 3.15. Let

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}, \vec{v} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \text{ and } \vec{w} = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}.$$

Then,

$$A\vec{v} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ -1 \end{pmatrix}.$$

Note that $A\vec{w}$ is undefined, since $\vec{w} \in \mathbb{R}^3$ but A only has two columns. In fact, the vector product $A\vec{u}$ is **only** defined for vectors $\vec{u}$ in $\mathbb{R}^2$.

Example 3.16. The vector equation

$$a \begin{pmatrix} 1 \\ 2 \end{pmatrix} + b \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

can be rewritten in matrix-vector form as

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

Observe that we now have multiple ways to represent and study our original problem of solving systems of linear equations. We should aim to become comfortable with all of the equivalent representations below, as we can learn something using each of these perspectives.

The following four representations each have identical sets of solutions.

Representation 1: Vector equation

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix} + \cdots + x_m \begin{pmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Why? Vector equations arose as our motivating problem for the course, the magic carpet ride problem. Vector equations are often helpful when representing a problem **geometrically**.

Representation 2: Systems of Linear Equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n. \end{aligned}$$

Why? This representation is familiar, and can help us understand our problem **algebraically**.

Representation 2: Augmented matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1m} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2m} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} & b_m \end{array}\right)$$

Why? This representation is convenient for computation. We have a concrete algorithm (Gauss-Jordan elimination) to reduce any augmented matrix to a convenient equivalent form.

Representation 4: Matrix-vector equation

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Why? In a future chapter, we'll see how matrices can be interpreted as "linear transformations". The representation above shows us that this matrix-vector operation sends a vector in $\mathbb{R}^m$ to another vector in $\mathbb{R}^m$. Keep this idea bookmarked for now.

In the following example, we demonstrate how to move between these different perspectives.

Example 3.17. Consider the matrix-vector equation

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

Observe that finding a vector solution to this matrix equation is equivalent to finding a solution x, y, z to the system of linear equations

$$x + 2y - z = 2$$

$$2x + y + z = 6.$$

This linear system has augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 2 \\ 2 & 1 & 1 & 6 \end{array}\right)$$

which is row equivalent to the matrix in reduced row echelon form

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 10/3 \\ 0 & 1 & -1 & -2/3 \end{array}\right).$$

So all solutions to our system of linear equations above can be written as

$$(x, y, z) = (10/3 - t, -2/3 + t, t).$$

Finding solutions to this system of linear equations is also equivalent to solving the vector equation

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

So, any solution (x, y, z) to our linear system should also satify the vector equation above. For example, if we choose the solution $(x, y, z) = (10/3, -2/3, 0)$ this shows that

$$\begin{pmatrix} 2 \\ 6 \end{pmatrix} = \frac{10}{3} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{2}{3} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

That is, the vector $\begin{pmatrix} 2 \\ 6 \end{pmatrix}$ is in the span of the vectors

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \text{ and } \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

The following result follows by connecting these various representations.

Theorem 3.18. *Let A be an $n \times m$ matrix. The following statements are equivalent.*

(1) *The matrix-vector equation $A\vec{x} = \vec{b}$ has a solution for every vector $\vec{b}$ in $\mathbb{R}^n$;*

(2) *The system of equations with augmented matrix*

$$\left(A \mid \vec{b} \right)$$

is consistent for any vector $\vec{b}$ in $\mathbb{R}^n$.

(3) *The reduced row echelon form of the matrix A has a pivot in each row;*

(4) *The span of the columns of A is equal to $\mathbb{R}^n$;*

Note: When we say the statements above are "equivalent", we mean that statement (i) is true if and only if statement (j) is true, for all combinations of $i, j \in \{1, 2, 3, 4\}$. More intuitively, we should think of this theorem as saying: if you want to show a matrix A satisfies any one of the properties itemized above, you can instead show that it satisfies any of the other properties itemized above. Let's look at a quick example before sketching this proof.

Example 3.19. Consider the matrix-vector equation

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

Since the matrix

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

is row equivalent to

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

and this matrix has a pivot in every row, we know that the matrix-vector equation

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

has a solution for any vector $\vec{b}$ in $\mathbb{R}^n$. We also know that

$$\text{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \end{pmatrix} \right) = \mathbb{R}^2$$

and that the system of linear equations

$$x + 2y - z = b_1$$

$$2x + y + 3z = b_2$$

has a solution for any $b_1, b_2 \in \mathbb{R}$.

Chapter 4

Vector Spaces

4.1. Subspaces of $\mathbb{R}^n$

In $\mathbb{R}^3$, we saw that the span of a set of vectors could either be

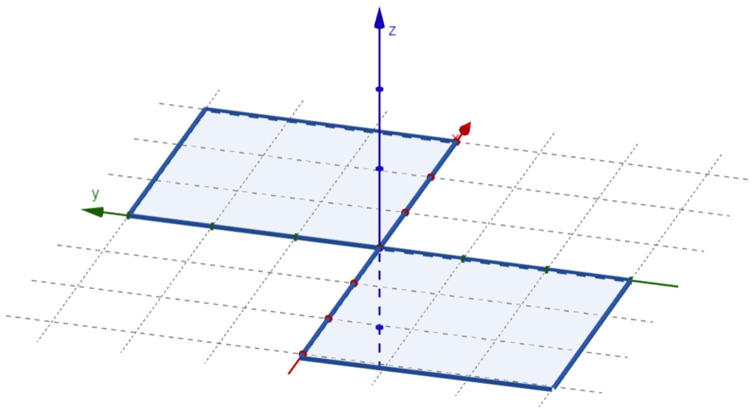
(1) a point (which looks like a "copy of $\mathbb{R}^0$")
(2) a line (which looks like a "copy of $\mathbb{R}^1$")
(3) a plane (which looks like a "copy of $\mathbb{R}^2$")
(4) all of $\mathbb{R}^3$

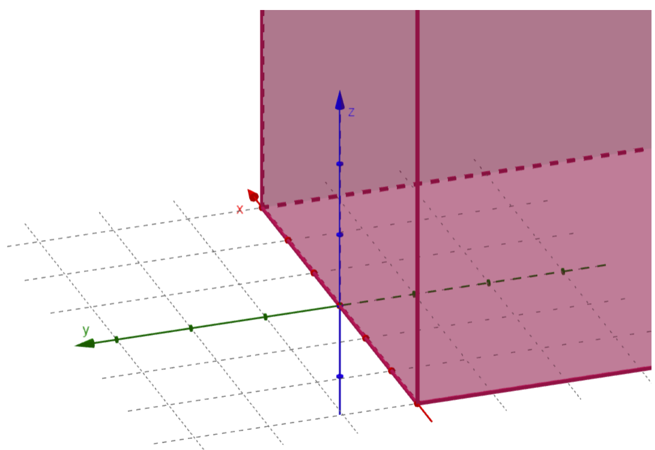
Furthermore, since $\vec{0}$ is an element of every span, each of these objects pass through the origin.

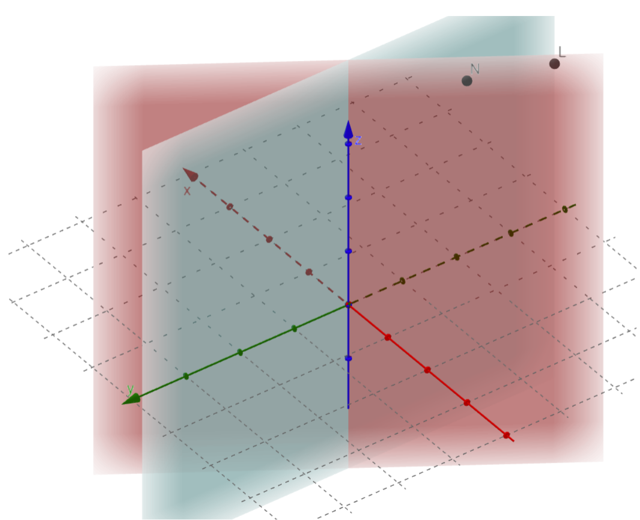
Our goal will be to extend the geometry we observed in $\mathbb{R}^3$ to general n-dimensional Euclidean space by capturing what we mean for a set to "look like a copy of $\mathbb{R}^m$. Oftentimes, mathematicians will classify a category of objects by considering what important properties these objects share (for example, characterizing integers by "even" or "odd" depending on whether they are divisible by 2). We'll take the same approach here.

Example 4.1. Consider the following subsets of $\mathbb{R}^3$. Explain why these sets **don't** look like a "copy of $\mathbb{R}^m$".

(1) Explain in normal language why these objects are not points, lines, planes, or all of $\mathbb{R}^3$.
(2) Using the language of vectors, what properties do lines and planes share that the objects below don't?


z
x
y


z
x
y


L
x
y

What we hope you observe in the previous example is that these sets don't look like copies of $\mathbb{R}^m$ because they're "missing" some pieces. A property we can observe these sets lack is the following: there exists vectors $\vec{v}$ and $\vec{w}$ so that $\vec{v}+\vec{w}$ is not in the set, and/or there exist a vector $\vec{v}$ so that some multiple $k\vec{v}$ is not in the set. This precisely captures the notion of "filling in" the missing pieces. We have the following definition.

Definition 4.2. A SUBSPACE V of $\mathbb{R}^n$ is any nonempty subset of $\mathbb{R}^n$ that satisfies both of the following properties

(1) V is closed under vector addition; that is for all $\vec{u}, \vec{v} \in V$ we have $\vec{u}+\vec{v} \in V$
(2) V is closed under scalar multiplication; that is, for all $\vec{u} \in V$ and $k \in \mathbb{R}$ we have $k\vec{u} \in V$

Observe that every vector subspace contains the zero vector; this follows because V is nonempty (so there exists some $\vec{v} \in V$) and is closed under scalars, so that $\vec{0} = 0\vec{v} \in V$.

Remark 4.3. A VECTOR SPACE can be defined more generally to be any set V satisfying the properties above, where we allow our scalars k to come from any "field" (which are algebraic objects analogous to $\mathbb{R}$). In this course, we'll stick with vector subspaces of $\mathbb{R}^n$. Note that, we will often refer to our subspaces as vector spaces, and remember that in this course they will always be subspaces of $\mathbb{R}^n$.

The following **very important** Theorem tells us that spans and subspaces of $\mathbb{R}^n$ are the same object.

Theorem 4.4. *A subset V is a vector subspace of $\mathbb{R}^n$ if and only if exists vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ so that*

$$V = \mathit{Span}\,(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)\,.$$

Proof. Suppose first the $V = \text{Span}\,(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$. Observe that $\vec{0} \in V$ since we can write

$$\vec{0} = 0\vec{v}_1 + \cdots + 0\vec{v}_m,$$

and so V is not empty. Furthermore, for $\vec{u}, \vec{v} \in V$ we can write

$$\vec{u} = c_1\vec{v}_1 + \cdots + c_m\vec{v}_m$$

$$\vec{v} = d_1\vec{v}_1 + \cdots + d_m\vec{v}_m$$

for scalars $c_i, d_i \in \mathbb{R}$ and so

$$\vec{u}+\vec{v} = (c_1+d_1)\vec{v}_1 + \cdots + (c_m+d_m)\vec{v}_m \in \text{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

Finally, for any $k \in \mathbb{R}$ we have

$$a\vec{u} = (ac_1)\vec{v}_1 + \cdots + (ac_m)\vec{v}_m \in \text{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

So, V is a subspace of $\mathbb{R}^n$.

Conversely, suppose that V is a vector subspace of $\mathbb{R}^n$. Observe that a subset of $\mathbb{R}^n$ contains between 0 and n linearly independent vectors. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$

be a linearly independent subset of V where m is maximal (note that such an m exists because $m \leq n$). Since V is a vector subspace, we know that

$$\mathrm{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m) \subseteq V.$$

For the opposite set inclusion, take any $\vec{v} \in V$. If $\vec{v} = \vec{v}_i \in B$ then certainly $\vec{v} \in \mathrm{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$ by writing

$$\vec{v} = 0\vec{v}_1 + \cdots + 1 \cdot \vec{v}_i + \cdots + 0\vec{v}_m.$$

So, suppose that $\vec{v} \notin \mathrm{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$. Since m is maximal, the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m, \vec{v}\}$ must be linearly **dependent**. So, there exist scalars $c_1, \ldots, c_{m+1} \in \mathbb{R}$ not all equal to zero so that

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m + c_{m+1}\vec{v} = \vec{0}.$$

Furthermore, $c_{m+1} \neq 0$ since the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ is linearly independent. So we have

$$\vec{v} = \frac{1}{c_{m+1}} \left(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_m\vec{v}_m\right) \in \mathrm{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m)$$

as needed. $\square$

4.2. Bases

Definition 4.5. Let V be a vector subspace of $\mathbb{R}^n$. A subset $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_m\}$ of $\mathbb{R}^n$ is called a BASIS if it's a linearly independent generating set. That is, the vectors $\vec{b}_1, \ldots, \vec{b}_m$ are linearly independent and we have $V = \mathrm{Span}(\vec{b}_1, \ldots, \vec{b}_m)$.

Example 4.6. Determine which of the following sets is a basis for $\mathbb{R}^3$

$$\mathcal{A} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right\}$$

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$$

$$\mathcal{C} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

$$\mathcal{D} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Solution: Note that $\mathcal{A}$ is not a linearly independent set, so this is not a basis for $\mathbb{R}^3$, and while $\mathcal{B}$ is linearly independent, it's not a generating set. We can see that both $\mathcal{C}$ and $\mathcal{D}$ are bases for $\mathbb{R}^3$. Note that in this example we're thinking of $\mathbb{R}^3$ as a vector subspace of itself.

Remark 4.7. In the example above, we saw that bases are not unique, however each set which was a basis had the same number of elements. This turns out to be true in general.

Theorem 4.8. *Let V be a vector subspace of $\mathbb{R}^n$. Then, the size of any basis for V is unique.*

Proof. Suppose that V is a vector subspace of $\mathbb{R}^n$ with bases

$$\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}, \text{ and } \mathcal{C} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}.$$

We need to show that $k = m$. To do this, we'll use a *proof by contradiction.*

By way of contradiction, suppose that $k \neq m$. Without loss of generality, suppose that $k < m$. For each $i \in \{1, \ldots, m\}$ write

$$\vec{u}_i = a_{i1}\vec{v}_1 + a_{i2}\vec{v}_2 + \cdots + a_{ik}\vec{v}_k. \tag{4.1}$$

Now, consider the vector equation

$$\vec{0} = x_1\vec{u}_1 + \cdots + x_m\vec{u}_m \tag{4.2}$$

with unknowns $x_1, \ldots, x_m$. Replacing each $\vec{u}_i$ with its representation in terms of the basis elements from B as in Equation (4.1) and collecting coefficients, we obtain

$$\begin{aligned}\vec{0} = \quad & (x_1a_{11} + x_2a_{21} + \cdots + x_ma_{m1})\vec{v}_1 \\ & + (x_1a_{12} + x_2a_{22} + \cdots + x_ma_{m2})\vec{v}_2 \\ & \vdots \\ & + (x_1a_{1k} + x_2a_{2k} + \cdots + x_ma_{mk})\vec{v}_k.\end{aligned}$$

Now, since the $\vec{v}_i$ are linearly independent, each of the coefficients above must be equal to zero. This yields the system of linear equations

$$\begin{gathered} x_1a_{11} + x_2a_{21} + \cdots + x_ma_{m1} = 0 \\ x_1a_{12} + x_2a_{22} + \cdots + x_ma_{m2} = 0 \\ \vdots \\ x_1a_{k1} + x_2a_{k2} + \cdots + x_ma_{mk} = 0, \end{gathered}$$

which is equivalent to the matrix-vector equation

$$A\vec{x} = \vec{0}$$

where

$$A = \begin{pmatrix} a_{ji} \end{pmatrix} \text{ and } \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

Since A is a $k \times m$ matrix and $k < m$, there are more columns than rows, which means we have a column with no pivot. Since a homogeneous system is always consistent, then Theorem 1.19 tells us that this system has infinitely many solutions. Namely, we have a nontrivial solution to (4.2), which means our set $\mathcal{C} = \{\vec{u}_1, \ldots, \vec{u}_m\}$ is linearly dependent, a contradiction. Hence, it must have been that $k = m$. $\square$

This gives rise to the following definition.

Definition 4.9. Let V be a vector subspace of $\mathbb{R}^n$. Then, the DIMENSION of V, denoted $\dim V$, is equal to the size of any basis for V. We define the dimension of the trivial subspace $\{\vec{0}\}$ to be 0.

Example 4.10. Let $V = \operatorname{Span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4)$ be a vector subspace of $\mathbb{R}^n$ and let

$$A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 \end{pmatrix}.$$

Say that we know the reduced row echelon form of A has a pivot in columns $1, 3$ and 4 and no pivot in column 2. Observe that we have the following:

(1) Show that $\vec{v}_2 \in \operatorname{Span}(\vec{v}_1, \vec{v}_3, \vec{v}_4)$

Solution: To see this, consider the vector equation

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_+ x_4\vec{v}_4 = \vec{0}.$$

Since we know that A doesn't have a pivot in the second column, the variable x_2 is free. That is, we can find a solution to the vector equation above *for any* real number x_2. So, there's a solution $(c_1, 1, c_3, c_4)$ to the vector equation above (where c_i are some real numbers, which we could find but we don't need to so let's be lazy). This gives

$$\vec{v}_2 = -c_1\vec{v}_1 - c_3\vec{v}_3 - c_4\vec{v}_4 \in \operatorname{Span}(\vec{v}_1, \vec{v}_3, \vec{v}_4).$$

(2) Explain how we know that $\{\vec{v}_1, \vec{v}_3, \vec{v}_4\}$ is a linearly independent set

Solution: Since A has pivots in columns 1, 3 and 4, the matrix $\begin{pmatrix} \vec{v}_1 & \vec{v}_3 & \vec{v}_4 \end{pmatrix}$ has a pivot in every column.

So, we have that $\vec{v}_1, \vec{v}_3, \vec{v}_4$ is a linearly independent generating set for V, which means it's a basis!

Example 4.11. Find a basis and the dimension of the following vector subspaces of $\mathbb{R}^3$.

(1) $V = \operatorname{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right)$

(2) $W = \operatorname{Span}\left(\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \\ -2 \end{pmatrix} \right)$

Solution: We can take

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

as a basis for V, and

$$\left\{ \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix} \right\}$$

as a basis for W

The following result generalizes our ideas from the previous examples.

Lemma 4.12. *Let A be an $n \times m$ matrix of the form*

$$A = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \end{pmatrix}$$

where the $\vec{v}_i$ are vectors in $\mathbb{R}^n$, and suppose that the reduced row echelon form of A is given by the matrix

$$X = \begin{pmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_m \end{pmatrix}.$$

If the column $\vec{x}_m$ of X does not have a pivot, then

$$\mathrm{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m) = \mathrm{Span}(\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{m-1}).$$

In general, we can remove any column of A that is not a pivot column and not change the span of its column vectors.

Proof. Consider the vector equation

$$c_1\vec{v}_1 + \cdots + c_m\vec{v}_m = \vec{0}.$$

Since there is no pivot in the mth column of X, we know that the variable x_m is free. That is, we can set x_m equal to *any real number* and obtain a solution to the vector equation above. Setting $c_m = -1$ gives

$$\vec{v}_m = c_1\vec{v}_1 + \cdots + c_{m-1}\vec{v}_{m-1} \in \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_{m-1}),$$

as needed. $\square$

By repeated use of Lemma 4.12 we obtain the following.

Theorem 4.13 (Finding Bases)**.** *Let V be the vector subspace of $\mathbb{R}^n$ given by*

$$V = \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

If A is the matrix with column vectors $\vec{v}_1, \ldots, \vec{v}_m$ then the pivot columns of A will form a basis for V. Furthermore, if the reduced row echelon form of A has k pivots, then $\dim(V) = k$.

Observe that 1-dimensional vector spaces are all lines, which look like "copies of $\mathbb{R}^1$. Similarly, 2-dimensional vector spaces are planes, which look like "copies of $\mathbb{R}^2$. This turns out to generalize – later on in the course we'll replace the words "copies of" with "is isomorphic to" to give a more precise meaning of this generalization.

4.3. Coordinate Systems

While the number of elements in a basis is fixed, we've seen that there are many (in fact, infinitely many) choices for a basis of a given vector space. In this section, we look at how our choice of basis impacts our geometric understanding of a given vector space.

In Core Problem 41, we saw that the two bases $\{\vec{e}_1, \vec{e}_2\}$ and $\{\vec{d}_1, \vec{d}_2\}$ gave us two different grids for $\mathbb{R}^2$, which in turn allowed us to give different directions to arrive at the same location. Let's introduce some terminology for this situation.

