MTH 261 Guided Notes

Damien Adams

June 5, 2024

Contents

1 Linear Systems

1.1 What is Linear Algebra?

Linear algebra is the branch of mathematics concerning linear equations such as

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

linear functions such as $(x_1, \ldots, x_n) \mapsto a_1x_1 + \cdots + a_nx_n$, and their representations in vector spaces and through matrices.

Wikipedia
Accessed March 20, 2020

If arithmetic is the foundational language of society, and if algebra is the foundational language of calculus, then linear algebra is the foundational language of STEM.

Each linear algebra course that is taught follows a very different flow. It is important to note that linear algebra is a complete branch of mathematics (as opposed to number theory or topology), and because it is so versatile and complete, any particular entry point or direction that the course tends to will generate a completely different course experience.

This particular version of the course will generate a very mathematics-based course. We will focus on the intricacies and interplay of theory and definition. **Computation** will simply be a vehicle to get to **interpretation**. We will basically learn one tool (only a slight exaggeration), but then we will spend a vast number of hours on

1. When to use this tool.
2. What to use this tool.
3. How to transform our information into something on which this tool can be used.
4. How to interpret the results of using this tool.

In this section, we will introduce a lot of terminology (in order to set a baseline for conversation), introduce matrices, and see how matrices play a role in solving linear systems.

1.2 Preliminary Definitions

Let's begin with a handful of definitions to put us all on the same page.

Definition

A **linear equation** is one of the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$. The numbers $a_1, a_2, \ldots, a_n, b \in \mathbb{C}$ (usually $\mathbb{R}$) are the coefficients.

Definition

A **linear system** is a collection of one or more simultaneous linear equations in the same variables.

Definition

A **solution** of the system is a list of numbers $(s_1, s_2, \ldots, s_n)$ that, when substituted into $x_1, x_2, \ldots, x_n$ respectively, satisfies all equations in the system simultaneously.

Definition

The set of all possible solutions of a system is the **solution set**.

Definition

Two systems that have the same solution set are called **equivalent** systems.

Now let's explore linear systems in two variables.

Example 1. In a linear system with two variables, what are the possible number of solutions for the system? Draw a graph to represent each possibility.

Proposition

A system of linear equations (in any number of equations and variables) has either

- No solutions,
- Exactly one solution, or
- Infinitely many solutions.

Definition

A system that has at least one solution is **consistent**.

Definition

A system that has no solutions is **inconsistent**.

We will spend a good amount of time determining either

- The solution set of a linear system,
- Whether a system is consistent or inconsistent (regardless of the actual solution), or
- If a system *is* consistent, then *how many* solutions does it have (regardless of the actual solution).

We can categorize linear systems as such.

Consistent	Inconsistent
Unique Solution	No Solutions
Infinitely Many Solutions	

1.3 Matrices

The object of study for most students who take a linear algebra class is the matrix. Any system can be compressed into its essential information using a rectangular array called a **matrix** (plural: **matrices**).

Definition

A rectangular array of entries (typically numbers) is called a **matrix**.

Consider the linear system

$$\begin{array}{rcrcrcr} 5x_1 & - & x_2 & + & 2x_3 & = & 7 \\ -2x_1 & + & 6x_2 & + & 9x_3 & = & 0 \\ -7x_1 & + & 5x_2 & - & 3x_3 & = & -7 \end{array}$$

This system can be represented by a matrix in several ways. Here are two ways.

$$\begin{bmatrix} 5 & -1 & 2 \\ -2 & 6 & 9 \\ -7 & 5 & -3 \end{bmatrix} \text{ or } \begin{bmatrix} 5 & -1 & 2 & 7 \\ -2 & 6 & 9 & 0 \\ -7 & 5 & -3 & -7 \end{bmatrix}$$

Definition

A matrix that represents the coefficients of variables in a linear system is called a **coefficient matrix** of the system.

Definition

A matrix that represents the coefficients of variables as well as the coefficient on the opposite side of an equals sign is called an **augmented matrix** of the system.

Definition

The **size** of a matrix tells how many rows and columns it has. The size of the matrix is given in $m \times n$ form, where m is the number of rows and n is the number of columns.