Definition 4.14. Let $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis for a vector space V. Recall that every vector $\vec{x}$ in V can be written in the form

$$\vec{x} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n.$$

The COORDINATES of $\vec{x}$ with respect to the basis $\mathcal{B}$ is given by

$$[\vec{x}]_{\mathcal{B}} := \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Conversely, we write

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_{\mathcal{B}} = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n.$$

Example 4.15. Note that $B = \{\vec{b}_1, \vec{b}_2\}$ is a basis for the vector space $\mathbb{R}^2$, where

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

Let's find the coordinates of $\vec{v} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ with respect to $\mathcal{B}$. To do this and we need to find coefficients x_1, x_2 in $\mathbb{R}$ so that

$$\vec{v} = x_1\vec{b}_1 + x_2\vec{b}_2.$$

This vector equation has augmented matrix

$$\left(\begin{array}{cc|c} 1 & 2 & 2 \\ 1 & 0 & 3 \end{array}\right)$$

which can be row reduced as

$$\left(\begin{array}{cc|c} 1 & 0 & 3 \\ 0 & 1 & -1/2 \end{array}\right)$$

and so we have

$$\vec{v} = 3\vec{b}_1 - \frac{1}{2}\vec{b}_2.$$

Using our new notation, we can write

$$[\vec{v}]_{\mathcal{B}} = \begin{pmatrix} 3 \\ -1/2 \end{pmatrix}.$$

Remark 4.16. Note that, in the previous example there was precisely one way to write $\vec{v}$ as a linear combination of $\vec{b}_1$ and $\vec{b}_2$, since the vector equation represented by

$$\left(\begin{array}{cc|c} \vec{b}_1 & \vec{b}_2 & \vec{v} \end{array}\right)$$

had exactly one solution. This turns out to be true in general (and is what allows us to define **the** coordinates of a vector with respect to a basis). We have the following.

Theorem 4.17. *Let V be a vector subspace of $\mathbb{R}^n$ and $\mathcal{B}$ a basis for V. Then, every vector in V has a unique representation in terms of the basis $\mathcal{B}$. That is, if $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_m\}$, then for every $\vec{v} \in V$ there are unique real numbers $x_1, \ldots, x_m$ so that*

$$\vec{v} = x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_m\vec{b}_m$$

Proof. Let $\vec{v} \in V$. Since $\{\vec{b}_1, \ldots, \vec{b}_m\}$ is a basis for V, this set is linearly independent, and so we know that reduced row echelon form of the matrix

$$\begin{pmatrix} \vec{b}_1 & \cdots & \vec{b}_m \end{pmatrix}$$

has a pivot in every column. Furthermore, since $V = \operatorname{Span}(\vec{b}_1, \ldots, \vec{b}_m)$ we know that the system

$$\left(\begin{array}{ccc|c} \vec{b}_1 & \cdots & \vec{b}_m & \vec{v} \end{array}\right)$$

must be consistent. Hence, the reduced row echelon form of the matrix above has a pivot in every column except for the last column, and so by Theorem 2.37 there is exactly one solution to the vector equation

$$\vec{v} = x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_m\vec{b}_m$$

as needed. □

Just for practice (and because I like the argument below), we give an alternate proof.

Proof of Theorem 4.17 (version 2). Let $\vec{v} \in V$ and suppose that we can write

$$\vec{v} = x_1\vec{b}_1 + \cdots + x_m\vec{b}_m$$

and

$$\vec{v} = y_1\vec{b}_1 + \cdots + y_m\vec{b}_m.$$

Subtracting these two equations gives

$$\vec{0} = (x_1 - y_1)\vec{v}_1 + \cdots + (x_m - y_m)\vec{v}_m.$$

Since our vectors $\vec{b}_1, \ldots, \vec{b}_m$ are linearly independent, we must have

$$x_i - y_i = 0 \Rightarrow x_i = y_i$$

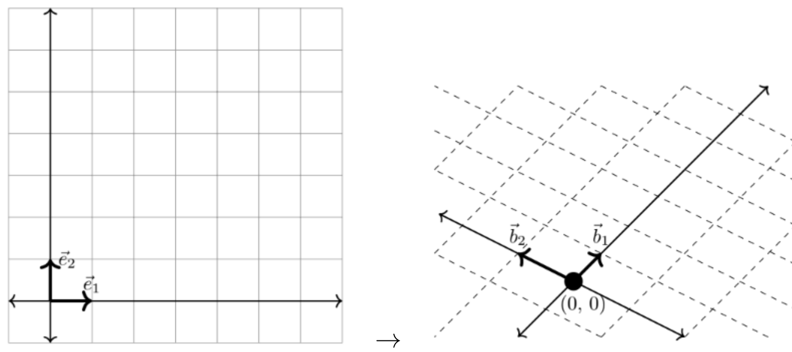
for all i. So, our two representations of $\vec{v}$ as a linear combination of the vectors $\vec{v}_1, \ldots, \vec{v}_m$ are the same. □

Remark 4.18. Observe that if $\mathcal{E} = \{\vec{e}_1, \ldots, \vec{e}_n\}$ is the standard basis for $\mathbb{R}^n$, then

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}_{\mathcal{E}} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

That is, when we talk about the coordinates of a vector (without referencing any specific basis), we really mean the coordinates of that vector *with respect to the standard basis.*

Here's a new perspective on what we've done in this section. We can think of the translation between different coordinate systems as a *transformation* (or function) from $\mathbb{R}^2$ to $\mathbb{R}^2$



Note that this transformation doesn't curve any of our lines (so lines stay lines), and the origin stays in place. In the next chapter, we'll study functions of this type more generally.

Chapter 5

Linear Transformations

5.1. Linearity

Definition 5.1. A function $F : \mathbb{R}^m \to \mathbb{R}^n$ is called LINEAR if it satisfies the following two properties:

(1) $F(\vec{x} + \vec{y}) = F(\vec{x}) + F(\vec{y})$, and
(2) $F(c\vec{x}) = cF(\vec{x})$

for any vectors $\vec{x}, \vec{y} \in \mathbb{R}^m$ and $c \in \mathbb{R}$. We'll often use the word *transformation* instead of the word *function* in this class. They mean exactly the same thing.

Remark 5.2. Note that our definition above assures that linear transformations leave the origin fixed. Indeed, if $\vec{x}$ is any vector in $\mathbb{R}^m$ then we have

$$F(\vec{0}) = F(\vec{x} - \vec{x}) = F(\vec{x}) - F(\vec{x}) = \vec{0}.$$

Example 5.3. Determine which of the following are linear transformations

(1) F: $\mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(\vec{x}) = 2\vec{x}$

Solution: We claim that F is linear. To see this, take any $\vec{x}, \vec{y} \in \mathbb{R}^2$. Then,

$$\begin{aligned} F(\vec{x} + \vec{y}) &= 2(\vec{x} + \vec{y}) \\ &= 2\vec{x} + 2\vec{y} \\ &= F(\vec{x}) + F(\vec{y}), \end{aligned}$$

and for any scalar $c \in \mathbb{R}$ we have

$$\begin{aligned} F(c\vec{x}) &= 2(c\vec{x}) \\ &= c(2\vec{x}) \\ &= cF(\vec{x}), \end{aligned}$$

as needed.

(2) $G : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$G\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^2 \\ y^2 \end{pmatrix}$$

Solution: This function is not linear. For example, we have

$$G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = G\left(\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 0 \end{pmatrix},$$

but

$$G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix},$$

and so

$$G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) \neq G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) + G\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right),$$

which shows that G is not linear.

(3) $H : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$H\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+y \\ x \\ 0 \end{pmatrix}$$

Solution: We claim that H is linear. To see this, take any

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \in \mathbb{R}^2.$$

Then,

$$\begin{aligned} H\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right) &= H\left(\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}\right) \\ &= \begin{pmatrix} (x_1 + x_2) + (y_1 + y_2) \\ x_1 + x_2 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} x_1 + y_1 \\ x_1 \\ 0 \end{pmatrix} + \begin{pmatrix} x_2 + y_2 \\ x_2 \\ 0 \end{pmatrix} \\ &= H\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) + H\left(\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}\right). \end{aligned}$$

Now, for any scalar $c \in \mathbb{R}$ we have

$$\begin{aligned} H\left(c\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) &= H\left(\begin{pmatrix} cx_1 \\ cy_1 \end{pmatrix}\right) \\ &= \begin{pmatrix} cx_1 + cy_1 \\ cx_1 \\ 0 \end{pmatrix} \\ &= c\begin{pmatrix} x_1 + y_1 \\ x_1 \\ 0 \end{pmatrix} \\ &= cH\left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right) \end{aligned}$$

as needed.

(4) $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+1 \\ y-1 \end{pmatrix}.$$

Solution: T is not linear. One way to see this is to note that $T(\vec{0}) \neq \vec{0}$ (we leave as an exercise to the reader that *any* linear function sends $\vec{0}$ to $\vec{0}$)

5.2. Matrix Transformations

Let A be an $n \times m$ matrix. Recall that the matrix-vector product $A\vec{x}$ is defined for a vector $\vec{x}$ in $\mathbb{R}^m$ and yields a vector $\vec{y} = A\vec{x}$ in $\mathbb{R}^n$. So, this product defines a function, which we'll give a special name and notation to.

Definition 5.4. Let A be an $n \times m$ matrix. Then, the MATRIX TRANSFORMATION associated to A is the function $T_A : \mathbb{R}^m \to \mathbb{R}^n$ defined by

$$T_A(\vec{x}) := A\vec{x}.$$

We have the following observation.

Proposition 5.5. Every matrix transformation is a linear transformation.

Proof. Let A be an $n \times m$ matrix, and write $A = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{pmatrix}$. Take any vectors

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \text{ and } \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

in $\mathbb{R}^m$. Then we have

$$\begin{aligned}
T_A(\vec{x}+\vec{y}) &= A(\vec{x}+\vec{y}) \\
&= A\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_m+y_m \end{pmatrix} \\
&= (x_1+y_1)\vec{v}_1 + (x_2+y_2)\vec{v}_2 + \cdots + (x_m+y_m)\vec{v}_m \\
&= (x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m) + (y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_m\vec{v}_m) \\
&= A\vec{x} + A\vec{y} \\
&= T_A(\vec{x}) + T_A(\vec{y}).
\end{aligned}$$

Now, take any $c \in \mathbb{R}$. Then we have

$$\begin{aligned}
T_A(c\vec{x}) &= A(c\vec{x}) \\
&= A\begin{pmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_m \end{pmatrix} \\
&= cx_1\vec{v}_1 + cx_2\vec{v}_2 + \cdots + cx_m\vec{v}_m \\
&= c(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_m\vec{v}_m) \\
&= c(A\vec{x}) \\
&= cT_A(\vec{x}),
\end{aligned}$$

as needed. $\square$

Example 5.6. Suppose that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation given by

$$F(\vec{e}_1) = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ and } F(\vec{e}_2) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

(1) Find $F\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$

(2) Find a formula for $F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right)$

(3) Find a 2×2 matrix A so that $F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = A\begin{pmatrix} x \\ y \end{pmatrix}$

What we did in the previous example generalizes.

Theorem 5.7. *Let $F : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Then, there exists a unique $n \times m$ matrix A so that $F(\vec{x}) = A\vec{x}$. That is, $F = T_A$ and so every linear transformation is a matrix transformation. Furthermore, we have*

$$A = \begin{pmatrix} F(\vec{e}_1) & F(\vec{e}_2) & \cdots & F(\vec{e}_m) \end{pmatrix}$$

where $\vec{e}_1, \dots, \vec{e}_m$ is the standard basis for $\mathbb{R}^m$.

Proof. Suppose that $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation. Define the $n \times m$ matrix A to be the matrix with column vectors equal to $F(\vec{e}_i)$ for $i = 1, \ldots, n$. We have

$$F(\vec{e}_i) = A\vec{e}_i \tag{5.1}$$

for every $i \in \{1, \ldots, m\}$. Now, take any $\vec{x} \in \mathbb{R}^m$. Since $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_m\}$ form a basis for $\mathbb{R}^m$ we can write $\vec{x}$ uniquely as

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_m\vec{e}_m.$$

So, we have

$$\begin{aligned}
F(\vec{x}) &= F(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_m\vec{e}_m) \\
&= x_1F(\vec{e}_1) + x_2F(\vec{e}_2) + \cdots + x_mF(\vec{e}_m) && \text{by linearity of } F \\
&= x_1A\vec{e}_1 + x_2A\vec{e}_2 + \cdots + x_mA\vec{e}_m && \text{by (5.1)} \\
&= A(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_m\vec{e}_m) && \text{by linearity of } T_A \\
&= T_A(\vec{x}).
\end{aligned}$$

So, we have $F(\vec{x}) = T_A(\vec{x})$ for every vector $\vec{x}$ in $\mathbb{R}^m$, which means $F = T_A$ as functions. □

Definition 5.8. Given a matrix A, we call T_A the MATRIX TRANSFORMATION corresponding to the matrix A. Given a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$, we call the $n \times m$ matrix A_T constructed in Theorem 5.7 the DEFINING MATRIX of the transformation T.

Example 5.9. To find the defining matrix of the linear function $F : \mathbb{R}^2 \to \mathbb{R}^2, \vec{x} \mapsto 2\vec{x}$ from Example 5.3, it suffices to find where this function maps the standard basis. We have

$$F(\vec{e}_1) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, F(\vec{e}_2) = \begin{pmatrix} 0 \\ 2 \end{pmatrix},$$

and so F has defining matrix

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Example 5.10. Find the defining matrix for the linear transformation $F : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates every vector $45°$ counterclockwise about the origin.

Solution: Observe that

$$F : \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}$$

and so F has defining matrix

$$A = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}.$$

When we know where a linear transformation sends the standard basis, finding the defining matrix can be done quickly using the method above. However, sometimes we might not know precisely where the standard basis is sent. We refer the reader to the example starting on page 112 of your textbook for problems of this type.

5.3. Function Composition and the Matrix Product

Given sets A, B, C, D and functions $f : A \to B$ and $g : B \to C$, recall that the composite function $f \circ g : A \to C$ is defined by

$$(g \circ f)(a) = g(f(a)).$$

Let's look at how function composition behaves on linear transformations. Given an $n \times k$ matrix A and an $\ell \times m$ matrix B, we have the associated linear transformations

$$T_A : \mathbb{R}^k \to \mathbb{R}^n, \vec{x} \mapsto A\vec{x}$$

$$T_B : \mathbb{R}^m \to \mathbb{R}^\ell, \vec{x} \mapsto B\vec{x}.$$

So, for the composition $T_A \circ T_B$ to be defined, we need $\mathbb{R}^k = \mathbb{R}^\ell$. That is, $k = \ell$. In this case, we have $T_A \circ T_B : \mathbb{R}^m \to \mathbb{R}^n$ defined by

$$(T_A \circ T_B)(\vec{x}) = A(B\vec{x}). \tag{5.2}$$

Observe that the composition of linear functions is linear (we'll show this in lecture together). So, by Theorem 5.7, there exists an $n \times m$ matrix C so that

$$T_A \circ T_B = T_C.$$

We'll call this matrix the product of A and B.

Definition 5.11. Let A be an $n \times k$ matrix and B a $k \times m$ matrix. Then, the PRODUCT of A and B is the $n \times m$ matrix C satisfying

$$T_A \circ T_B = T_C.$$

We write $C = AB$.

Let's develop a method to calculate the matrix product AB. Write

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

and suppose B has column vectors

$$\vec{b}_1, \vec{b}_2, \dots, \vec{b}_m.$$

Then we have

$$\begin{aligned} A(B\vec{x}) &= A\left(x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_m\vec{b}_m\right) \\ &= x_1A\vec{b}_1 + x_2A\vec{b}_2 + \cdots + x_mA\vec{b}_m \\ &= C\vec{x}, \end{aligned}$$

where $C = \begin{pmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_m \end{pmatrix}$. Since the defining matrix of a linear transformation is unique, we must have

$$\boxed{AB = \begin{pmatrix} A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_m \end{pmatrix}}.$$

Let's look at some examples.

Example 5.12. Let

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & -1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 2 \\ 0 & 3 \\ -1 & 1 \end{pmatrix}.$$

Since A is 2×3 and B is 3×2, the matrix product AB is a 2×2 matrix. We have

$$AB = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 \end{pmatrix}$$

where

$$\vec{a}_1 = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \text{ and } \vec{a}_2 = \begin{pmatrix} 1 & 0 & 3 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

So,

$$AB = \begin{pmatrix} -2 & 5 \\ 2 & 4 \end{pmatrix},$$

With practice, we can perform this computation a bit more quickly. Let's perform the steps above by just keeping track of how we're generating each entry:

$$\begin{pmatrix} \color{red}{1} & \color{red}{0} & \color{red}{3} \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \color{red}{1} & 2 \\ \color{red}{0} & 3 \\ \color{red}{-1} & 1 \end{pmatrix} = \begin{pmatrix} \color{red}{-2} & \\ & \end{pmatrix}$$

$$\begin{pmatrix} \color{red}{1} & \color{red}{0} & \color{red}{3} \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & \color{red}{2} \\ 0 & \color{red}{3} \\ -1 & \color{red}{1} \end{pmatrix} = \begin{pmatrix} -2 & \color{red}{5} \\ & \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 3 \\ \color{red}{1} & \color{red}{1} & \color{red}{-1} \end{pmatrix} \begin{pmatrix} \color{red}{1} & 2 \\ \color{red}{0} & 3 \\ \color{red}{-1} & 1 \end{pmatrix} = \begin{pmatrix} -2 & 5 \\ \color{red}{2} & \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 3 \\ \color{red}{1} & \color{red}{1} & \color{red}{-1} \end{pmatrix} \begin{pmatrix} 1 & \color{red}{2} \\ 0 & \color{red}{3} \\ -1 & \color{red}{1} \end{pmatrix} = \begin{pmatrix} -2 & 5 \\ 2 & \color{red}{4} \end{pmatrix}$$

Remark 5.13. Note that if A is an $m \times k$ matrix and B is $\ell \times n$ matrix, the matrix product AB is only defined when $k = \ell$. So, in the example above, the matrix BA also happens to be defined and gives a 3×3 matrix.

Let's practice a few more examples.

Example 5.14. Let

$$A = \begin{pmatrix} 1 & 0 & 1 & -1 \\ 0 & 2 & 3 & 1 \end{pmatrix}, B = \begin{pmatrix} 0 & -1 \\ 3 & 0 \\ 1 & 0 \\ -1 & 4 \end{pmatrix} \text{ and } C = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & -2 & 1 \end{pmatrix}.$$

Calculate all possible matrix products of the three matrices above.

5.4. The Kernel and Image

Recall, for a function $F : A \to B$, we call A the DOMAIN and B the CODOMAIN. We define two more sets that will help important for us to understand linear transformations.

Definition 5.15. Given a function $F : \mathbb{R}^m \to \mathbb{R}^n$, the KERNEL of F is the subset of $\mathbb{R}^m$ given by

$$\ker(F) := \{\vec{x} \in \mathbb{R}^m \mid F(\vec{x}) = \vec{0}\}.$$

The IMAGE (aka the RANGE) of F is the subset of $\mathbb{R}^n$ given by

$$\operatorname{im}(F) := \{\vec{y} \in \mathbb{R}^n \mid F(\vec{x}) = \vec{y} \text{ for some } \vec{x} \in \mathbb{R}^m\}.$$

Example 5.16. Consider the function

$$F : \mathbb{R}^3 \to \mathbb{R}^2, \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x - y \\ z \end{pmatrix}.$$

Observe that $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \in \ker(F)$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \operatorname{im}(F)$.

In Core Exercise 52, we showed the following.

Proposition 5.17. Let $F : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation with defining matrix A. Then, $\ker(F)$ and $\operatorname{im}(F)$ are vector subspaces of $\mathbb{R}^m$.

Since we know $\ker(F)$ and $\operatorname{im}(F)$ are subspaces, it makes sense to talk about their dimension. We have the following.

Definition 5.18. The RANK of a linear transformation F is the dimension of $\operatorname{im}(F)$. The NULLITY of a linear transformation F is the dimension of $\ker(F)$.

Example 5.19. Find the rank and nullity of the linear transformation from Example 5.16.

Solution. Observe that this function has defining matrix

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

To find $\operatorname{rank}(A)$, let's find a basis for $\operatorname{im}(T_A)$. Observe that we have

$$\begin{aligned} \operatorname{im}(T_A) &= \{\vec{y} \in \mathbb{R}^2 \mid A\vec{x} = \vec{y} \text{ for some } \vec{x} \in \mathbb{R}^3\} \\ &= \{A\vec{x} \mid \vec{x} \in \mathbb{R}^3\} \\ &= \left\{ x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mid x_1, x_2, x_3 \in \mathbb{R} \right\} \\ &= \operatorname{Span}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right). \end{aligned}$$

Since the pivot columns of A are in the first and third column, our work above tells us that we can take

$$\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$$

as a basis for $\mathrm{im}(T_A)$. Hence, $\mathrm{rank}(A) = \dim(\mathrm{im}(T_A)) = 2$.

Next, observe that

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \ker(F) \text{ if and only if } A\vec{x} = \vec{0}.$$

This vector equation has corresponding augmented matrix

$$\begin{pmatrix} 31 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

and so x_2 is free and so

$$\begin{aligned} \ker(F) &= \left\{ \begin{pmatrix} x_2 \\ x_2 \\ 0 \end{pmatrix} : x_2 \in \mathbb{R} \right\} \\ &= \left\{ x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} : x_2 \in \mathbb{R} \right\} \\ &= \mathrm{Span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right). \end{aligned}$$

Since this generating set only has one element, it must be a linearly independent set, and hence $\mathrm{nullity}(F) = \dim(\ker(F)) = 1$.

Chapter 6

Fundamental Subspaces and the Geometry of Systems

6.1. Three Fundamental Subspaces

Note: this was previously Section 5.5

Suppose that F is linear with defining matrix A. That is, $F(\vec{x}) = A\vec{x}$. We introduce new terminology to describe the same sets from the previous section.

Definition 6.1. Given an $n \times m$ matrix A, the NULL SPACE of A, denoted $\mathrm{Nul}(A)$, is equal to $\ker(T_A)$. That is,

$$\mathrm{Nul}(A) := \{\vec{x} \in \mathbb{R}^m \mid A\vec{x} = \vec{0}\}.$$

The NULLITY OF A, denoted $\mathrm{nullity}(A)$ is equal to the nullity of T_A. That is,

$$\mathrm{nullity}(A) := \dim(\mathrm{Nul}(A)).$$

Definition 6.2. Given an $n \times m$ matrix A, the COLUMN SPACE of A, denoted $\mathrm{Col}(A)$, is equal to $\mathrm{im}(T_A)$. That is, if $A = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_m \end{pmatrix}$, then $\mathrm{Col}(A)$ is the subspace of $\mathbb{R}^n$ defined by

$$\mathrm{Col}(A) := \mathrm{Span}(\vec{v}_1, \ldots, \vec{v}_m).$$

The RANK OF A, denoted $\mathrm{rank}(A)$, is equal to the rank of T_A. That is,

$$\mathrm{rank}(A) = \dim(\mathrm{Col}(A)).$$

Example 6.3. Let $A = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. From our work in Example 5.19 we have

$$\mathrm{rank}(A) = 2 \text{ and } \mathrm{nullity}(A) = 1.$$

Example 6.4. Let $A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 3 & 2 & -1 \\ 0 & -1 & -1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix}$. Find $\operatorname{rank}(A)$ and $\operatorname{nullity}(A)$.

(Solution coming soon!)

6.1.1. The Row Space and Matrix Transpose. There's one more matrix operation we'll introduce here, which we'll get some geometric intuition for in the last chapter of these notes. For now, let's just practice getting comfortable with these definitions.

Definition 6.5. Let A be the $m \times n$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Then, the TRANSPOSE of A is the $n \times m$ matrix $A^\top$ given by

$$A^\top = \begin{pmatrix} a_{11} & a_{11} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

That is, $A^\top$ is the matrix with column vectors equal to the rows of A.

Example 6.6. Let

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

Then

$$A^\top = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Definition 6.7. The ROW SPACE of an $m \times n$ matrix A is the vector subspace $\operatorname{Row}(A)$ of $\mathbb{R}^n$ given by

$$\operatorname{Row}(A) = \operatorname{Col}(A^\top).$$

That is, $\operatorname{Row}(A)$ is the subspace of $\mathbb{R}^n$ spanned by the row vectors of A.

We have the following observation.

Theorem 6.8. *Let A be an $m \times n$ matrix. Then,*

$$\dim(\operatorname{Col}(A)) = \dim(\operatorname{Row}(A)).$$

Proof. First, observe that if A is row equivalent to B, then $\operatorname{Row}(A) = \operatorname{Row}(B)$. To see this, note that two matrices being row equivalent means that the rows of one can be written as linear combinations of rows of the other (I'll leave this formal justification as an exercise for the reader). Now, suppose that X is the reduced row echelon form of A. By above, we have that $\dim(\operatorname{Row}(A)) = \dim(\operatorname{Row}(X))$. Observe, by definition of a matrix being in reduced row echelon form, that the dimension of Row(X) is precisely the number of nonzero rows of X (since these rows are linearly independent). The result follows because there is exactly one pivot in every nonzero row of X, and the number of pivots of X is the dimension of the column space of A. □

6.2. Rank-Nullity

Observe in our examples from the previous section that the rank was equal to the number of basic variables of A, and the nullity was the number of free variables. This turns out to be true in general.

Theorem 6.9. *Let A be an $n \times m$ matrix with r pivot columns. Then, $\operatorname{rank}(A) = r$ and $\operatorname{nullity}(A) = m - r$. That is, the rank of A is equal to the number of pivot columns of A, and the nullity of A is the number of non-pivot columns of A.*

Proof. Suppose that A has reduced row echelon form equal to X. If X contains r columns with pivots, we know that Col(A) has a basis with r elements by Theorem 4.13. Furthermore, since X has r pivots we know that X will have $m - r$ free variables. To make our notation simpler, let's assume the pivots are in the first r columns (note that the argument follows identically by an appropriate relabeling of the column vectors). So, if

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \in \operatorname{Nul}(A)$$

then we can write $x_1, \ldots, x_r$ in terms of the remaining $m-r$ free variables $x_{r+1}, \ldots, x_m$. That is, there are real numbers a_i with

$$\begin{aligned} x_1 &= a_{1,r+1}x_{r+1} + a_{1,r+2}x_{r+2} + \cdots + a_{1,m}x_m \\ x_2 &= a_{2,r+1}x_{r+1} + a_{2,r+2}x_{r+2} + \cdots + a_{2,m}x_m \\ &\vdots \\ x_r &= a_{r,r+1}x_{r+1} + a_{r,r+2}x_{r+2} + \cdots + a_{r,m}x_m. \end{aligned}$$

So, we have

$$\operatorname{Nul}(A) = \operatorname{Span}\left(\begin{pmatrix} a_{1,r+1} \\ a_{2,r+1} \\ \vdots \\ a_{r,r+1} \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} a_{1,r+2} \\ a_{2,r+2} \\ \vdots \\ a_{r,r+2} \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} a_{1,m} \\ a_{2,m} \\ \vdots \\ a_{r,m} \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} . \right).$$

Finally, observe that the $m - r$ vectors above are linearly independent, since the matrix with the vectors above as its columns has a pivot in every column. □

We have the following consequence to Theorem 6.9, often referred to as the RANK-NULLITY THEOREM. Observe that this gives a connection between the geometry of the column space and the geometry of the null space. This connection will be revisited in the next chapter as well.

Corollary 6.10 (The Rank-Nullity Theorem). Let A be an $n \times m$ matrix. Then,

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = m.$$

Proof. Suppose that the matrix A has r pivots. By theorem 4.13 we know that $\operatorname{rank}(A) = r$. So the result follows by Theorem 6.9. $\square$

6.2.1. Geometric Rank-Nullity.

Note that the remainder this section is optional, and just meant to help you build some geometric intuition.

Recall the Rank-Nullity Theorem (Corollary 6.10) tells us that if A is an $n \times m$ matrix then

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = m.$$

Suppose that $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation with defining matrix A and recall that

$$\ker(F) = \operatorname{Nul}(A) \text{ and } \operatorname{im}(F) = \operatorname{Col}(A)$$

and $\dim(\ker(F)) = \operatorname{nullity}(A)$ and $\dim(im(F)) = \operatorname{rank}(A)$. This gives the following restatement of rank-nullity.

Proposition 6.11 (Geometric Rank-Nullity). Let $F : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation with defining matrix A. Then,

$$m = \dim(\ker(F)) + \dim(\operatorname{im}(F)).$$

The following example will illustrate the geometric interpretation of rank-nullity. Note that you will not be tested on anything from the rest of this section, since it's a bit beyond the scope of this course. Also note that I'm not going to be careful with a new definition here, my hope is that this example helps you build some geometric intuition.

Example 6.12. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation with defining matrix

$$F = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Observe that $\ker(F) = \operatorname{Span}\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)$. This tells us that under the function F, everything on the line $x = y$ gets "identified" with $\vec{0}$. Let's see what else gets identified under this map. We have

$$\begin{aligned} F(\vec{x}) &= F(\vec{y}) \\ \Leftrightarrow F(\vec{x}) - F(\vec{y}) &= \vec{0} \\ \Leftrightarrow F(\vec{x} - \vec{y}) &= \vec{0} \\ \Leftrightarrow \vec{x} - \vec{y} &\in \ker(F). \end{aligned}$$

That is, the vectors $\vec{x}$ and $\vec{y}$ will be identified under F if and only if

$$\vec{x} = \vec{y} + (\text{something in the kernel}).$$

So, the line passing through $\vec{x}$ that's parallel to $\ker(F)$ all gets collapsed onto one single point. Once you perform all of these identifications, what you end up with is the image. Here's a clip from a 3blue1brown video for what this identification looks like.

In general, if $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear map, we get

$$\mathbb{R}^m / \ker(F) \cong \operatorname{im}(F)$$

where the thing on the left-hand side is called a "quotient", and we should think of it as the "smooshing down" we saw when we identified everything in the kernel (and then everything else in $\mathbb{R}^2$ that also needed to be identified). The equation above is an example of something called the "first isomorphism theorem" that you would see in an abstract algebra course.

To see what this has to do with the dimension, observe that when we identify vectors in the kernel, we "lose" the dimension of the kernel. That is,

$$\dim(\mathbb{R}^m / \ker(F)) = \dim(\mathbb{R}^m) - \dim(\ker(F)).$$

Since the quotient gives us something that looks like the image, we obtain the equality

$$\dim(\mathbb{R}^m) - \dim(\ker(F)) = \dim(\operatorname{im}(F)).$$

Noting that $\dim(\mathbb{R}^m) = m$ and rearranging the equation above precisely gives the geometric version of the rank-nullity theorem (Proposition 6.11).

6.3. Homogeneous Systems and the Geometry of Systems

Now that we've developed some machinery, we can return to our problem of describing the solution set to a system of linear equations. We define the following.

Definition 6.13. A system of linear equations is called HOMOGENEOUS if the constant coefficients are all equal to zero. For example, the system

$$2x - y = 0$$

$$x + 3y = 0$$

is homogeneous.

On your Webwork assignments, you saw several examples of solution sets to homogeneous systems that could be described as a span. From what we did in the previous sections, we can now prove that this is always the case.

Theorem 6.14. *The solution set to any homogeneous system of equations is a vector space.*

Proof. Observe that a homogeneous system can be represented in matrix-vector form as

$$A\vec{x} = \vec{0}.$$

So, the set of solutions to a homogeneous system is equal to $\operatorname{Nul}(A)$, which we know is a vector space by Core Exercise 52. □

Theorem 6.15. *The solution set to a system of n linear equations in m variables can be described by*

$$\vec{p} + V = \{\vec{p} + \vec{v} \mid \vec{v} \in V\}$$

where $\vec{p}$ is any particular vector solution to the system of linear equations and V is a vector subspace of $\mathbb{R}^m$. Furthermore, if the system of linear equation has matrix-vector form

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

then $V = \mathrm{Nul}(A)$.

Proof. Suppose that our system of linear equations has matrix-vector representation $A\vec{x} = \vec{b}$ for a vector $\vec{b} \in \mathbb{R}^n$, and suppose that $\vec{y}$ is any solution. Then,

$$\begin{aligned} A(\vec{y} - \vec{p}) &= A\vec{y} - A\vec{p} \\ &= \vec{b} - \vec{b} \\ &= \vec{0}. \end{aligned}$$

Hence, $\vec{y} - \vec{p}$ is a solution to $A\vec{x} = 0$ and so we can write $\vec{y} - \vec{p} = \vec{v}$ for some $\vec{v} \in V$. This gives $\vec{y} = \vec{v} + \vec{p}$ for $\vec{v} \in V$ as desired. Conversely, suppose that $\vec{v} \in V$. That is, $A\vec{v} = \vec{0}$. Then,

$$\begin{aligned} A(\vec{p} + \vec{v}) &= A\vec{p} + A\vec{v} \\ &= \vec{b} + \vec{0} \\ &= \vec{b}. \end{aligned}$$

So, $\vec{p} + \vec{v}$ is a solution to $A\vec{x} = \vec{b}$. □

Remark 6.16. We call the sets $\vec{p} + V$ TRANSLATED VECTOR SPACES (or TRANSLATED SPANS). Geometrically, these sets look like a vector space translated by some fixed vector $\vec{p}$ as in the examples pictured below

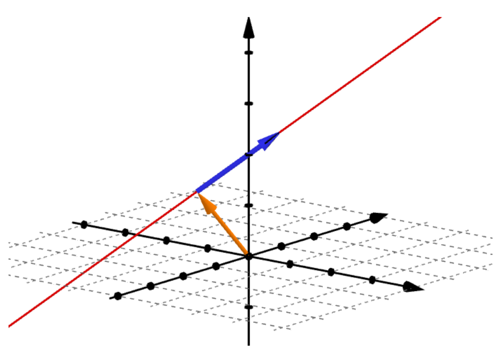


Figure 1. A translated one-dimensional vector space in $\mathbb{R}^3$ is a line

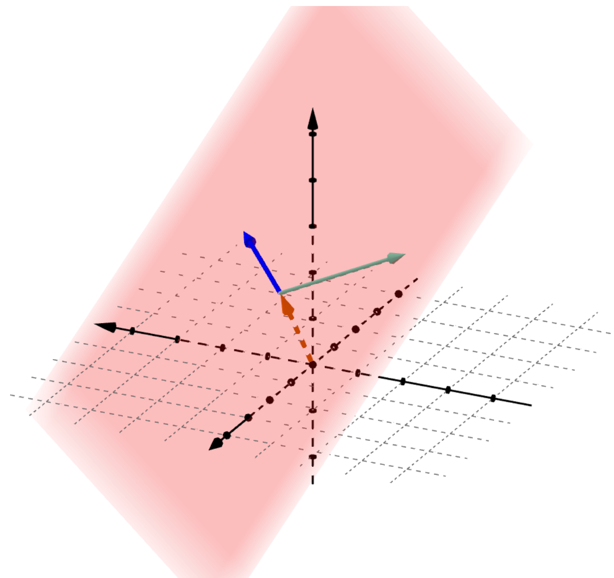


Figure 2. A translated two-dimensional vector space in $\mathbb{R}^3$ is a plane

Example 6.17. Write the solution set to the following system of linear equations as a translated vector space. Describe what this space looks like geometrically.

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

s

Chapter 7

Inverses

7.1. Inverse Functions

Recall that functions can be thought of as machines which send a set of inputs to a set of outputs. For some functions, it makes sense to "undo" the operations that sent our inputs to our outputs. For example, if f is the function which inputs a real number, to undo our operation we could divide the result by two. More formally, we could define

$$f : \mathbb{R} \to \mathbb{R}, f(x) = 2x \text{ and } g : \mathbb{R} \to \mathbb{R}, g(y) = y/2.$$

We see that the function g "undoes" the operations performed by f by looking at the composition

$$(g \circ f)(x) = g(f(x)) = g(2x) = 2x/2 = x.$$

Also note that g is in fact a function, since every element in the domain of g has a valid output, and furthermore that f also undoes g, since

$$(f \circ g) = f(g(x)) = f(x/2) = 2(x/2) = x.$$

Let's look at another example. Consider the function $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = x^2$. We might guess that a function which "undoes" h might look like $k : \mathbb{R} \to \mathbb{R}, k(x) = \sqrt{x}$. There are two issues with this. First, note that k is not actually a function since it's not defined everywhere on its domain (for example $k(-1) = \sqrt{-1}$ is not a valid real number). So maybe we alter our functions so that $h : \mathbb{R} \to \mathbb{R}_{\geq 0}$ and $k : \mathbb{R}_{\geq 0} \to \mathbb{R}$. But then we still run into trouble because for example

$$(k \circ h)(-4) = k(16) = 4$$

and so k didn't undo h like we wanted.

In this section, we review what it means for a function to have an inverse, and the properties needed to guarantee that an inverse exists.

Definition 7.1. Let X be any set. The *identity function* is the function $\mathrm{id} : X \to X$ defined by $\mathrm{id}(x) = x$ for every $x \in X$.

Definition 7.2. Let $f : X \to Y$ be a function. We say that f is INVERTIBLE if there exists a function $g : Y \to X$ so that $f \circ g = \mathrm{id}$ and $g \circ f = \mathrm{id}$. In this case, we call g an *inverse* of f.

Remark 7.3. Let's unpack the definition above. If we want g to be an inverse of f, g needs to itself be a function. That is, every input for g needs to have a unique corresponding output. What this means for f is that every **output** needs to have a unique corresponding **input**. We have the following.

We define the following.

Definition 7.4. Let $f : X \to Y$ be a function.

(1) F is called ONE-TO-ONE (or INJECTIVE) if the following property holds: for every $y \in Y$, there is *at most* one input $x \in X$ so that $f(x) = y$. We often use the arrow $f : X \hookrightarrow Y$ to indicate when a function is injective.
(2) F is called ONTO (or SURJECTIVE) if the following property holds: for every vector $\vec{b}$ in $\mathbb{R}^n$ there is *at least* one vector $\vec{x}$ in $\mathbb{R}^m$ so that $F(\vec{x}) = \vec{b}$. We often use the arrow $F : \mathbb{R}^m \twoheadrightarrow \mathbb{R}^n$ to indicate when a function is surjective.
(3) F is called BIJECTIVE if it is both one-to-one and onto.

Observation 7.5. A function is invertible if and only if it's bijective.

Determining when an arbitrary function is injective/surjective/bijective can be challenging in general. But we can exploit what we've learned about *linear* functions to simplify this question.

Example 7.6. Determine which of the following linear functions are injective, surjective and bijective.

(1) $F : \mathbb{R}^2 \to \mathbb{R}^3$, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$

Solution. We claim that F is injective but not surjective. To see this, observe that F has defining matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Take any vector $\vec{u} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{R}^3$ and consider the augmented matrix

$$\left(\begin{array}{cc|c} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & c \end{array}\right).$$

Observe that $\vec{u} \in \operatorname{im} F$ if and only if $c = 0$. This tells us that $\operatorname{im} F \neq \mathbb{R}^3$, since for example $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \notin \operatorname{im} F$ and so F is not surjective. Another way to see this is to note that $\operatorname{rank}(A) = 2$ and so $\operatorname{im}(F)$ is a two dimensional subspace of $\mathbb{R}^3$ and so $\operatorname{im}(F) \neq \mathbb{R}^3$. Next, since A has a pivot in every column, then by Theorem 2.37 the vector equation $A\vec{x} = \vec{u}$ has at most one solution, and so F is injective.

(2) $G : \mathbb{R}^3 \to \mathbb{R}^3$, $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x-y \\ y+z \\ x+z \end{pmatrix}$

Solution. We claim that G is neither surjective nor injective. To see this, observe that G has defining matrix

$$B = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

So, $\operatorname{rank}(B) = 2$ which means that $\operatorname{im}(G)$ is a two-dimensional subspace of $\mathbb{R}^3$ and so we must have $\operatorname{im}(G) \neq \mathbb{R}^3$. Hence, G is not surjective. Next, since B has a column without a pivot, then the matrix-vector equation $B\vec{x} = \vec{0}$ has infinitely many solutions. So, by Theorem 2.37 G is not one-to-one.

(3) $H : \mathbb{R}^3 \to \mathbb{R}^3$, $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x-y \\ y+z \\ z \end{pmatrix}$

Solution. We claim that H is bijective. To see this, observe that H has defining matrix

$$C = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Observe that $\operatorname{rank}(C) = \mathbb{R}^3$ and so $\operatorname{im}(H)$ is a 3-dimensional subspace of $\mathbb{R}^3$, which gives $\operatorname{im}(H) = \mathbb{R}^3$, and so H is surjective. Since C has a pivot in every column, then by Theorem 2.37 the matrix-vector equation $A\vec{x} = \vec{u}$ always has at most one solution, and so H is injective as well. Hence, H is bijective.

Theorem 7.7. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with defining matrix A. Then,*

(1) F is injective if and only if every column in the reduced row echelon form of A has a pivot.

(2) F is surjective if and only if every row in the reduced row echelon form of A has a pivot.

Proof. Take any $\vec{y} \in \mathbb{R}^m$ and consider the system

$$\left(A \mid \vec{y}\right).$$

By Theorem 2.37 this system has at most one solution if and only if A has a pivot in every column, as needed.