Example 2. What are the sizes of each of the matrices below?

$$\begin{bmatrix} 5 & -1 & 2 \\ -2 & 6 & 9 \\ -7 & 5 & -3 \end{bmatrix} \qquad \begin{bmatrix} 5 & -1 & 2 & 7 \\ -2 & 6 & 9 & 0 \\ -7 & 5 & -3 & -7 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 2 \\ 0 & -1 \end{bmatrix} \qquad \begin{bmatrix} \alpha & \beta & \gamma \\ \delta & \epsilon & \zeta \\ \eta & \theta & \iota \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \qquad \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix}$$

1.4 Solving a System of Linear Equations

In algebra classes, we are typically taught to solve systems of linear equations using three strategies: substitution, elimination (sometimes called the addition method), and replacement (often referred to as simplifying).

Example 3. Solve the following system *without* substitution.

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 \qquad\quad - 5x_3 &= 10 \end{aligned}$$

Continued...

1.5 Elementary Row Operations

The strategies used in the previous example are used thoroughly in a process we will come to know as *row reduction*. Three basic operations can be used on matrices to produce a different matrix (with particular properties).

Definition

The **Elementary Row Operations** are Scaling, Interchange, and Replacement.

- **Scaling** – Multiply all entries in a row by a nonzero constant.
- **Interchange** – Interchange two rows.
- **Replacement** – Replace one row by the sum of itself and a multiple of another row ("Add to one row a multiple of another row").

Definition

Two matrices are **row equivalent** if there is a sequence of elementary row operations that transforms one matrix into the other.

Note: The elementary row operations are *reversible*!

Proposition

If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

This means that instead of solving systems of equations using the algebra strategies you may have used before, *we* will solve linear systems using matrices!

2 Row Reduction & Echelon Forms

2.1 REF & RREF Forms

In the last section, we performed an algorithm to achieve a goal called "row reduction". In this section, we'll introduce more terminology, discuss the advantages and disadvantages of different types of row reduced forms, and we'll introduce a process called Gaussian Elimination.

Definition

In a matrix, the **leading entry** of a row is the leftmost nonzero entry in a nonzero row.

Example 1. Identify the leading entry of each of the rows of the following matrix or determine if no leading entry exists.

$$\begin{bmatrix} 0 & 0 & 3 & 2 & 0 \\ 0 & 1 & - & 0 & 1 \\ -9 & 3 & 0 & -4 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Definition

A matrix is in **echelon form** (or **row echelon form** or REF) if it has the following properties:

- All nonzero rows are above any zero rows.
- Each leading entry of a row is in a column to the right of the leading entry of the row above it.
- All entries in a column beneath a leading entry are zero.

Example 2. The matrix

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

is not in echelon form. Explain in as much detail as possible why.

On the other hand, $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is in echelon form.

Definition

A matrix is in **reduced echelon form** (or **row reduced echelon form** or RREF) if it has the following properties:

- The matrix is in echelon form.
- The leading entry in each nonzero row is 1.
- Each leading 1 is the only nonzero entry in its column.

Example 3. The matrix

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

is not in reduced echelon form. Explain in as much detail as possible why.

Example 4. The matrix

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

is not in reduced echelon form. Explain in as much detail as possible why.

On the other hand, $\begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is in reduced echelon form.

Definition

A matrix that is in echelon form is called an **echelon matrix** (or **row echelon matrix**). A matrix that is in reduced echelon form is called a **reduced echelon matrix** (or **row reduced echelon matrix**).

In particular,

$$\underset{\text{Row Echelon Matrix}}{\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}} \qquad \underset{\text{Row Reduced Echelon Matrix}}{\begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}}$$

Note: The word "echelon" is French for "ladder". An echelon matrix (or reduced echelon matrix) should look like a ladder (or staircase) of 0's. Here are some more examples in pictorial form. In these examples, $\blacksquare$ represents a leading term (so it is any nonzero number), and $*$ represents any real number (it can be zero or nonzero).

$$\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(Row) Echelon Matrix

$$\begin{bmatrix} 1 & 0 & * & 0 \\ 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(Row) Reduced Echelon Matrix

$$\begin{bmatrix} \blacksquare & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \blacksquare \end{bmatrix}$$

(Row) Echelon Matrix

$$\begin{bmatrix} 1 & * & * & * & 0 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

(Row) Reduced Echelon Matrix

Theorem

Each matrix is row equivalent to one and only one reduced echelon matrix.