Next, by Theorem 6.9 we know that $\operatorname{rank}(A)$ is equal to the number of pivot columns of A. If A has a pivot in every row, then since A is an $m \times n$ matrix we have that $\operatorname{rank}(A) = m$ and so $\operatorname{im} F$ if an m-dimensional subspace of $\mathbb{R}^m$, which gives that $\operatorname{im} F = \mathbb{R}^m$ as needed. Now, if A has a row without a pivot, then $\operatorname{rank}(A) < m$

and so $\operatorname{im}(F)$ is a k-dimensional subspace of $\mathbb{R}^m$ with $k < m$, which means that $\operatorname{im}(F) \neq \mathbb{R}^m$, and so F is not surjective. $\square$

7.2. Isomorphisms

We have the following observation.

Proposition 7.8. Suppose that $F : \mathbb{R}^m \to \mathbb{R}^n$ is a linear transformation. If F is a bijection then we must have $m = n$.

Proof. Let A be the defining matrix for F. That is, let A be an $n \times m$ matrix so that $F(\vec{x}) = A\vec{x}$. Suppose that A has reduced row echelon form equal to X. By Theorem 7.7, X must have a pivot in every row and every column. If $n > m$ then there are at most m pivots (since X is in reduced row echelon form, each column can have at most one pivot). If $n < m$ than there could be at most n pivots (since each row can have at most one pivot), and so X would have a column without a pivot, which violates F being one-to-one. So, we must have $n = m$. $\square$

This proposition tells us that bijective linear functions can only map between "identical" Euclidean spaces because they must preserve the size and structure of the space.

This idea generalizes to bijective linear maps between vector spaces more generally. So far, we've been defining linear maps between Euclidean spaces $\mathbb{R}^m \to \mathbb{R}^n$. Note that we can also define a linear transformation between vectors subspaces $V \subseteq \mathbb{R}^m$ and $W \subseteq \mathbb{R}^n$.

Definition 7.9. Let V be a subspace of $\mathbb{R}^m$ and W a subspace of $\mathbb{R}^n$. An ISOMORPHISM between V and W is any linear bijective map $F : V \to W$. If an isomorphism exists between two vector spaces, we say these spaces are ISOMORPHIC and write $V \cong W$.

We have the following.

Theorem 7.10. *Let V and W be vector subspaces. Then $V \cong W$ if and only if* $\dim(V) = \dim(W)$.

Proof. Suppose that $F : V \to W$ is an isomorphism, and suppose that V has basis $\{\vec{v}_1, \ldots, \vec{v}_d\}$. Note that this implies $\dim(V) = d$. Let's show that the set $\{F(\vec{v}_1), \ldots, F(\vec{v}_d)\}$ is a basis for W. To see that this set is linearly independent, suppose that

$$x_1 F(\vec{v}_1) + \cdots + x_d F(\vec{v}_d) = \vec{0}.$$

Since F is linear, this gives

$$F(x_1 \vec{v}_1 + \cdots + x_d \vec{v}_d) = \vec{0}.$$

But since F is one-to-one and $F(\vec{0}) = \vec{0}$ we must have

$$x_1 \vec{v}_1 + \cdots + x_d \vec{v}_d = \vec{0}.$$

Since the $\vec{v}_i$ are linearly independent, we must have $x_1 = \cdots = x_d = 0$. Hence, $F(\vec{v}_1), \ldots, F(\vec{v}_d)$ are linearly independent as well. Next, to see that this set spans

W, take any $\vec{w} \in W$. Since F is onto, there exists $\vec{v} \in V$ so that $F(\vec{v}) = \vec{w}$. Since the $\vec{v}_i$ form a basis for V we can write $\vec{v} = c_1\vec{v}_1 + \cdots + c_d\vec{v}_d$ and so

$$\begin{aligned} \vec{w} &= F(c_1\vec{v}_1 + \cdots + c_d\vec{v}_d) \\ &= c_1F(\vec{v}_1) + \cdots + c_dF(\vec{v}_d) \\ &\in \operatorname{Span}(F(\vec{v}_1), \ldots, F(\vec{v}_d)), \end{aligned}$$

where the second equality follows by linearity of F. Hence, $W = \operatorname{Span}(F(\vec{v}_1), \ldots, F(\vec{v}_d))$ and so the $F(\vec{v}_i)$ form generating set. Hence, $\{F(\vec{v}_1), \ldots, F(\vec{v}_d)\}$ is a basis for W, and so $\dim(W) = d$.

Conversely, suppose that $\dim(V) = \dim(W)$. Let

$$\{\vec{v}_1, \ldots, \vec{v}_d\}$$

be a basis for V and

$$\{\vec{w}_1, \ldots, \vec{w}_d\}$$

be a basis for W. Define the map $F : V \to W$ by

$$F(x_1\vec{v}_1 + \cdots + x_d\vec{v}_d) = x_1\vec{w}_1 + \cdots + x_d\vec{w}_d.$$

(that is, $F : \vec{v}_i \mapsto \vec{w}_i$ and we extend linearly). It can be checked by definition that F is an isomorphism. $\square$

Example 7.11. Consider the spaces

$$V = \operatorname{Span}\left(\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\right) \text{ and } W = \operatorname{Span}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right).$$

Then $\dim(V) = \dim(W)$ and so $V \cong W$. Furthermore, both V and W are isomorphic to $\mathbb{R}^2$. In this case $V = \mathbb{R}^2$, and while W isn't **equal** to $\mathbb{R}^2$ on the nose (since it contains three-dimensional vectors) it "looks like" $\mathbb{R}^2$ (which is what isomorphisms are meant to capture).

7.3. Matrix Inverses

We have the following observation.

Proposition 7.12. If $F : \mathbb{R}^m \to \mathbb{R}^n$ is an invertible linear transformation, then its inverse $F^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ is also linear.

Proof. Take any $\vec{y}, \vec{z} \in \mathbb{R}^n$. Since F is surjective, there exists vectors $\vec{u}, \vec{v} \in \mathbb{R}^m$ so that $F(\vec{u}) = \vec{y}$ and $F(\vec{v}) = \vec{z}$. So, we have

$$\begin{aligned} F^{-1}(\vec{y} + \vec{z}) &= F^{-1}(F(\vec{u}) + F(\vec{v})) \\ &= F^{-1}(F(\vec{u} + \vec{v})), \text{ since } F \text{ is linear} \\ &= \vec{u} + \vec{v}, \text{ since } F^{-1} \circ F = \mathrm{id} \\ &= F^{-1}(\vec{y}) + F^{-1}(\vec{z}), \end{aligned}$$

where the final equality follows because $F(\vec{u}) = \vec{y} \Rightarrow F^{-1}(\vec{y}) = \vec{u}$ and $F(\vec{v}) = \vec{z} \Rightarrow F^{-1}(\vec{z}) = \vec{v}$. Similarly, for any $c \in \mathbb{R}$ we have

$$\begin{aligned} F^{-1}(c\vec{y}) &= F^{-1}(cF(\vec{u})) \\ &= F^{-1}(F(c\vec{u})) \\ &= c\vec{u} \\ &= cF^{-1}(\vec{y}), \end{aligned}$$

as needed. □

In this section, we derive a method to find the defining matrix of F^{-1}.

Definition 7.13. The IDENTITY MATRIX I_n is the defining matrix of the identity transformation $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n, \mathrm{id}(\vec{x}) = \vec{x}$.

Observe that I_n is the $n \times n$ matrix with 1s on the diagonal, and zeros everywhere else. For example,

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

Suppose that $F = T_A$ and $F^{-1} = T_B$. Then we have

$$F \circ F^{-1} = \mathrm{id} \Rightarrow AB = T_A \circ T_B = \mathrm{id} = I_n$$

and similarly

$$F^{-1} \circ F = \mathrm{id} \Rightarrow BA = I_n.$$

Let's use this observation to define inverses of matrices.

Definition 7.14. Let A be an $n \times n$ matrix. Then INVERSE MATRIX OF A, if it exists, is the $n \times n$ matrix B satisfying

$$AB = BA = I_n$$

If such a matrix B exists, we say that the matrix A is INVERTIBLE, and we write $B = A^{-1}$. As we discussed above, observe that

$$T_{A^{-1}} = T_A^{-1}.$$

Example 7.15. Let

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

Observe that if A^{-1} exists, it must satisfy

$$AA^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so A^{-1} must have column vectors $\vec{b}_1, \vec{b}_2, \vec{b}_3$ satisfying the matrix-vector equations

$$A\vec{b}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, A\vec{b}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \text{ and } A\vec{b}_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

We can use row reduction to solve the first matrix-vector equation, as below

$$\begin{aligned}
\left(\begin{array}{ccc|c}1&1&1&1\\0&2&1&0\\0&1&1&0\end{array}\right) &\sim \left(\begin{array}{ccc|c}1&1&1&1\\0&1&0&0\\0&1&1&0\end{array}\right), \text{ subtracting } R_3 \text{ from } R_2\\
&\sim \left(\begin{array}{ccc|c}1&0&0&1\\0&1&0&0\\0&1&1&0\end{array}\right), \text{ subtracting } R_3 \text{ from } R_1\\
&\sim \left(\begin{array}{ccc|c}1&0&0&1\\0&1&0&0\\0&0&1&0\end{array}\right), \text{ subtracting } R_2 \text{ from } R_3
\end{aligned}$$

This gives

$$\vec{b}_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}.$$

Observe that we can use exactly the same row operations to solve for $\vec{b}_2$ and $\vec{b}_3$:

$$\begin{aligned}
\left(\begin{array}{ccc|c}1&1&1&0\\0&2&1&1\\0&1&1&0\end{array}\right) &\sim \left(\begin{array}{ccc|c}1&1&1&0\\0&1&0&1\\0&1&1&0\end{array}\right), \text{ subtracting } R_3 \text{ from } R_2\\
&\sim \left(\begin{array}{ccc|c}1&0&0&0\\0&1&0&1\\0&1&1&0\end{array}\right), \text{ subtracting } R_3 \text{ from } R_1\\
&\sim \left(\begin{array}{ccc|c}1&0&0&0\\0&1&0&1\\0&0&1&-1\end{array}\right), \text{ subtracting } R_2 \text{ from } R_3
\end{aligned}$$

so that $\vec{b}_2 = \begin{pmatrix}0\\1\\-1\end{pmatrix}$ and similarly we compute

$$\begin{aligned}
\left(\begin{array}{ccc|c}1&1&1&0\\0&2&1&0\\0&1&1&1\end{array}\right) &\sim \left(\begin{array}{ccc|c}1&1&1&0\\0&1&0&-1\\0&1&1&1\end{array}\right), \text{ subtracting } R_3 \text{ from } R_2\\
&\sim \left(\begin{array}{ccc|c}1&0&0&-1\\0&1&0&-1\\0&1&1&1\end{array}\right), \text{ subtracting } R_3 \text{ from } R_1\\
&\sim \left(\begin{array}{ccc|c}1&0&0&-1\\0&1&0&-1\\0&0&1&2\end{array}\right), \text{ subtracting } R_2 \text{ from } R_3
\end{aligned}$$

so that $\vec{b}_3 = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$, which gives

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

Note that, since we performed the same row operations for every matrix-vector equation, we could performed the same operations as above by instead looking at the augmented matrix $\big(A \mid I_3\big)$, as follows

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_2$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{array}\right), \text{ subtracting } R_3 \text{ from } R_1$$

$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & -1 & 2 \end{array}\right), \text{ subtracting } R_2 \text{ from } R_3$$

and observe that the matrix on the right is A^{-1}. The following Theorem tells us that in fact this always works as a general strategy.

Theorem 7.16. *Let A be an $n \times n$ matrix. If $\big(A \mid I_n\big)$ is row equivalent to $\big(I_n \mid B\big)$ for an $n \times n$ matrix B, then then A is invertible with $A^{-1} = B$.*

Proof. Suppose that $\big(A \mid I_n\big)$ is row equivalent to $\big(I_n \mid B\big)$ for an $n \times n$ matrix B, and and write

$$B = \begin{pmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{pmatrix}$$

Let $\vec{u}_i$ be the vector in $\mathbb{R}^n$ with 1 in the ith component and 0s everywhere else. From above, we have that $\big(A \mid \vec{u}_i\big)$ is row equivalent to $\big(I_n \mid \vec{b}_i\big)$, and so $\vec{b}_i$ is a solution to the matrix-vector equation $A\vec{x} = \vec{u}_i$. This gives $AB = I_n$. Next, observe that $\big(A \mid I_n\big)$ being row equivalent to $\big(I_n \mid B\big)$ implies that $\big(B \mid I_n\big)$ is row equivalent to $\big(I_n \mid A\big)$ (this is not immediate, you may try convincing yourself of this with some examples). This proof will then be completed with the following lemma. □

Lemma 7.17. *For $n \times n$ matrices A, B, if $AB = I_n$ then $B = A^{-1}$.*

Note that this lemma tells us that we only need to check for "one-sided" inverses, which saves us time. We record the proof of this result below.

Proof. Observe that $\mathrm{Nul}(B) \subseteq \mathrm{Nul}(AB)$, since if $B\vec{x} = \vec{0}$ then $(AB)\vec{x} = A(B\vec{x}) = A\vec{0} = \vec{0}$. So, $\mathrm{nullity}(B) \leq \mathrm{nullity}(AB) = \mathrm{nullity}(I_n) = 0$. Hence, every column of $\mathrm{rref}(B)$ has a pivot. But since B is square this implies that every row of B must also have a pivot. So, B is invertible. This gives

$$A = AI_n = A(BB^{-1}) = (AB)B^{-1} = IB^{-1} = B^{-1}$$

and so $A = B^{-1}$. □

Note that we also have the following.

Proposition 7.18. If a matrix A is invertible, then its inverse A^{-1} is unique.

Proof. Suppose that B and C are two matrices satisfying

$$AB = BA = I_n$$

$$AC = CA = I_n.$$

This gives

$$B = BI_n = B(AC) = (BA)C = I_nC = C$$

and so $B = C$. □

7.4. The Invertible Matrix Theorem

In Theorem 7.16 we saw that if an $n \times n$ matrix A is invertible then $A \sim I_n$. The converse also turns out to be true. The main theorem in this section gives equivalent conditions to determine when an $n \times n$ matrix is invertible. Along the way, we'll develop another method to calculate matrix inverses. We first need a definition.

Definition 7.19. An $n \times n$ matrix is called ELEMENTARY if it can be obtained by performing exactly one row operation to the identity matrix.

Since we have three elementary row operations, there should be three types of elementary matrices. We have the following:

Row-switching matrices: let S_{ij} be the matrix which is obtained by swapping the ith and jth rows of the identity matrix. For example, for 3×3 matrices we have

$$S_{1,3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Observe that multiplying a matrix on the left by $S_{1,3}$ swaps the first and third row:

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{pmatrix}.$$

In general, $\boxed{\text{multiplying a matrix on the left by } S_{i,j} \text{ swaps the } i\text{th and } j\text{th rows}}$.

Row-multiplying matrix: Let $M_i(c)$ be the matrix which is obtained by multiplying the ith row of the identity matrix by a constant c. For example, for 3×3 matrices we have

$$M_2(5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Observe that multiplying a matrix on the left by $M_2(5)$ multiplies the 2nd row of that matrix by 5:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 5a_{21} & 5a_{22} & 5a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

In general, $\boxed{\text{multiplying a matrix on the left by } M_i(c) \text{ multiplies the } i\text{th row of that matrix by c}}$.

Row-addition matrix: Let $A_{i,j}(c)$ be the matrix with 1's on the diagonal, c in the (i, j) entry, and zeros everywhere else. That is, $A_{i,j}(c)$ is the matrix which is obtained by adding c times the jth row to the ith row of the identity matrix. For example, for 3×3 matrices we have

$$A_{1,2}(5) = \begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Observe that multiplying a matrix on the left by $A_{1,2}(5)$ adds 5 times the 2nd row to the first row:

$$\begin{pmatrix} 1 & 5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} + 5a_{21} & a_{12} + 5a_{22} & a_{13} + 5a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

In general, $\boxed{\text{multiplying a matrix on the left by } A_{i,j}(c) \text{ adds c times the } j\text{th row to the ith row}}$.

Observation 7.20. Note that every elementary matrix is invertible. Indeed, we have

$$S_{ij}^{-1} = S_{ij} = I_n$$
$$M_i(c)^{-1} = M_i(1/c)$$
$$A_{i,j}(c)^{-1} = A_{i,j}(-c)$$

We need a lemma before we prove one of our main results.

Lemma 7.21. *For $n \times n$ invertible matrices A, B we have*

$$(AB)^{-1} = B^{-1}A^{-1}$$

Proof. We have

$$(AB)(B^{-1}A^{-1}) = AI_nA^{-1} = AA^{-1} = I_n$$
$$(B^{-1}A^{-1})(AB) = B^{-1}I_nB = B^{-1}B = I_n$$

so $(AB)^{-1} = B^{-1}A^{-1}$. □

We have the following.

Theorem 7.22. *A matrix A is invertible if and only if it's a product of elementary matrices.*

Proof. Suppose first that A is invertible. By Theorem 7.16 we know that $A \sim I_n$. So, there is a series of elementary row operations which transform A to I_n. This is equivalent to the equality

$$I_n = E_k \cdots E_2 E_1 A$$

where E_i are elementary matrices. Since elementary matrices are invertible (as seen in Observation 7.20), then by repeated use of Lemma 7.21 we get

$$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$$

and so the result follows because the inverse of an elementary matrix is an elementary matrix, as seen in Observation 7.20. The converse follows by Observation 7.20 and noting that the product of invertible matrices is invertible. $\square$

The following result combines various perspectives we've collected so far this semester.

Theorem 7.23 (Invertible Matrix Theorem)**.** *Let A be an $n \times n$ matrix. The following are equivalent:*

(1) A is invertible;

(2) The reduced row echelon form of A is I_n;

(3) A is a product of elementary matrices.

(4) $\ker(T_A) = \{\vec{0}\}$

(5) T_A is injective

(6) The matrix-vector equation $A\vec{x} = \vec{0}$ only has the solution $\vec{x} = \vec{0}$;

(7) The system of linear equations with augmented matrix $\left(A \mid \vec{b}\right)$ has exactly one solution for any vector $\vec{b} \in \mathbb{R}^n$

(8) The columns of A are linearly independent;

(9) T_A is surjective;

(10) The row vectors of A are linearly independent;

(11) $\operatorname{rank}(A) = n$*;*

(12) The span of the column vectors of A is equal to $\mathbb{R}^n$;

(13) $\dim(\operatorname{Row}(A)) = \dim(\operatorname{Col}(A)) = n$*;*

(14) $\operatorname{im}(T_A) =^n$*;*

(15) $\operatorname{nullity}(A) = 0$*;*

(16) T_A is an isomorphism.

Activity: Add at least one equivalent condition to Theorem 7.22 that's not listed above (note: all conditions starting at (9) were generated by you!)

Chapter 8

Change of Basis and Matrix Similarity

8.1. Change of Basis Matrices

Example 8.1. Consider the basis $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ for $\mathbb{R}^2$ where

$$\vec{b}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \text{ and } \vec{b}_2 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

Let $[\vec{x}]_{\mathcal{E}} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$. Let's find $[\vec{x}]_{\mathcal{B}}$.

Note: recall that when we don't indicate a basis, we assume the coordinates of a vector are written with respect to the standard basis. For clarity in this section, we'll use the notation $[\vec{x}]_{\mathcal{E}}$ to emphasize when our coordinates are respect to the standard basis.

Recall that $[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ means that $\vec{x} = x_1\vec{b}_1 + x_2\vec{b}_2$. Since we also know that $\vec{x} = -\vec{e}_1 + 3\vec{e}_2$ we have

$$-\vec{e}_1 + 3\vec{e}_2 = x_1\vec{b}_1 + x_2\vec{b}_2.$$

Writing this equality in standard coordinates gives

$$\begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

But since the column vectors of $\begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}$ are the elements of the basis $\mathcal{B}$, they must be linearly independent. Hence, by the invertible matrix theorem, this matrix is invertible, and so we have

$$\begin{aligned}[\vec{x}]_{\mathcal{B}} &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} -1 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} -18 \\ 7 \end{pmatrix}.\end{aligned}$$

Example 8.2. Consider the basis $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$ for $\mathbb{R}^3$ where

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \text{ and } \vec{b}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Using a similar method to above, find $[\vec{x}]_{\mathcal{B}}$, where $[\vec{x}]_{\mathcal{E}} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$.

Observe that our examples above generalize to n-dimensions. We have the following.

Theorem 8.3. *Suppose that $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$ is a basis for $\mathbb{R}^n$ and define the matrix*

$$\big(\mathcal{B}\big) := \begin{pmatrix} [\vec{b}_1]_{\mathcal{E}} & \cdots & [\vec{b}_n]_{\mathcal{E}} \end{pmatrix}.$$

Then, the matrix $\big(\mathcal{B}\big)$ is invertible and we have

$$[\vec{x}]_{\mathcal{B}} = \big(\mathcal{B}\big)^{-1} [\vec{x}]_{\mathcal{E}}.$$

Proof. This proof will generalize what we did in the previous examples. Suppose that

$$[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Then we have

$$\vec{x} = x_1 \vec{b}_1 + \cdots + x_n \vec{b}_n.$$

Writing this equation in standard coordinates gives

$$[\vec{x}]_{\mathcal{E}} = \begin{pmatrix} [\vec{b}_1]_{\mathcal{E}} & \cdots & [\vec{b}_n]_{\mathcal{E}} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Since $\mathcal{B}$ is linearly independent, the matrix $\big(\mathcal{B}\big) = \begin{pmatrix} [\vec{b}_1]_{\mathcal{E}} & \cdots & [\vec{b}_1]_{\mathcal{E}} \end{pmatrix}$ is invertible, and so the result follows by multiplying the equation above by $\big(\mathcal{B}\big)^{-1}$. $\square$

Example 8.4. Let $\mathcal{A} = \{\vec{a}_1, \vec{a}_2\}$ be a basis for $\mathbb{R}^2$ where

$$\vec{a}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{a}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Suppose that we know

$$[\vec{x}]_{\mathcal{A}} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Letting $\mathcal{B}$ be the matrix from Example 8.1, let's find $[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Since we have $\vec{x} = x_1\vec{b}_1 + x_2\vec{b}_2$ and $\vec{x} = 2\vec{a}_1 + 3\vec{a}_2$ we get

$$x_1\vec{b}_1 + x_2\vec{b}_2 = 2\vec{a}_1 + 3\vec{a}_2.$$

Writing this equality in standard coordinates gives

$$\begin{pmatrix} \vec{b}_1 & \vec{b}_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \vec{a}_1 & \vec{a}_2 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

or using the notation above

$$(\mathcal{B}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (\mathcal{A}) \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

and so

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= (\mathcal{B})^{-1} (\mathcal{A}) \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 3 & -5 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 20 \\ -7 \end{pmatrix}. \end{aligned}$$

Example 8.5. Let $\mathcal{B}$ be the basis from Example 8.2 and let $\mathcal{A} = \{\vec{a}_1, \vec{a}_2, \vec{a}_3\}$ be the basis for $\mathbb{R}^3$ with

$$\vec{a}_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \vec{a}_2 = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \vec{a}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Using a similar method to above, find $[\vec{x}]_{\mathcal{B}}$, where $[\vec{x}]_{\mathcal{A}} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$.

Observe that our examples above generalize to n-dimensions. We have the following.

Theorem 8.6. *Suppose that $\mathcal{A}$ and $\mathcal{B}$ are bases for $\mathbb{R}^n$. Then,*

$$[\vec{x}]_{\mathcal{B}} = (\mathcal{B})^{-1} (\mathcal{A}) [\vec{x}]_{\mathcal{A}}.$$

Proof. This proof will follow similarly to our example above. Suppose that

$$[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

and that

$$[\vec{x}]_{\mathcal{A}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Then we have

$$\vec{x} = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n$$

and

$$\vec{x} = y_1\vec{a}_1 + \cdots + y_n\vec{a}_n.$$

This gives

$$x_1\vec{b}_1 + \cdots + x_n\vec{b}_n = \vec{a}_1 + \cdots + y_n\vec{a}_n$$

which can be written in standard coordinates as

$$\begin{pmatrix} [\vec{b}_1]_{\mathcal{E}} & \cdots & [\vec{b}_n]_{\mathcal{E}} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} [\vec{a}_1]_{\mathcal{E}} & \cdots & [\vec{a}_n]_{\mathcal{E}} \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

or using the notation above

$$\big(\mathcal{B}\big) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \big(\mathcal{A}\big) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

As before, we know that $\big(\mathcal{B}\big)$ is invertible and so we can multply the equality above by $\big(\mathcal{B}\big)^{-1}$ to obtain the desired result. □

Remark 8.7. In Theorem 8.6, we found a matrix M so that $[\vec{x}]_{\mathcal{B}} = M[\vec{x}]_{\mathcal{A}}$. That is, M is a matrix which translates a vector in $\mathcal{A}$-coordinates to a vector in $\mathcal{B}$-coordinates. We call this matrix the CHANGE OF BASIS MATRIX from $\mathcal{A}$ to $\mathcal{B}$. This is sometimes denoted $M = M_{\mathcal{B}\leftarrow\mathcal{A}}$ or as in your textbook by $M = [\mathcal{B} \leftarrow \mathcal{A}]$.

8.2. Matrix Similarity

Recall that for any linear transformation $F : \mathbb{R}^n \to \mathbb{R}^m$ there's a defining matrix A so that $F(\vec{x}) = A\vec{x}$. As in the last section, let's be careful again to keep track of the coordinates we're considering. Recall that if we set

$$A = \begin{pmatrix} F(\vec{e}_1) & \cdots & F(\vec{e}_n) \end{pmatrix}$$

then we have

$$[F(\vec{x})]_{\mathcal{E}} = A[\vec{x}]_{\mathcal{E}}.$$

Let's look at what happens if we instead consider our transformation with respect to a basis other than the standard one.

Suppose that $\mathcal{B}$ is any basis for $\mathbb{R}^n$. Let's prove the analogue of Theorem 5.7 with respect to a basis other than the standard basis.

Theorem 8.8. *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and $\mathcal{B}$ be any basis for $\mathbb{R}^n$. Then, there exists a unique $n \times n$ matrix M so that $[F(\vec{x})]_{\mathcal{B}} = M[\vec{x}]_{\mathcal{B}}$. Furthermore, we have*

$$M = \begin{pmatrix} [F(\vec{b}_1)]_{\mathcal{B}} & \cdots & [F(\vec{b}_n)]_{\mathcal{B}} \end{pmatrix}.$$

Proof. This proof will follow similarly to Theorem 5.7. Observe that $[\vec{b}_i]_{\mathcal{B}}$ is a vector in $\mathbb{R}^n$ with 1 in the i-th coordinate and zeros everywhere else. So,

$$\begin{pmatrix} [F(\vec{b}_1)]_{\mathcal{B}} & \cdots & [F(\vec{b}_n)]_{\mathcal{B}} \end{pmatrix} [\vec{b}_i]_{\mathcal{B}} = [F(\vec{b}_i)]_{\mathcal{B}}.$$

Hence, $[F(\vec{b}_i)]_{\mathcal{B}} = M[\vec{b}_i]_{\mathcal{B}}$ for every $i \in \{1, \dots, n\}$. Next, take any $\vec{x} \in \mathbb{R}^m$ and write

$$\vec{x} = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n.$$

Then,

$$\begin{aligned} [F(\vec{x})]_{\mathcal{B}} &= [F(x_1\vec{b}_1 + \cdots + x_n\vec{b}_n)]_{\mathcal{B}} \\ &= x_1[F(\vec{b}_1)]_{\mathcal{B}} + \cdots + x_n[F(\vec{b}_n)]_{\mathcal{B}} && \text{by linearity of } F \\ &= x_1 M[\vec{b}_1]_{\mathcal{B}} + \cdots + x_m M[\vec{b}_n]_{\mathcal{B}} && \text{by above} \\ &= M(x_1[\vec{b}_1]_{\mathcal{B}} + \cdots + x_n[\vec{b}_n]_{\mathcal{B}}) \\ &= M[\vec{x}]_{\mathcal{B}}, \end{aligned}$$

as needed. □

Definition 8.9. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation, and $\mathcal{B}$ be any basis for $\mathbb{R}^n$. Then, the DEFINING MATRIX OF F WITH RESPECT TO THE BASIS $\mathcal{B}$ is the matrix M so that

$$[F(\vec{x})]_{\mathcal{B}} = M[\vec{x}]_{\mathcal{B}}.$$

We use the notation $M = [F]_{\mathcal{B}}$.

Example 8.10. Let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2\}$ be the basis with

$$\vec{b}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \vec{b}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

Consider the linear transformation $F : \mathbb{R}^2 \to \mathbb{R}^2$ which stretches vectors in the $\vec{b}_1$ direction by 2 and leaves vectors in the $\vec{b}_2$ direction fixed. That is,

$$F(x_1\vec{b}_1 + x_2\vec{b}_2) = 2x_1\vec{b}_1 + x_2\vec{b}_2.$$

Find $[F]_{\mathcal{B}}$ and $[F]_{\mathcal{E}}$. (Observe that it's "easier" to view F in terms of the basis $\mathcal{B}$).

Note that the defining matrices we obtained in the previous example were different, but we don't want to lose track of the fact that these matrices are related to each other. We define the following.

Definition 8.11 (Geometric definition of matrix similarity)**.** Two $n \times n$ matrices C and D are called *similar* if they represent the same function, but in possibly different bases. That is, there is a single linear function $F : \mathbb{R}^n \to \mathbb{R}^n$ so that

$$[F]_{\mathcal{A}} = C \text{ and } [F]_{\mathcal{B}} = D,$$

where $\mathcal{A}$ and $\mathcal{B}$ are bases for $\mathbb{R}^n$.

Let's derive an algebraic method to detect matrix similarity using our results from the previous section. For simplicity, let's assume that $[F]_{\mathcal{E}} = D$. By definition, we have

$$[F(\vec{x})]_{\mathcal{B}} = C[\vec{x}]_{\mathcal{B}}.$$

Let's change bases to view this equality with respect to the standard basis. By Theorem 8.3 we get

$$\big(\mathcal{B}\big)^{-1}[F(\vec{x})]_{\mathcal{E}} = C\big(\mathcal{B}\big)^{-1}[\vec{x}]_{\mathcal{E}}.$$

Multiplying both sides of this equation by $\big(\mathcal{B}\big)$ gives

$$[F(\vec{x})]_{\mathcal{E}} = \big(\mathcal{B}\big)\,C\,\big(\mathcal{B}\big)^{-1}[\vec{x}]_{\mathcal{E}}.$$

Hence, we have $D = [F]_{\mathcal{E}} = \big(\mathcal{B}\big)\,C\,\big(\mathcal{B}\big)^{-1}$.

Definition 8.12 (Algebraic definition of matrix similarity)**.** Two $n \times n$ matrices C and D are called *similar* if there exists an invertible $n \times n$ matrix X so that

$$D = XCX^{-1}.$$

This definition will turn out to be useful for us later on when we investigate special bases that are useful for viewing linear transformations (this is related to the eigenstory we'll discuss later on in the course).

Chapter 9

Determinants

9.1. Oriented Volumes

In this section we'll define the DETERMINANT. The determinant is a powerful tool, as we'll see over the next several sections, and arises from a key observation.

Observation 9.1. A linear transformation $F : \mathbb{R}^n \to \mathbb{R}^m$ is completely determined by the (standard) basis. That is, if we know where a linear function F sends $\vec{e}_1, \dots, \vec{e}_n$ (or any other basis for that matter) then we know how the function behaves on all of $\mathbb{R}^n$.

Using this observation, we can view how a linear transformation behaves by just considering how one special subset of our domain space changes. We define the following.

Definition 9.2. The UNIT n-CUBE is the n-dimensional cube with sides given by the standard basis vectors, and lower-left corner located at the origin. That is,

$$C_n = \{\alpha_1 \vec{e}_1 + \cdots + \alpha_n \vec{e}_n : 0 \le \alpha_i \le 1 \text{ for } i = 1, \dots, n\}.$$

Observe that C_n always has volume 1.

The determinant will be a way to measure how C_n is transformed under F. To measure this accurately, we need one more definition.

Definition 9.3. The ordered basis $\mathcal{B} = \{\vec{b}_1, \dots, \vec{b}_n\}$ is RIGHT-HANDED or POSITIVELY ORIENTED if it can be continuously transformed to the standard basis while remaining linearly independent throughout the transformation. Otherwise, $\mathcal{B}$ is called LEFT-HANDED or NEGATIVELY ORIENTED.

In Core Exercise 43, we'll look at some examples of positive and negatively oriented bases. Another way to detect orientation in $\mathbb{R}^2$ and $\mathbb{R}^3$ is by using the "right-hand rule". We can now define the following.

Definition 9.4. The DETERMINANT of a linear transformation $F : \mathbb{R}^n \to \mathbb{R}^n$, denoted $\det(F)$ or $|F|$, is the oriented volume of the image of the unit n-cube. That is,

$$\det(F) = \begin{cases} \mathrm{vol}(F(C_n)), & \text{if } \{F(\vec{e}_1, \ldots, F(\vec{e}_n)\} \text{ is positively oriented} \\ -\mathrm{vol}(F(C_n)), & \text{if } \{F(\vec{e}_1, \ldots, F(\vec{e}_n)\} \text{ is negatively oriented} \end{cases}$$

Note that we will often write $\det(A)$ to mean the determinant of the linear transformation defined by A. That is, $\det(A) := \det(T_A)$.

Note that the volume of a set S in $\mathbb{R}^n$ can be defined formally using integration, but we won't need to do any calculus here. In the Core Exercises of Module 14, we'll see that we can use trigonometry to calculate the determinant of transformations from $\mathbb{R}^2$ to $\mathbb{R}^2$. In the next section, we'll develop a method to "reduce" determinants of higher-dimension transformations to two-dimensions.

9.2. Parallelepipeds

Example 9.5. Suppose that $F : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$F\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x + 2y \\ x - y \end{pmatrix}.$$

Recall that $\det(F)$ is equal to the oriented volume of $F(C_2)$. We have

$$F(\vec{e}_1) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } F(\vec{e}_2) = \begin{pmatrix} 2 \\ -1 \end{pmatrix}.$$

So, for any $\alpha_1\vec{e}_1 + \alpha_2\vec{e}_2 \in C_2$ we get

$$F(\alpha_1\vec{e}_1 + \alpha_2\vec{e}_2) = \alpha_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \end{pmatrix},$$

where $0 \le \alpha_1, \alpha_2 \le 1$. So, to calculate $\det(F)$ we'd like to calculate the oriented volume of

$$F(C_2) = \left\{\alpha_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \end{pmatrix} : 0 \le \alpha_1, \alpha_2 \le 1\right\}.$$

Observe that $F(C_2)$ is the parallelogram in $\mathbb{R}^2$ obtained by the column vectors of the defining matrix of A. Let's look at how this generalizes.

Definition 9.6. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a subset of $\mathbb{R}^n$. Then, the PARALLELEPIPED defined by the set B is given by

$$P_B := \{\alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n \mid 0 \le \alpha_i \le 1\}.$$

For an $n\times n$ matrix A, we'll let P_A be the parallelepiped defined by the columns of A.

As in the example above, we have the following.

Theorem 9.7 (Volume Theorem I)**.** *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with defining matrix A. Then $\det(F)$ is equal to the oriented volume of the parallelepiped $\mathcal{P}_A$.*

We have the following key observation.

Lemma 9.8. *Let $B = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a subset of $\mathbb{R}^n$, and let C be the subset of B with $\vec{v}_i$ removed. Then, the volume of P_B is equal to the volume of the "base" P_C times the distance of the vector $\vec{v}_n$ from the "base" P_C.*

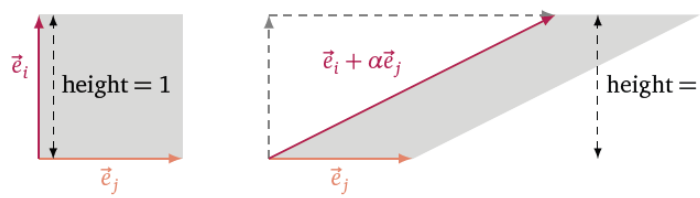
We omit the proof of this lemma since the formal proof is a bit messy, and this is geometrically intuitive enough for us to accept. Note that this could be proven formally by using the definition of the volume as an integral in $\mathbb{R}^n$.

Using this observation will allow us to calculate the determinant of elementary matrices.

Example 9.9. Use geometric reasoning to calculate the determinant of the following elementary matrices.

(1) Calculate $\det(E_1)$ where E_1 is the 2×2 elementary matrix which replaces R_1 with $R_1 + \alpha R_2$

Solution: observe that the transformation T_{E_1} horizontally shears by a factor of α, which does not change the base or height, as seen in the image below



So, $\det(E_1) = \det(I_2) = 1$.

(2) Calculate $\det(E_2)$ where E_2 is the 2×2 elementary matrix which replaces R_1 with α times R_1.

(3) Calculate $\det(E_3)$ where E_3 is the 2×2 elementary matrix which swaps R_1 and R_2.

Our observations from the previous example generalize to n dimensions. We have the following.

Theorem 9.10. *The determinant of the $n \times n$ elementary matrices are given below.*

(1) (Row addition matrices) Let E_1 be the elementary matrix formed by replacing the ith row of I_n with the ith row of I_n plus the sum adding the jth row. Then, $\det(E_1) = \det(I_n) = 1$.

(2) (Row multiplication matrices) Let E_2 be the elementary matrix formed by replacing the ith row of I_n with α times the ith row of I_n. Then, $\det(E_2) = \alpha \det(I_n) = \alpha$.

(3) (Row switching matrices) Let E_3 be the elementary matrix formed by swapping the ith and jth rows of I_n. Then, $\det(E_3) = -\det(I_n) = -1$.

We need two more observations, and then we'll have all of the tools we need to calculate the determinant of any matrix.

Theorem 9.11. *A matrix is invertible if and only if* $\det(A) \neq 0$.

Informal proof sketch. Rather than prove this formally, let's understand this geometrically. Note that when A is not invertible, the columns of A are linearly dependent. So, the parallelepiped $\mathcal{P}_A$ is "missing a dimensions" and will be "flat" in $\mathbb{R}^n$. Hence, $\det(A) = 0$. Conversely, if A is invertible, then $\mathrm{Col}(A) \cong \mathbb{R}^n$ and so $\mathcal{P}_A$ will have nonzero volume. $\square$

Theorem 9.12. *Let A and B be matrices. Then* $\det(AB) = \det(A)\det(B)$.

Informal proof sketch. Again, rather than proving this formally, let's understand this geometrically. Suppose first that A and B are both invertible. If T_A changes the volume of $\mathcal{P}_A$ by a factor of α and T_B changes the volume of $\mathcal{P}_B$ by a factor of β, then we can intuit that $T_A \circ T_B$ changes the volume by a factor of $\alpha\beta$. This gives $\det(AB) = \det(A)\det(B)$ by keeping track of the orientation and recalling that the defining matrix of $T_A \circ T_B$ is equal to AB.

Now, if A or B isn't invertible, observe that neither is AB (for example, by noting that $T_A \circ T_B$ cannot be bijective). Hence $\det(AB) = 0 = \det(A)\det(B)$. $\square$

Example 9.13. Using our results above, calculate

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Proposition 9.14. We have

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Proof. We use a similar method to the previous example. We have

$$\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix} &\sim \begin{pmatrix} a & b \\ 0 & d - \frac{c}{a}b \end{pmatrix} && R_2 - \frac{c}{a}R_1 \\
&\sim \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} && \frac{a}{ad-bc}R_2 \\
&\sim \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} && R_1 - bR_2 \\
&\sim I_2 && \frac{1}{a}R_1.
\end{aligned}$$

Reversing these operations (to write our matrix as a product of elementary matrices) and applying Theorems 9.12 and 9.10 gives

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a \cdot \frac{ad - bc}{a} = ad - bc.$$

$\square$

9.3. Cofactor Expansion

In this section, we'll derive an iterative method to calculate the determinant of any $n \times n$ matrix. We need one more property.

Proposition 9.15. The determinant is row additive. That is,

$$\det \begin{pmatrix} x_{11}+y_{11} & x_{12}+y_{12} & \cdots & x_{1n}+y_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} = \det \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} + \det \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

We're going to omit this proof, since it's quite messy. If you're interested in this proof, you can checkout the multilinearity property in these notes.

Let's look at how to apply these properties in the following example.

Example 9.16. Let's calculate the determinant of the 3×3 matrix

$$A = \begin{pmatrix} a & b & c \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

Then, using Proposition 9.15 as well as Theorems 9.10 and 9.12 we have

$$\begin{aligned}
\det(A) &= \det \begin{pmatrix} a & 0 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \det \begin{pmatrix} 0 & b & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \det \begin{pmatrix} 0 & 0 & c \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \\
&= a \det \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + b \det \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + c \det \begin{pmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \\
&= a \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 5 & 6 \end{pmatrix} + b \det \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 3 \\ 4 & 0 & 6 \end{pmatrix} + c \det \begin{pmatrix} 0 & 0 & 1 \\ 1 & 2 & 0 \\ 4 & 5 & 0 \end{pmatrix} \\
&= a \det \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + b \det \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + c \det \begin{pmatrix} 0 & 0 & 1 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \\
&= a \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 5 & 6 \end{pmatrix} - b \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 4 & 6 \end{pmatrix} + c \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 4 & 5 \end{pmatrix}
\end{aligned}$$

Note that, if $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & x & y \\ 0 & z & w \end{pmatrix}$ we have

$$\det(B) = \det \begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

To see this, observe that P_B has height equal to 1 and base given by the parallelogram defined by vectors

$$\begin{pmatrix} 0 \\ x \\ z \end{pmatrix}, \begin{pmatrix} 0 \\ y \\ w \end{pmatrix}$$

which has oriented volume equal to $\det \begin{pmatrix} x & y \\ z & w \end{pmatrix}$.

So, we have

$$\begin{aligned} \det(A) &= a \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 3 \\ 0 & 5 & 6 \end{pmatrix} - b \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 4 & 6 \end{pmatrix} + c \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 4 & 5 \end{pmatrix} \\ &= a \det \begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix} - b \det \begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix} + c \det \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}, \text{ by (c)} \\ &= \boxed{-3a + 6b - 3c}. \end{aligned}$$

The method we derived above generalizes to n dimensions. We first need a definition.

Definition 9.17. For an $n \times n$ matrix $A = (a_{ij})$, the ij-MINOR of A is defined to be the $(n-1) \times (n-1)$ matrix A_{ij} with the ith row and jth column deleted.

Theorem 9.18 (Cofactor Expansion)**.** *Let A be the $n \times n$ matrix with ij-entry equal to a_{ij}. Then we have*

$$\det(A) = a_{11} \det(A_{11}) - a_{12} \det(A_{12}) + \cdots + (-1)^{n+1} a_{1n} \det(A_{1n}).$$

We give one more result that can simplify our determinant calculations.

Proposition 9.19. The determinant is invariant under the matrix transpose. That is, for any $n \times n$ matrix A we have $\det(A) = \det(A^\top)$

Proof. Suppose first that A is not invertible. If A is $n \times n$ then $\operatorname{rank}(A) < n$, which gives

$$\dim(\mathrm{Row}(A)) = \dim(\mathrm{Col}(A)) = \operatorname{rank}(A) < n.$$

and so by Theorem 9.11 we have $\det(A) = \det(A^\top) = 0$.

Next, suppose that A is invertible. Then by Theorem 7.23 we know that A is a product of elementary matrices, say

$$A = E_1 E_2 \cdots E_\ell.$$

It can be shown that for matrices X and Y we have $(XY)^\top = Y^\top X^\top$ and so

$$\begin{aligned} A^\top &= E_\ell^\top \cdots E_2^\top E_1^\top \\ \Rightarrow \det(A^\top) &= \det(E_1^\top) \cdots \det(E_\ell^\top). \end{aligned}$$

We leave it to the reader to check that for any elementary matrix E we have $\det(E) = \det(E^\top)$. So, by above we get $\det(A^\top) = \det(E_1) \cdots \det(E_\ell) = \det(A)$, as needed. $\square$

Example 9.20. Find the determinant of the following matrix

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

Observe that

$$A^\top \sim \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ 3 & 0 & 2 & 0 \end{pmatrix}$$

by the row operation $R_2 \leftrightarrow R_1$, so we have

$$\begin{aligned}
\det(A) &= \det(A^\top) \\
&= -\det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ 3 & 0 & 2 & 0 \end{pmatrix} \\
&= -\det \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 2 & 0 \end{pmatrix}, \text{ by Theorem 9.18} \\
&= \det \begin{pmatrix} 0 & 2 & 0 \\ 1 & -1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, R_1 \leftrightarrow R_3 \\
&= -2 \det \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \text{ by Theorem 9.18} \\
&= \boxed{2}.
\end{aligned}$$

Chapter 10

Eigenvalues and Eigenvectors

10.1. The Characteristic Equation

In this Chapter, we'll see how understanding the "stretch factors" of a linear transformation can help us gain a geometric understanding of how a linear function transforms a vector space. We have the following definition.

Definition 10.1. Let A be an $n \times n$ matrix. A non-zero vector $\vec{x}$ is an EIGENVECTOR of A if there is a scalar λ such that $A\vec{x} = \lambda\vec{x}$. The scalar λ is called an EIGENVALUE of A.

Geometrically, this means that when we apply the matrix transformation T_A to an eigenvector $\vec{x}$, this is the same thing as stretching the vector $\vec{x}$ by the eigenvalue λ, as visualized in this 3Blue1Brown video.

Example 10.2. For each of the following matrix-vector pairs, determine whether $\vec{x}$ is an eigenvector of the matrix A. If it is find the corresponding eigenvalue λ.

(1) $A = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$.

(2) $B = \begin{pmatrix} 3 & 2 \\ 3 & 8 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

(3) $C = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ and $\vec{x} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

In this section, we'll develop a method to calculate all eigenvalues and eigenvectors of a given matrix. We have the following.

Proposition 10.3. For an $n \times n$ matrix A, the set of eigenvectors of A corresponding to an eigenvalue λ is equal to the nonzero vectors in $\mathrm{Nul}(A - \lambda I_n)$.

Proof. Observe that any $\vec{x} \in \mathrm{Nul}(A - \lambda I_n)$ if and only if

$$
\begin{aligned}
(A - \lambda I_n)\vec{x} &= \vec{0} \\
\Leftrightarrow A\vec{x} - \lambda\vec{x} &= \vec{0} \\
\Leftrightarrow A\vec{x} &= \lambda\vec{x}.
\end{aligned}
$$

□

Definition 10.4. We call the space $\mathrm{Nul}(A - \lambda I_n)$ the λ-EIGENSPACE of A, and use the notation $E_\lambda := \mathrm{Nul}(A - \lambda I_n)$. By Proposition 10.3 the nonzero vectors in E_λ is equal to the set of all eigenvectors with corresponding eigenvalue λ. Geometrically, E_λ is the set of vectors $\vec{x} \in \mathbb{R}^n$ so that the matrix transformation T_A stretches $\vec{x}$ by a factor of λ.

Since we know how to find bases for null spaces, what's left is to find a method to calculate the eigenvalues of a matrix. We have the following.

Proposition 10.5. A (complex) number λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof. Observe that $\vec{x}$ is an eigenvalue of A if and only if it's a nonzero solution to the matrix-vector equation

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

which can be rewritten as

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

By the Invertible Matrix Theorem (Theorem 7.23), the matrix-vector equation above having a nontrivial solution is equivalent to the matrix $A - \lambda I_n$ not being invertible. So, the result follows by Theorem 9.11. □

So, to find eigenvalues, we need to find all solutions to the equation $\det(A - \lambda I_n) = 0$. We define the following.

Definition 10.6. For an $n \times n$ matrix A,

$$\chi_A(\lambda) = \det(A - \lambda I_n).$$

is called the CHARACTERISTIC POLYNOMIAL of A. Note that in your textbook, the notation $\mathrm{char}(A)$ is used instead of χ_A.

By Proposition 10.5 the eigenvalues of a matrix A are the solutions to

$$\chi_A(\lambda) = 0.$$

We have the following observation.

Proposition 10.7. For any $n \times n$ matrix A, the characteristic polynomial $\chi_A(\lambda)$ is a polynomial of degree n.

The formal proof of Proposition 10.7 would use an inductive argument, along with the cofactor expansion formula for the determinant. Instead of worrying about understanding this proof formally, note that:

(1) If we look at the cofactor formula for the determinant, we see that the only operations happening are addition and multiplication, and so we end up with some algebraic expression made up of sums and products of real numbers and our unknown x, which precisely defines a polynomial.

(2) If an $n \times n$ matrix A has diagonal entries $d_1, d_2, \ldots, d_n$, then the highest degree term coming out of the cofactor exapansion will be $(d_1 - \lambda)(d_2 - \lambda) \cdots (d_n - \lambda)$ (convince yourself of this in the 3×3 case). So, the degree of $\chi_A(\lambda)$ will be at most n, and in fact in can be argued that the degree is equal to n (by noting that the remaining summands each have degree strictly smaller than n).

This resolves our problem of finding eigenvectors. For an $n \times n$ matrix A

- First, we can find all eigenvalues of A by solving the polynomial equation $\chi_A(\lambda) = 0$. This has at most n solutions (since the number of solutions to a polynomial equation is less than the degree of the polynomial).
- For each eigenvalue λ, we can find all eigenvectors with eigenvalue λ by finding the null space
$$E_\lambda = \mathrm{Nul}(A - \lambda I_n).$$

Example 10.8. Let
$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$
Then
$$\chi_A(\lambda) = \begin{pmatrix} 1-\lambda & -1 \\ 1 & 1-\lambda \end{pmatrix} = (1-\lambda)^2 + 1$$
So, $\chi_A(\lambda) = (1-\lambda)^2 + 1$ (let's leave it in this form), which means the eigenvalues of A will be solutions to the equation
$$(1-\lambda)^2 + 1 = 0$$
$$\Rightarrow (1-\lambda)^2 = -1.$$
While this equation has no *real* solutions, it does have complex ones. Namely,
$$\lambda = i - 1, -i - 1.$$
Note that in general, eigenvalues can be complex. This story is a bit tricky, because the corresponding eigenvectors will also be complex, and we've only been working with *real* vector spaces in this class. We're going to sweep this story under the rug for now, and only focus on examples where our eigenvalues are real. Time permitting, we'll return to this story (in particular, we'll spend time thinking about what this means geometrically).

The following example illustrates how eigenvalues can help us understand how a given matrix transforms a vector.

Example 10.9. Consider the matrix
$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Let's use some eigentheory to understand the matrix transformation T_A. First, we compute

$$\chi_A(x) = \det \begin{pmatrix} 1-x & 0 & 1 \\ 0 & 1-x & 1 \\ 0 & 0 & 2-x \end{pmatrix} = (1-x)^2(2-x)$$

and so $\chi_A(x) = (1-x)^2(2-x)$. This tells us that A has eigenvalues $\lambda = 1$ and $\lambda = 2$. Once we've found our eigenvalues, we can now compute the set of eigenvectors as the null space of $A - \lambda I_3$. With a bit of work, we compute

$$E_1 = \text{Nul}(A - I_3) = \text{Span}\left(\vec{v}_1, \vec{v}_2\right),$$

where

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

and $E_2 = \text{Nul}(A - 2 \cdot I_3) = \text{Span}\left(\vec{v}_3\right)$, where

$$\vec{v}_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Observe that $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a basis for $\mathbb{R}^3$. For any $\vec{v} \in \mathbb{R}^3$ write

$$\vec{v} = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3.$$

Using linearity of the matrix transformation T_A gives

$$\begin{aligned} T_A(\vec{x}) &= x_1 T_A(\vec{v}_1) + x_2 T_A(\vec{v}_2) + x_3 T_A(\vec{v}_3) \\ &= (1)(x_1\vec{v}_1 + x_2\vec{v}_2) + 2(x_3\vec{v}_3). \end{aligned}$$

where the second equality follows by recalling that $\vec{v}_1, \vec{v}_2 \in E_1$ and $\vec{v}_3 \in E_2$. Observe that this transformation is simpler to understand with $\mathcal{B}$ coordinates, since we have

$$T_A\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}_{\mathcal{B}}\right) = \begin{pmatrix} x_1 \\ x_2 \\ 2x_3 \end{pmatrix}_{\mathcal{B}}.$$

That is, T_A is the transformation that does nothing to the first two $\mathcal{B}$-coordinates, and scales the third $\mathcal{B}$-coordinate by 2. Observe in this case we have

$$[T_A]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

In the following section, we'll investigate how to find "nice bases" so that we can more easily understand a given matrix transformation.

10.2. Diagonalization

We define the following.

Definition 10.10. A matrix is called DIAGONAL if the only nonzero entries in the matrix appear on the diagonal. We write

$$D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$$

to notate the diagonal matrix

$$D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ 0 & 0 & \cdots & d_n \end{pmatrix}.$$

Note that if $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ then

$$D \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1x_1 \\ d_2x_2 \\ \vdots \\ d_nx_n \end{pmatrix},$$

and so the matrix transformation T_D can be understood as the transformation which stretches each coordinate by a of factor d_i. We define the following.

Definition 10.11. An $n \times n$ matrix A is called DIAGONALIZABLE if it is similar to a diagonal matrix.

Observe that a matrix A is similar if it represents a function $F : \mathbb{R}^n \to \mathbb{R}^n$ so that $[F]_{\mathcal{B}}$ is a diagonal matrix for some basis $\mathcal{B}$ of $\mathbb{R}^n$. The remaining results in this section give a characterization of such matrices. The following result generalizes what we observed in the example in the previous section.

Theorem 10.12 (The Diagonalization Theorem)**.** *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. Furthermore, if A is diagonalizable with linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$ then $D = C^{-1}AC$ where*

$$D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \text{ and } C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

Proof. First, suppose that A has n linearly independent eigenvectors $\vec{v}_1, \ldots, \vec{v}_n$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$ and let C be the matrix

$$C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

Since the columns of C are linearly independent, we know that C is invertible. To compute the matrix $C^{-1}AC$, we'll multiply by the standard basis vectors $\vec{e}_i$. So, if $\vec{e}_i$ is the ith standard basis vector, note that

$$C\vec{e}_i = \vec{v}_i \Rightarrow \vec{e}_i = C^{-1}\vec{v}_i. \tag{10.1}$$

We have

$$\begin{aligned}
C^{-1}AC\vec{e}_i &= C^{-1}A\vec{v}_i, \text{ by Equation (10.1)} \\
&= C^{-1}\lambda_i\vec{v}_i, \text{ since } \vec{v}_i \text{ is an eigenvector of } \lambda_i \\
&= \lambda_i C^{-1}\vec{v}_i \\
&= \lambda_i\vec{e}_i, \text{ by Equation (10.1)}.
\end{aligned}$$

Since the ith column of $C^{-1}AC = \lambda_i\vec{e}_i$ we have that

$$C^{-1}AC = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

as desired. Conversely, suppose that A is diagonalizable, so that

$$C^{-1}AC = D$$

for an invertible matrix C and diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$. Suppose that C has column vectors equal to $\vec{c}_i$. Since C is invertible, we know that $\{\vec{c}_1, \ldots, \vec{c}_n\}$ is a linearly independent set. So, we just need to show that $\vec{c}_i$ is an eigenvector with eigenvalue d_i. As we observed before, note that

$$C\vec{e}_i = \vec{c}_i \Rightarrow \vec{e}_i = C^{-1}\vec{c}_i. \tag{10.2}$$

So we have

$$\begin{aligned}
A\vec{c}_i &= CDC^{-1}\vec{c}_i \\
&= CD\vec{e}_i, \text{ by Equation (10.2)} \\
&= Cd_i\vec{e}_i, \text{ since the } i\text{th column of } D \text{ is } d_i\vec{e}_i \\
&= d_iC\vec{e}_i \\
&= d_i\vec{c}_i, \text{ by Equation (10.2)}
\end{aligned}$$

So, $\vec{c}_i$ is an eigenvector of A with eigenvalue d_i as desired. □

Example 10.13. Determine which of the following matrices A_i are diagonalizable. For those that are, find an invertible matrix C_i and diagonal matrix D_i so that $D_i = C_i^{-1}A_iC_i$.

$$A_1 = \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}\ A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{pmatrix}\ A_3 = \begin{pmatrix} 4 & -3 & 0 \\ 2 & -1 & 0 \\ 1 & -1 & 1 \end{pmatrix}.$$

Solution. A_1 has eigenvalues 2 and -1 and

$$E_2 = \operatorname{Span}\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)$$

$$E_{-1} = \operatorname{Span}\left(\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right).$$

Observe that

$$\vec{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } \vec{w} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

are linearly independent eigenvectors, and so by the Diagonalization Theorem, the matrix A_1 is diagonalizable. If we let

$$D_1 = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix} \text{ and } C_1 = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$$

then we can see that $D_1 = C_1^{-1} A_1 C_1$.

A_2 has eigenvalues 2 and 1 and

$$E_2 = \text{Span}\left(\begin{pmatrix}1\\2\\4\end{pmatrix}\right)$$

$$E_1 = \text{Span}\left(\begin{pmatrix}1\\1\\1\end{pmatrix}\right).$$

Since the eigenvectors is E_2 (resp E_1) are linearly dependent, it's not possible to have three linearly independent eigenvectors. So, by the Diagonalization Theorem, we know that A_2 is not diagonalizable.

A_3 has eigenvalues 2 and 1 and

$$E_2 = \text{Span}\left(\begin{pmatrix}3\\2\\1\end{pmatrix}\right)$$

$$E_1 = \text{Span}\left(\begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}\right).$$

Since the vectors

$$\vec{v}_1 = \begin{pmatrix}3\\2\\1\end{pmatrix}, \vec{v}_2 = \begin{pmatrix}0\\0\\1\end{pmatrix}, \text{ and } \vec{v}_3 \begin{pmatrix}1\\1\\0\end{pmatrix}$$

are linearly independent, then by the Diagonalization theorem A_3 is Diagonalizable. If we let

$$D_3 = \begin{pmatrix}2 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{pmatrix} \text{ and } C_3 = \begin{pmatrix}3 & 0 & 1\\2 & 0 & 1\\1 & 1 & 0\end{pmatrix}$$

then we can see that $D_3 = C_3^{-1} A_3 C_3$.

Note that if A is diagonalizable with $D = C^{-1}AC$ we can write $A = CDC^{-1}$. This gives a useful decomposition of our matrix A. We have the following definition.

Definition 10.14. Suppose that A is an $n \times n$ diagonalizable matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ and corresponding linearly independent eigenvectors $\vec{v}_1, \dots, \vec{v}_n$. Let $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ and $C = \begin{pmatrix}\vec{v}_1 & \cdots & \vec{v}_n\end{pmatrix}$. We call the equality

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

the EIGENDECOMPOSITION of the matrix A.

Example 10.15. Let's find a transformation $F : \mathbb{R}^2 \to \mathbb{R}^2$ which stretches every vector in the $\begin{pmatrix}1\\1\end{pmatrix}$ direction by 2 and in the $\begin{pmatrix}0\\-1\end{pmatrix}$ direction by 3.

Observing that F is linear, we know that $F = T_A$ for a matrix A. Furthermore, we know that A has eigenvalues $2, 3$ with corresponding eigenvectors

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

By the Diagonalization Theorem, A has eigendecomposition

$$A = \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 2 & 0 \\ -1 & 3 \end{pmatrix}.$$

The eigendecomposition can also help us compute large powers of a matrix. We have the following.

Proposition 10.16. Let A be a diagonalizable matrix with eigendecomposition $A = CDC^{-1}$. Then,

$$A^n = CD^nC^{-1}$$

for any integer n.

Proof. We have

$$\begin{aligned} A^n &= (CDC^{-1})^n \\ &= \underbrace{(CDC^{-1})(CDC^{-1}) \cdots (CDC^{-1})}_{n \text{ times}} \\ &= \underbrace{CD(C^{-1}C)DC^{-1} \cdots CDC^{-1}}_{n \text{ times}} \\ &= C \underbrace{DD \cdots D}_{n \text{ times}} C^{-1} \\ &= CD^nC^{-1}. \end{aligned}$$

□

To use the Diagonalization Theorem, we need a method to determine whether A has enough linearly independent eigenvectors. We have the following Proposition, which gives a special case.

Proposition 10.17. Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of a matrix A, and suppose that $\vec{v}_i \in E_{\lambda_i}$ for each $i \in \{1, \ldots, k\}$. Then $\{\vec{v}_1, \ldots, \vec{v}_k\}$ is a linearly independent set.

Proof Outline. We'll fill in the details of this outline in class together.

- Let λ_1 and λ_2 be distinct eigenvalues of a matrix A, and suppose that $\vec{v}_1 \in E_{\lambda_1}$ and $\vec{v}_2 \in E_{\lambda_2}$. If $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{0}$ observe that we have the equalities

$$x_1\lambda_1\vec{v}_1 + x_2\lambda_2\vec{v}_2 = \vec{0}$$

$$x_1\lambda_1\vec{v}_1 + x_2\lambda_1\vec{v}_2 = \vec{0}.$$

- Taking the difference of the equalities above shows that $x_2 = 0$. Similarly we can show that $x_1 = 0$.

- Now, suppose that $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are eigenvectors of a matrix A with corresponding eigenvalues $\lambda_1, \lambda_2, \lambda_3$. If $\lambda_i \neq \lambda_j$ for any $i \neq j$ (that is, the λ_i are all distinct real numbers), then we can use a similar strategy to what we've done above to show that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a linearly independent set.

In the outline above, we showed that any two eigenvectors with distinct eigenvalues are linearly independent. We then used this to show that any three eigenvectors with distinct eigenvalues are linearly independent. We could proceed using the same method to show any *four* eigenvectors with distinct eigenvalues are linearly independent, and so on. The formal version of this argument is called proof by induction. We omit the details here, since this isn't a proof method we've discussed. □

Corollary 10.18. If an $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.

Proof. If A has n distinct eigenvalues, then by Proposition 10.17 A must then have n distinct eigenvectors, and so A is diagonalizable by the Diagonalization Theorem. □

Remark 10.19. Note that the converse of Proposition 10.17 does not hold. That is, it's not the case that if an $n \times n$ matrix A is diagonalizable then A must have n distinct eigenvalues. For example, if we let A be the matrix defined in Example 10.9 is diagonalizable but only has two eigenvalues. The following definition will help us characterize $n \times n$ diagonalizable matrices with less than n eigenvalues.

Definition 10.20. For an eigenvalue λ of a matrix A, the GEOMETRIC MULTIPLICITY of λ is defined to be the dimension of the λ-eigenspace E_λ.

Example 10.21. Let A be as in Example 10.9. Recall that A has eigenvalues $\lambda = 1$ and $\lambda = 2$ and that $\dim E_1 = 2$ and $\dim E_2 = 1$. So the geometric multiplicity of the eigenvalue $\lambda = 1$ is equal to 2, and the geometric multiplicity of the eigenvalue $\lambda = 2$ is equal to 1.

Example 10.22. Recall the setting of Example 10.9. If

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

then we found that

$$\chi_A(x) = (1-x)^2(2-x),$$

and that E_1 had dimension 2 and E_2 had dimension 1. In this case, observe that the dimension of the λ-eigenspace exactly matches the power on the term $(\lambda - x)$ in χ_A. This turns out to always be the case for diagonalizable matrices. We have the following definition.

Definition 10.23. Suppose that a matrix A has eigenvalue λ. The ALGEBRAIC MULTIPLICITY of λ is the largest integer m so that $(x-\lambda)^m$ divides the characteristic polynomial χ_A of A.

Example 10.24. In Example 10.9, the algebraic multiplicity of $\lambda = 1$ is 2, and the algebraic multiplicity of $\lambda = 2$ is 1.

We can now add the our Diagonalization Theorem. We have the following.

Theorem 10.25 (Diagonalization Theorem, final version)**.** *Let A be an $n \times n$ matrix whose eigenvalues are all real. The following are equivalent.*

(1) A is diagonalizable;

(2) The sum of the geometric multiplicities of A is equal to n;

(3) The geometric multiplicity of every eigenvalue λ is equal to the algebraic multiplicity of λ.

Furthermore, if A is diagonalizable with linearly independent eigenvectors $\vec{v}_1, \dots, \vec{v}_n$ corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$ then $D = C^{-1}AC$ where

$$D = \operatorname{diag}(\lambda_1, \dots, \lambda_n) \text{ and } C = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}.$$

Note that, because of Proposition 10.17, (1) $\Leftrightarrow$ (2) can be seen as a rephrasing of our original Diagonalization Theorem (Theorem 10.12). All that's left to show in this final version is (2) $\Leftrightarrow$ (3). We first need a Lemma.

Lemma 10.26. *The geometric multiplicity of any eigenvalue is less than or equal to its algebraic multiplicity.*

We'll omit the proof of this Lemma (there are ways to prove this with the information we have, but they're all a bit complicated... unfortunately all resources I've seen site the proof of this result as "beyond the scope of this course"). Let's see how this proves our Theorem.

Proof of Theorem 10.25. Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \dots, \lambda_k$ and let λ_i have geometric multiplicity d_i and algebraic multiplicity m_i. By the previous lemma, we know that $d_i \le m_i$.

(2) $\Rightarrow$ (3): Suppose that $d_1 + \cdots + d_k = n$. Note that the characteristic polynomial of A is of degree n, and since we can write

$$\chi_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$$

this gives $m_1 + \cdots + m_k = n$. So, if $d_i < m_i$ for any i we would have

$$d_1 + \cdots + d_k < m_1 + \cdots + m_k < n,$$

contradicting our assumption. Since by our Lemma $d_i \le m_i$, we must then have the equality $d_i = m_i$ for every i.

(3) $\Rightarrow$ (2): Conversely, suppose that $m_i = d_i$ for all i. Then we have

$$n = m_1 + \cdots + m_k = d_1 + \cdots + d_k,$$

as desired. □

Remark 10.27. Note that the addition of part (3) to our Theorem doesn't quite simplify our computation when our matrix *is* diagonalizable: no matter what we do, we still need to compute the dimension of the λ-eigenspace for each value of λ. But it does give the potential to simplify our justification that a matrix *is not* diagonalizable: all we need to do is find ONE eigenvalue where the geometric and algebraic multiplicities do not agree.

Example 10.28. Consider the matrix

$$A_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & -5 & 4 \end{pmatrix}.$$

The characteristic polynomial of A_2 is given by

$$\chi_{A_2}(x) = -(x-2)(x-1)^2$$

and so $\lambda = 1$ has algebraic multiplicity 2. However,

$$A_2 - 1 \cdot I_3 = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 2 & -5 & 3 \end{pmatrix}$$

which is row equivalent to a matrix with 2 pivots. So, $\text{nullity}(A - I_3) = 1$ which tells us that the geometric multiplicity of $\lambda = 1$ is equal to 1. So, by our final version of the Diagonalization Theorem, we know that A_2 is not diagonalizable.

Chapter 11

Orthogonality

In this chapter, we'll add more "geometric structure" to our understanding of Euclidean space, which will help us connect several ideas from the course.

11.1. The Dot Product

Definition 11.1. Let

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \text{ and } \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

be vectors in $\mathbb{R}^n$. The DOT PRODUCT of $\vec{u}$ and $\vec{v}$ is the scalar

$$\vec{u} \cdot \vec{v} := u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Note that is we think of the vector $\vec{u}$ as an $n \times 1$ matrix, we can define the dot product instead as the matrix-vector product $\vec{u} \cdot \vec{v} = \vec{u}^\top \vec{v}$.

Example 11.2. Let

$$\vec{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \text{ and } \vec{v} = \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}.$$

Then,

$$\vec{u} \cdot \vec{v} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} = 1(-1) + 2(0) + 3(2) = \boxed{5}.$$

Observe that the dot product satisfies the following properties.

Proposition 11.3. Let $\vec{u}, \vec{v}$ and $\vec{w}$ be vectors in $\mathbb{R}^n$ and let c be a scalar. Then,

(1) *Commutativity:* $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$
(2) *Distributivity with Addition:* $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$

(3) *Distributivity with Scalar Multiplication:* $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v})$

This operation is will help us add some geometric structure to $\mathbb{R}^n$. Formally, the dot product defines something called an INNER PRODUCT on $\mathbb{R}^n$. We'll omit these details here, but inner products are the operations needed to define notions of distance and angles on vector spaces more generally. We have the following.

Definition 11.4. The NORM of a vector $\vec{u}$ is defined by

$$\|\vec{u}\| := \sqrt{\vec{u} \cdot \vec{u}}.$$

Observe that, for any vector $\vec{u}$, we have

$$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

and so the norm measures the length of a given vector. This allows us to define the distance between vectors.

Definition 11.5. Let $\vec{u}$ and $\vec{v}$ be vectors in $\mathbb{R}^n$. Then, the DISTANCE between $\vec{u}$ and $\vec{v}$, denoted $d(\vec{u}, \vec{v})$ is equal to the length of the vector $\vec{u} - \vec{v}$. That is,

$$d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|.$$

Definition 11.6. Let $\vec{u}$ and $\vec{v}$ be nonparallel vectors in $\mathbb{R}^n$. Then, the ANGLE between vectors $\vec{u}$ and $\vec{v}$ is the smaller of the two angles between $\vec{u}$ and $\vec{v}$ when drawn in the plane spanned by $\{\vec{u}, \vec{v}\}$. If $\vec{u}$ and $\vec{v}$ are parallel, then the angle between them is 0.

Example 11.7. Find the angle between the vectors

$$\vec{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } \vec{v} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

(Hint: use law of cosines).

What we found in the previous example generalizes. We have the following.

Proposition 11.8. Let $\vec{u}$ and $\vec{v}$ be nonparallel vectors in $\mathbb{R}^n$. Then, the angle between $\vec{u}$ and $\vec{v}$ is the value $\theta \in (0, \pi]$ given by

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

Proof. By Law of Cosines we have

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\| \cos\theta.$$

Furthermore, observe that for any vector $\vec{x}$ we have

$$\|\vec{x}\|^2 = \vec{x} \cdot \vec{x}.$$

So, we get

$$(\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) = \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2\|\vec{u}\|\vec{v}\| \cos\theta.$$

Using Proposition 11.3 we have

$$\begin{aligned} \vec{u} \cdot \vec{u} - 2\vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{v} &= \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} - 2\|\vec{u}\|\vec{v}\| \cos\theta \\ \Rightarrow -2\vec{u} \cdot \vec{v} &= -2\|\vec{u}\|\vec{v}\| \cos\theta \\ \Rightarrow \cos\theta &= \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}. \end{aligned}$$

□

Corollary 11.9. Vectors $\vec{u}$ and $\vec{v}$ are perpendicular (that is, the angle between them is equal to 90°) if and only if $\vec{u} \cdot \vec{v} = 0$.

11.2. Orthogonal Complements

Definition 11.10. Given a vector subspace V of $\mathbb{R}^n$, the ORTHOGONAL COMPLEMENT of V is the set

$$V^\perp := \{\vec{x} \in \mathbb{R}^n \mid \vec{x} \cdot \vec{v} = 0 \text{ for all } \vec{v} \in V\}.$$

Example 11.11. Use geometric reasoning to find the orthogonal complement of the xy-plane. That is, find $V^\perp$ where

$$V = \operatorname{Span}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right).$$

The following result will allow us to calculate $V^\perp$ for any vector space.

Theorem 11.12. *Let V be a vector subspace of $\mathbb{R}^n$ with basis $\{\vec{v}_1, \dots, \vec{v}_m\}$ and let $A = \begin{pmatrix}\vec{v}_1 & \cdots & \vec{v}_m\end{pmatrix}$. Then, $V^\perp = \operatorname{Nul}(A^\top)$.*

Proof Outline. We'll fill in the details of this outline in class together.

- Observe that if $A = \begin{pmatrix}\vec{v}_1 & \cdots & \vec{v}_n\end{pmatrix}$ then we have
$$A^\top \vec{x} = \begin{pmatrix}\vec{v}_1^\top \vec{x}\\ \vdots \\ \vec{v}_n^\top \vec{x}\end{pmatrix} = \begin{pmatrix}\vec{v}_1 \cdot \vec{x}\\ \vdots \\ \vec{v}_n \cdot \vec{x}\end{pmatrix}.$$
- So, if we take any $\vec{x} \in V^\perp$ then $\vec{v}_i \cdot \vec{x} = 0$. By above we get $A^\top \vec{x} = \vec{0}$. Hence, $\vec{x} \in \operatorname{Nul}(A^\top)$.
- Now, if we take any $\vec{x} \in \operatorname{Nul}(A^\top)$ then by our first item we see that $\vec{v}_i \cdot \vec{x} = 0$ for all $i = 1, \dots, n$. We can use this to show that $\vec{v} \cdot \vec{x} = 0$ for any vector $\vec{v} \in V$, and so $\vec{x} \in V^\perp$.

□

Definition 11.13. Let $\vec{x}$ be a vector in $\mathbb{R}^n$ and V a subspace of $\mathbb{R}^n$. The ORTHOGONAL PROJECTION of $\vec{v}$ onto V is the closet vector in V to $\vec{x}$, and is denoted $\vec{x}_V$ (sometimes people use the notation $\operatorname{proj}_V(\vec{x})$).

Example 11.14. Let $\vec{x} = \begin{pmatrix}1\\2\\3\end{pmatrix}$. If V is the xy-plane, find $\vec{x}_V$

The following result will allow us to calculate orthogonal projections more generally.

Theorem 11.15. *Let V be a vector subspace of $\mathbb{R}^n$ with basis $\{\vec{v}_1, \dots, \vec{v}_m\}$, and let $A = \begin{pmatrix}\vec{v}_1 & \cdots & \vec{v}_m\end{pmatrix}$. Then, for all vectors $\vec{x}$ in $\mathbb{R}^m$ we have*

$$\vec{x}_V = A(A^\top A)^{-1} A^\top \vec{x}.$$

We first need a lemma.

Lemma 11.16. *Let A be an $n \times m$ matrix with linearly independent columns. Then, the $m \times m$ matrix $A^\top A$ is invertible.*

Proof. We'll show that the null space of $A^\top A$ is trivial. Suppose that $\vec{x} \in \mathrm{Nul}(A^\top A)$. Then, $A^\top(A\vec{x}) = \vec{0}$. So, we get that $A\vec{x} \in \mathrm{Nul}(A^\top)$, but by Theorem 11.12 we have that $\mathrm{Nul}(A^\top) = \mathrm{Col}(A)^\perp$. Also note that $A\vec{x} \in \mathrm{Col}(A)$. So, if we let $V = \mathrm{Col}(A)$ and $\vec{v} = A\vec{x}$ we see that

$$\vec{v} \in V \cap V^\perp.$$

But this means $\|\vec{v}\|^2 = \vec{v} \cdot \vec{v} = 0 \Rightarrow \vec{v} = \vec{0}$. Hence, $A\vec{x} = 0$. But since A has linearly independent columns, we must have $\vec{x} = \vec{0}$. Hence, $\mathrm{nullity}(A^\top A) = 0$. By rank-nullity this gives $\mathrm{rank}(A^\top A) = 0$ and by the invertible matrix theorem we see that $A^\top A$ is invertible. □

We can now prove our theorem.

Proof of Theorem 11.15. Suppose that V has basis $\{\vec{v}_1, \ldots, \vec{v}_n\}$ and let $A = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}$. Since $\vec{x} - \vec{x}_V \in V^\perp$, and by Theorem 11.12 we know that $V^\perp = \mathrm{Nul}(A^\top)$, we have

$$A^\top(\vec{x} - \vec{x}_V) = \vec{0} \Rightarrow A^\top\vec{x} = A^\top\vec{x}_V.$$

Unfortunately since A is not necessarily square, $A^\top$ may not be invertible. So we have to work a little harder to solve for $\vec{x}_V$. Since $\vec{x}_V \in V = \mathrm{Col}(A)$ there exists a vector $\vec{y} \in \mathbb{R}^n$ so that $A\vec{y} = \vec{x}_V$. Substituting this into the equation above yields

$$A^\top\vec{x} = A^\top A\vec{y}.$$

By Lemma 11.16 we know that $A^\top A$ is invertible and so we get

$$\vec{y} = (A^\top A)^{-1}A^\top\vec{x}.$$

Multiplying both sides of this equation on the left by A gives

$$A\vec{y} = A(A^\top A)^{-1}A^\top\vec{x},$$

and recalling that $\vec{y}$ satisfied $A\vec{y} = \vec{x}_V$ gives

$$\vec{x}_V = A(A^\top A)^{-1}A^\top\vec{x},$$

as needed. □

This result gives us the following Corollary, which can also be shown geometrically (as in Module 5 of your textbook).

Corollary 11.17. Let $V = \mathrm{Span}(\vec{v})$ be a 1-dimensional subspace of $\mathbb{R}^n$. Then, for any $\vec{x} \in \mathbb{R}^n$ we have

$$\vec{x}_V = \frac{\vec{v} \cdot \vec{x}}{\vec{v} \cdot \vec{v}}\vec{v}.$$

Proof. This follows by Theorem 11.15 by noting that $A = \begin{pmatrix} \vec{v} \end{pmatrix}$. □

Remark 11.18. When projecting a vector onto a one-dimensional subspace $V = \mathrm{Span}(\vec{v})$, the notation $\mathrm{proj}_{\vec{v}}(\vec{x})$ is sometimes used to mean $\vec{x}_V$.

11.3. Orthogonal Decomposition

Observe that if V is any subspace of $\mathbb{R}^n$ we have the $\vec{x}-\vec{x}_V \in V^\perp$. This by the Pythagorean theorem, since $\vec{x}_V$ is defined to be the vector $\vec{y}$ so that $\|\vec{x}-\vec{y}\|$ is minimal. This gives the following.

Theorem 11.19. *For any vector subspace V of $\mathbb{R}^n$ we have $\mathbb{R}^n = V \oplus V^\perp$ (this is called the* DIRECT SUM*). That is,*

(1) $\mathbb{R}^n = V + V^\perp$

(2) $V \cap V^\perp = \{\vec{0}\}$*, and*

(3) every vector $\vec{x} \in \mathbb{R}^n$ can be written uniquely in the form $\vec{x} = \vec{v} + \vec{w}$ for $\vec{v} \in V$ and $\vec{w} \in V^\perp$.

Proof. Take any $\vec{x} \in \mathbb{R}^n$. From our observation above, we know that there exists a vector $\vec{w} \in V^\perp$ so that $\vec{x}-\vec{x}_V = \vec{w}$ which gives

$$\vec{x} = \vec{x}_V + \vec{w}.$$

This shows that $\mathbb{R}^n = V + V^\perp$. Note that if $\vec{v} \in V \cap V^\perp$ we have

$$\vec{v}\cdot\vec{v} = 0 \Rightarrow \|\vec{v}\| = 0,$$

and since the components of vectors in $\mathbb{R}^n$ are real this gives $\vec{v} = \vec{0}$. Finally, from above we know that we can write

$$\vec{x} = \vec{x}_V + \vec{w},$$

for $\vec{w} \in V^\perp$ and noting that $\vec{x}_V \in V$ (by definition). So, suppose we could write

$$\vec{x} = \vec{v} + \vec{w}'$$

for some $\vec{v} \in V$ and $\vec{w} \in V^\perp$. Then we have

$$\vec{x}_V - \vec{v} = \vec{w} - \vec{w}'$$

and so $\vec{w}-\vec{w}' \in V \cap V^\perp$ which we know is trivial, by above. Hence $\vec{w} = \vec{w}'$. Similarly, we get $\vec{v} = \vec{x}_V$. □

As a Corollary, we obtain the following, sometimes called the "fundamental theorem of linear algebra".

Corollary 11.20. Suppose that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with defining matrix A. Then

$$\mathbb{R}^n = \mathrm{Nul}(A) \oplus \mathrm{Row}(A), \text{ and}$$

$$\mathbb{R}^m = \mathrm{Nul}(A^\top) \oplus \mathrm{Col}(A).$$

Proof. We have

$$\begin{aligned}\mathbb{R}^n &= \mathrm{Nul}(A) \oplus \mathrm{Nul}(A)^\perp, \text{ by Theorem 11.19}\\ \\ &= \mathrm{Nul}(A) \oplus \mathrm{Col}(A^\top), \text{ by Theorem 11.12}\\ &= Nul(A) \oplus \mathrm{Row}(A), \text{ by definition.}\end{aligned}$$

Next, note that $A^\top$ defines a transformation from $\mathbb{R}^m \to \mathbb{R}^n$, and so $\mathrm{Nul}(A^\top)$ is a subspace of $\mathbb{R}^m$. Arguing as above gives

$$\begin{aligned}\mathbb{R}^m &= \mathrm{Nul}(A^\top) \oplus \mathrm{Nul}(A^\top)^\perp \\ &= \mathrm{Nul}(A^\top) \oplus \mathrm{Col}((A^\top)^\top) \\ &= \mathrm{Nul}(A^\top) \oplus \mathrm{Col}(A),\end{aligned}$$

as needed. □

Note that the following sections are included for reference only. You will not be tested on anything beyond Section 11.3.

11.4. Orthogonal Matrices and the Gram-Schmidt Process

In the previous chapters, we saw that the fundamental object needed to understand a vector space is a basis. We learned that real vector spaces of dimension n are all isomorphic to $\mathbb{R}^n$, and we saw how different bases define coordinate systems on our vector spaces which can help us better understand certain linear transformations (this is the eigen-story from the previous chapter). Let's look at how the dot product interacts with different bases for $\mathbb{R}^n$.

Definition 11.21. A basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is ORTHOGONAL if $\vec{v}_i \cdot \vec{v}_j = 0$ for every $i \neq j$. If it's also the case that $\|\vec{v}_i\| = 1$ for every i we call $\mathcal{B}$ an ORTHONORMAL BASIS for $\mathbb{R}^n$.

Example 11.22. The basis

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

is an orthogonal basis for $\mathbb{R}^3$, but it is not orthonormal. The standard basis for $\mathbb{R}^n$ is an orthonormal basis.

The following Proposition tells us that orthonormal bases preserve dot products, and as a consequence preserve distances and angles.

Proposition 11.23. Let $\mathcal{B}$ be an orthonormal basis for $\mathbb{R}^n$ and take any vectors $\vec{x}, \vec{y}$ in $\mathbb{R}^n$. Then

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

In particular, we have $\|\vec{x}\| = \|[\vec{x}]_{\mathcal{B}}\|$.

Proof. Suppose that $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ and write

$$[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ and } [\vec{y}]_{\mathcal{B}} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

That is,

$$\vec{x} = x_1\vec{v}_1 + \cdots + x_n\vec{v}_n$$

$$\vec{y} = y_1\vec{v}_1 + \cdots + y_n\vec{v}_n.$$

Then we have

$$\vec{x} \cdot \vec{y} = (x_1\vec{v}_1 + \cdots + x_n\vec{v}_n) \cdot (y_1\vec{v}_1 + \cdots + y_n\vec{v}_n)$$

Using the distributive properties of the dot product (Proposition 11.3), we'll end up with a sum of terms of the form

$$x_iy_j\vec{v}_i\vec{v}_j.$$

But, since we know that $\vec{v}_i \cdot \vec{v}_j = 0$ whenever $i \neq j$ then we have

$$\vec{x} \cdot \vec{y} = x_1y_1\vec{v}_1 \cdot \vec{v}_1 + \cdots x_ny_n\vec{v}_n \cdot \vec{v}_n.$$

But we also know that $\vec{v}_i \cdot \vec{v}_i = \|\vec{v}\|^2 = 1$, since our basis is orthonormal. So, we have

$$\vec{x} \cdot \vec{y} = x_1y_1 + \cdots + x_ny_n$$

as desired. $\square$

Example 11.24. Consider the basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$ where

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \text{ and } \vec{v}_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix},$$

which is **not orthonormal**, and consider a vector $\vec{x}$ written in terms of this basis, say $\vec{x} = -\vec{v}_1 + 2\vec{v}_2$. This gives

$$[\vec{x}]_{\mathcal{B}} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \Rightarrow \|[\vec{x}]_{\mathcal{B}}\| = \sqrt{5}.$$

But if we write $\vec{x}$ in the standard basis, we get

$$\vec{x} = -\vec{v}_1 + 2\vec{v}_2 = \begin{pmatrix} 3 \\ -1 \end{pmatrix} \Rightarrow \|\vec{x}\| = \sqrt{10}.$$

So, to calculate the actual norm, it was necessary for us to calculate the coordinates of $\vec{x}$ in the standard basis.

Proposition 11.23 tells us that orthonormal bases are the "right" kinds of bases to use when we want to understand vector spaces as so called "inner product spaces". So, the next natural question is, how can we find an orthonormal basis of a given vector space? It turns out that we transform any basis into an orthonormal basis using the Gram-Schmidt Process outlined below. First, we set some notation for convenience.

Suppose that $V = \text{Span}(\vec{v})$. Then we write $\text{proj}_{\vec{v}}\vec{x} := \vec{x}_V$.

Theorem 11.25 (The Gram-Schmidt Process)**.** *Let V be a vector subspace of $\mathbb{R}^n$ with basis $\{\vec{u}_1, \ldots, \vec{u}_m\}$. Define the vectors*

$$\begin{aligned}
\vec{v}_1 &= \vec{u}_1 \\
\vec{v}_2 &= \vec{u}_2 - proj_{\vec{v}_1}\vec{u}_2 \\
\vec{v}_3 &= \vec{u}_3 - proj_{\vec{v}_1}\vec{u}_3 - proj_{\vec{v}_2}\vec{u}_3 \\
&\vdots \\
\vec{v}_m &= \vec{u}_m - proj_{\vec{v}_1}\vec{u}_m - proj_{\vec{v}_2}\vec{u}_m - \cdots - proj_{\vec{v}_{m-1}}\vec{u}_m.
\end{aligned}$$

Then, $\{\vec{v}_1, \ldots, \vec{v}_m\}$ *is an orthogonal basis for* V*.*

We'll talk through why this method produces an orthogonal basis together in class rather than reading through a formal proof (which can be lengthy and tedious). Note that we can normalize the orthogonal basis above to obtain an orthonormal basis. That is, if $\{\vec{v}_1, \ldots, \vec{v}_m\}$ is an orthogonal basis, then

$$\left\{\frac{\vec{v}_1}{\|\vec{v}_1\|}, \ldots, \frac{\vec{v}_n}{\|\vec{v}_n\|}\right\}$$

is an orthonormal basis.

We have the following.

Proposition 11.26. Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis for $\mathbb{R}^n$ and let A be the matrix with column vectors $\vec{v}_1, \ldots, \vec{v}_n$. Then $\mathcal{B}$ is orthonormal if and only if $A^{-1} = A^\top$.

Proof. Let A be the matrix with column vectors $\vec{v}_i$. Then $A^\top$ is the matrix with *rows* $\vec{v}_i$, and so we can observe that

$$A^\top A = \begin{pmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 & \cdots & \vec{v}_1 \cdot \vec{v}_n \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 & \cdots & \vec{v}_2 \cdot \vec{v}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{v}_n \cdot \vec{v}_1 & \vec{v}_n \cdot \vec{v}_2 & \cdots & \vec{v}_n \cdot \vec{v}_n \end{pmatrix}.$$

So, $A^\top A = I_n$ if and only if

$$\vec{v}_i \cdot \vec{v}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

which occurs precisely when $\{\vec{v}_1, \ldots, \vec{v}_n\}$ forms an orthonormal basis for $\mathbb{R}^n$. □

This gives rise to the following (annoying) definition.

Definition 11.27. We call a matrix A ORTHOGONAL if $A^{-1} = A^\top$.

Remark 11.28. This definition is annoying, because orthogonal matrices aren't just those matrices with *orthogonal* column vectors, but rather with *orthonormal* column vectors. I don't know why we don't just call them orthonormal matrices. My guess is because matrices with column vectors that are orthogonal, but not orthonormal, don't have many nice properties so they don't get their own name.

Just for fun and in case you're interested, Hadamard matrices are matrices with orthogonal (but not orthonormal) column vectors which only have entries equal to ± 1. It can be shown that for an $n \times n$ Hadamard matrix H we have $HH^\top = nI_n$ so that $H^{-1} = (1/n)H^\top$.

Remark 11.29. Note that Proposition 11.26 doesn't give a more convenient method of checking whether a set of vectors is orthonormal – in fact, it's often the case that computing the inverse of a matrix is *more* computationally expensive than just computing dot products – but it does tell us that if you know a matrix is orthonormal, its inverse is easy to compute.

We have the following observation.

Corollary 11.30. Let A be an orthogonal matrix. Then $\det(A) = \pm 1$.

Proof. Since $AA^\top = I_n$ then $\det(AA^\top) = \det(I_n) = 1$ but we know that

$$\det(AA^\top) = \det(A)\det(A) = (\det(A))^2.$$

So, $(\det(A))^2 = 1 \Rightarrow \det(A) = \pm 1$. □

Example 11.31. Consider the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}.$$

Observe that A has eigenvalues $\lambda_1 = -3$ and $\lambda_2 = 2$ and we have

$$E_{-3} = \operatorname{Span}(\vec{v}_1) \text{ and } E_2 = \operatorname{Span}(\vec{v}_2)$$

where

$$\vec{v}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \text{ and } \vec{v}_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Since $\|\vec{v}_1\| = \sqrt{5}$ and $\|\vec{v}_2\| = \sqrt{5}$ we can normalize these vectors to obtain a basis $\mathcal{B} = \{\vec{u}_1, \vec{u}_2\}$ for $\mathbb{R}^2$ where

$$\vec{u}_1 = \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{pmatrix} \text{ and } \vec{u}_2 = \begin{pmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}.$$

Furthermore, since $\vec{u}_1 \in E_{-3}$ and $\vec{u}_2 \in E_2$ then the basis $\mathcal{B}$ is **both** orthonormal (which is good for preserving distances and angles) and consists of eigenvectors (which is good for understanding the linear transformation T_A).

We have the following.

Definition 11.32. An $n \times n$ matrix A is ORTHOGONALLY DIAGONALIZABLE if there exists an orthogonal matrix Q and a diagonal matrix D so that $Q^\top AQ = D$. In this case, we have

$$Q = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}, \text{ and } D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

where $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors $\vec{v}_i$ for A and λ_i are the corresponding eigenvalues of A.

Note that, an $n \times n$ matrix A being orthogonally diagonalizable is equivalent to the existence of an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A. This is a particularly nice situation, since orthonormal bases preserve the dot product, and bases consisting of eigenvectors help us understand the linear transformation T_A. The following result completely characterizes when we're in this situation. Recall that a matrix A is SYMMETRIC when $A = A^\top$.

Theorem 11.33 (The Spectral Theorem)**.** *An $n \times n$ matrix A is orthogonally diagonalizable if and only if it is symmetric (that is $A = A^\top$).*

In Activity **??** we proved the forward direction of the Spectral theorem, and showed that the backward direction holds if we know that A is diagonalizable. Showing that a symmetric matrix is always diagonalizable takes a bit more work than we have time for, and typically is done with an inductive argument. so we're going to sweep this under the rug. Interested readers can find this proof in most linear algebra texts (here are some notes you could look at).

Note that we needed the following useful lemma.

Lemma 11.34. *Let $\vec{x}, \vec{y}$ be vectors in $\mathbb{R}^n$. Then, thinking of these as $n \times 1$ matrices, we have*

$$\vec{x} \cdot \vec{y} = \vec{x}^\top \vec{y}.$$

Proof. This follows by definition. Let

$$\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ and } \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Then

$$\vec{x}^\top \vec{y} = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + \cdots + x_n y_n,$$

as needed. □

Remark 11.35. The Spectral Theorem tells us that any symmetric matrix A has a SPECTRAL DECOMPOSITION. That is, we can write

$$A = QDQ^\top$$

for an orthogonal matrix Q and diagonal matrix D. On Activity **??** we saw that orthogonal matrices define linear transformations that act as rotations or reflections. Furthermore, note that multiplication by a diagonal matrix behaves like a "dilation" (that is, diagonal matrices stretch along each basis vector by some fixed amount). So, transformations defined by a symmetric matrix must look like some rotation/reflection, followed by a dilation (multiplying by a diagonal matrix), and then followed by the opposite rotation/reflection.

While the Spectral Theorem might seem like a special edge case, we can actually use this result to obtain a similar geometric understanding of *any* matrix transformation. We explore this in our final section of the semester.

11.5. The Singular Value Decomposition

In this section, we'll derive an important decomposition for *any* $m \times n$ matrix. Geometrically, this decomposition will show as that any linear transformation can be decomposed into a composition of three transformations: a rotation/reflection, followed by a dilation, followed by another rotation/reflection (not necessarily inverse to the original rotation/reflection).

This decomposition rests on the observation that for any $m \times n$ matrix A, the $n \times n$ matrix $A^\top A$ is symmetric, since

$$(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A = A.$$

We have the following key proposition.

Proposition 11.36. Let A be an $m \times n$ matrix, and let $\{\vec{v}_1, \ldots, \vec{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of $A^\top A$ (which exists by the Spectral Theorem). Then, $\{A\vec{v}_1, \ldots, A\vec{v}_n\}$ is an orthogonal subset of $\mathbb{R}^m$. Furthermore, if we reindex so that $A\vec{v}_1, \ldots, A\vec{v}_r$ are nonzero, and $A\vec{v}_{r+1} = \cdots = A\vec{v}_n = \vec{0}$, then

$$\{A\vec{v}_1, \ldots, A\vec{v}_r\}$$

forms an orthogonal basis for $\mathrm{Col}(A)$.

Proof. Suppose that the orthonormal eigenvectors $\vec{v}_i$ of $A^\top A$ have corresponding eigenvalue λ_i. Then, for any $i \neq j$ we have

$$\begin{aligned}
(A\vec{v}_i) \cdot (A\vec{v}_j) &= (A\vec{v}_i)^\top (A\vec{v}_j) \\
&= \vec{v}_i^\top A^\top A\vec{v}_j \\
&= \vec{v}_i^\top (\lambda_j \vec{v}_j) \\
&= \vec{v}_i \cdot (\lambda_j \vec{v}_j) \\
&= \lambda_j (\vec{v}_i \cdot \vec{v}_j) \\
&= 0,
\end{aligned}$$

where the final equality follows because $\vec{v}_i$ and $\vec{v}_j$ are perpendicular when $i \neq j$.

Next, if we reindex as written in the Theorem statement, we see that $A\vec{y} \in \mathrm{Col}(A)$ if and only if

$$A\vec{y} = A(x_1 + \cdots + x_n\vec{v}_n) = x_1 A\vec{v}_1 + \cdots + x_r A\vec{v}_r + \vec{0},$$

and so $\mathrm{Col}(A) = \mathrm{Span}(A\vec{v}_1, \ldots, A\vec{v}_r)$. To show this set forms a basis, we'll observe that any orthogonal set of nonzero vectors forms a linearly independent set. To see, this, suppose that $\{\vec{y}_1, \ldots, \vec{y}_r\}$ is an orthogonal set of nonzero vectors, and consider the vector equation

$$c_1\vec{y}_1 + \cdots + c_r\vec{y}_r = \vec{0}.$$

Taking the dot product on both sides of the equation with $\vec{y}_i$ gives

$$c_i\vec{y}_i \cdot \vec{y}_i = 0,$$

noting that $\vec{y}_i \cdot \vec{y}_j = 0$ when $i \neq j$. Since $\vec{y}_i \neq \vec{0}$ we get $\vec{y}_i \cdot \vec{y}_i = \|\vec{y}_i\|^2 \neq 0$ and so we must have $c_i = 0$ for all i. Hence, this set is linearly independent, as needed. □

Note that we can normalize the set $\{A\vec{v}_1, \ldots, A\vec{v}_n\}$ to obtain an orthonormal set by dividing through by $\|A\vec{v}_i\|$. We define the following.

Definition 11.37. Let A be an $m \times n$ matrix and $\vec{v}_1, \ldots, \vec{v}_n$ be an orthonormal basis for $\mathbb{R}^n$ of eigenvectors for $A^\top A$, as above. The SINGULAR VALUES of A are given by $\sigma_i := \|A\vec{v}_i\|$.

We have the following, which gives a more practical method to calculate the singular values of a matrix.

Proposition 11.38. Let A be an $m \times n$ matrix and $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A^\top A$. Then, $\lambda_i > 0$ and the singular values of A are given by $\sigma_i = \sqrt{\lambda_i}$.

Proof. We have

$$\begin{aligned}
\sigma_i^2 &= \|A\vec{v}_i\|^2 \\
&= (A\vec{v}_i) \cdot (A\vec{v}_i) \\
&= (A\vec{v}_i)^\top (A\vec{v}_i) \\
&= \vec{v}_i^\top A^\top A\vec{v}_i \\
&= \vec{v}_i^\top \lambda_i \vec{v}_i \\
&= \lambda_i \vec{v}_i \cdot \vec{v}_i \\
&= \lambda_i \|\vec{v}_i\|^2 \\
&= \lambda_i,
\end{aligned}$$

where the final equality follows because the $\vec{v}_i$ form an ortho**normal** set. Hence, $\sigma_i^2 = \lambda_i$ and so $\lambda_i > 0$ and $\sigma_i = \sqrt{\lambda_i}$. □

Now, if we let $\vec{u}_i = A\vec{v}_i/\sigma_i$ then

$$\{\vec{u}_1, \ldots, \vec{u}_n\}$$

forms an orthonormal basis for $\mathrm{Col}(A)$. We have the following result.

Theorem 11.39 (The Singular Value Decomposition)**.** *Let A be an $m \times n$ matrix. Then, there exists an orthogonal $m \times m$ matrix U, an orthogonal $n \times n$ matrix V and a "block diagonal" matrix Σ so that*

$$A = U\Sigma V^\top.$$

Before we prove this formally, let's look at example to see where this decomposition is coming from geometrically.

Example 11.40. Consider the 3×2 matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & -1 \end{pmatrix}.$$

Then

$$A^\top A = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}$$

which eigenvalues $\lambda_1 = 7$ and $\lambda_2 = 2$ with corresponding orthonormal eigenvectors

$$\vec{v}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \text{ and } \vec{v}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

So, A has singular values $\sigma_1 = \sqrt{7}$ and $\sigma_2 = \sqrt{2}$. From Proposition 11.38 we know that $\{A\vec{v}_1, A\vec{v}_2\}$ forms an orthogonal basis for $\mathrm{Col}(A)$, and so we obtain the following geometric understanding of the transformation $T_A : \mathbb{R}^2 \to \mathbb{R}^3$

Note in this example, $\mathrm{Col}(A)$ is a two dimensional subspace of $\mathbb{R}^3$. Observe that we could obtain this operation in several steps. First, rotate $\mathbb{R}^2$ to send the vectors $\vec{v}_i \mapsto \vec{e}_i$, which can be performed by the map with defining matrix

$$V^\top = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix}^{-1}$$

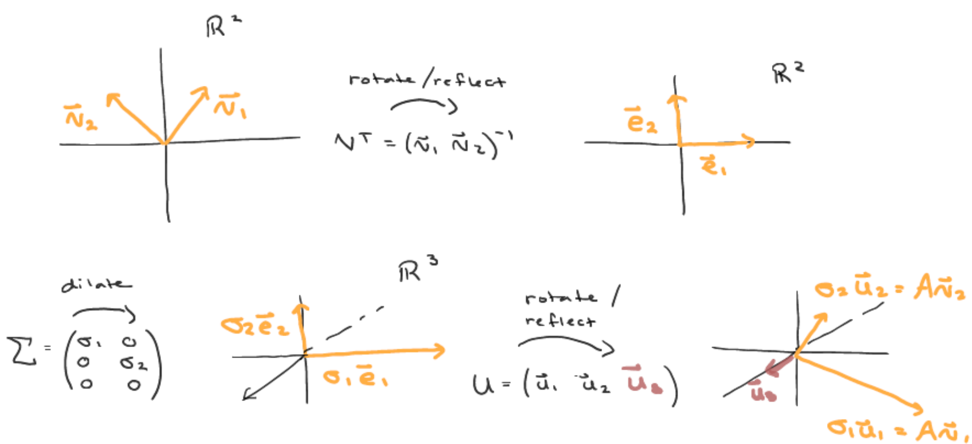
(noting that $V^\top = V^{-1}$ since the $\vec{v}_i$ are orthonormal). After that, we can dilate our map to stretch our standard basis vectors $\vec{e}_i \mapsto \sigma_i \vec{e}_i$, which can be performed by the map with defining matrix

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \\ 0 & 0 \end{pmatrix}.$$

Finally, we can rotate/reflect to send $\sigma_i \vec{e}_i \mapsto A\vec{v}_i$. Writing $\sigma_i \vec{u}_i = A\vec{v}_i$ and extending this to an orthonormal basis for $\mathbb{R}^3$ we can describe this by the map with defining matrix

$$U = \begin{pmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \end{pmatrix}.$$

This process is pictured below



This gives the desired decomposition $A = U\Sigma V^\top$. Now that we have a conceptual understanding for why this decomposition holds, let's look at the formal proof.

Proof of the Singular Value Decomposition. Suppose that A is an $m \times n$ matrix, and let $\{\vec{v}_1, \dots, \vec{v}_n\}$ be the eigenvalues of $A^\top A$ which form an orthonormal basis for $\mathbb{R}^n$. By Proposition 11.26 we know that

$$V = \begin{pmatrix} \vec{v}_1 & \cdots & \vec{v}_n \end{pmatrix}$$

is an orthogonal matrix. Furthermore, we have

$$AV = \begin{pmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{pmatrix}.$$

By Proposition 11.38 if we set $A\vec{v}_i = \sigma_i \vec{u}_i$, then (up to reindexing) we know that

$$\{\vec{u}_1, \dots, \vec{u}_r\}$$

forms an orthonormal basis for $\mathrm{Col}(A)$. By Gram-Schmidt, we can extend this to an orthonormal basis of $\mathbb{R}^m$, say $\{\vec{u}_1, \dots, \vec{u}_m\}$. Since we have $\vec{sigma}_i = 0$ for $i = r+1, \dots, n$ we get that $A\vec{v}_i = \sigma_i \vec{u}_i$ for all i. We have a few cases.

If $n = m$, then we have

$$AV = \begin{pmatrix} \sigma_1 \vec{u}_1 & \cdots & \sigma_n \vec{u}_n \end{pmatrix} = U\Sigma,$$

where

$$U = \begin{pmatrix} \vec{u}_1 & \cdots & \vec{u}_n \end{pmatrix} \text{ and } \Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_n).$$

Next, if $n > m$ then we have

$$AV = \begin{pmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{pmatrix} = U\Sigma$$

where

$$U = \begin{pmatrix} \vec{u}_1 & \cdots & \vec{u}_m \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m & 0 & \cdots & 0 \end{pmatrix}$$

where there are an additional $n - m$ columns of zeros at the end of the matrix Σ.

Finally, if $n < m$ then we have

$$AV = \begin{pmatrix} A\vec{v}_1 & \cdots & A\vec{v}_n \end{pmatrix} = U\Sigma$$

where

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

where there are an additional $n - m$ rows of zeros at the bottom of the matrix Σ.

□