Definition

If a matrix A is row equivalent to an echelon matrix U, then we call U an **echelon form of** A (or **row echelon form of** A). In this case, we write $U = \text{REF}(A)$.
If a matrix A is row equivalent to a reduced echelon matrix U, then we call U a **reduced echelon form of** A (or **row reduced echelon form of** A). In this case, we write $U = \text{RREF}(A)$.

It is important to note that $\text{REF}(A)$ is *not* unique, but $\text{RREF}(A)$ *is* unique. Moreover, we will find that if A is an augmented matrix representing a system of equations, then $\text{RREF}(A)$ will give us the solution set of the system.

We should now set a goal: **Given a matrix** A**, how can we produce** $\text{REF}(A)$ **(of which there are (typically) infinitely many possibilities) and** $\text{RREF}(A)$ **(of which there is only one possibility).**

2.2 Pivots

In the last many examples, it can be noted that the leading entries are always in the same position – we can then give this position a name.

Definition

A **pivot position** in a matrix A is a location in A that corresponds to a leading 1 in $\text{RREF}(A)$.
A **pivot column** is a column of A that contains a pivot position.
The numbers in the pivot positions of $\text{REF}(A)$ are called **pivots**.

Example 5. Let $A = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 3 & 5 & 7 & 9 \\ 5 & 7 & 9 & 11 \end{bmatrix}$. Where are the pivots of the matrix A, provided that $\text{RREF}(A)$ is described below?

$$\text{RREF}(A) = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

We will use pivots extensively in producing

$$A \longrightarrow \text{REF}(A) \longrightarrow \text{RREF}(A)$$

There are several ways to produce these forms, but we will be using one.

2.3 Gauss-Jordan Elimination

This process has six steps divided into two parts. The process of using the first four steps is called *Gaussian Elimination* and produces $A \longrightarrow \text{REF}(A)$. Including the last two steps is called *Gauss-Jordan Elimination* and produces $A \longrightarrow \text{RREF}(A)$.

Gauss-Jordan Elimination

This algorithm is used to produce $\text{RREF}(A)$ given a matrix A.

1. Leftmost nonzero column is a pivot column with pivot position at the top.
2. Choose a nonzero entry in the pivot column to be a pivot. Interchange rows so that the pivot is in the top row.
3. Use row operations to create all zeros below the pivot.
4. Ignore the pivot row and repeat for the remaining rows ad terminum.
5. Make all pivots 1.
6. Beginning with the rightmost pivot working upward and left, use row operations to zero all entries above each pivot.

Example 6. Use Gauss-Jordan Elimination to find $\text{RREF}(A)$ for the matrix A below.

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 5 & 7 \\ 1 & 4 & 7 & 10 \end{bmatrix}$$

2.4 Gauss-Jordan Elimination & Linear Systems

Now let's explore using Gauss-Jordan Elimination to solve a system of equations. When we solve, there are several ways to describe a solution set.

Definition

Variables corresponding with pivot columns are called **basic variables**.
Variables not represented by a pivot column are **free variables**.

We will be using these definitions in forming solution sets.

- If there is no solution, we write $\emptyset$.
- If there is a unique solution, we list that one solution in a set.
- If there are infinitely many solutions, we form a **general solution** by expressing the basic variables in terms of the free variables and denoting the free variables as "free". This solution set is also called a **parametric solution** where the parameters are the free variables.

Example 7. Consider the linear system below.

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 4 \\ x_1 + 3x_2 + 5x_3 &= 7 \\ x_1 + 4x_2 + 7x_3 &= 10 \end{aligned}$$

a. Convert the system into an augmented matrix.

b. Use Gauss-Jordan Elimination to find RREF(A).

c. Identify the basic and free variables of the system.

d. Find a general solution to the linear system.

e. Find three particular solutions to the linear system.

Continued...

Given a linear system, we will want to learn to "read" a matrix and interpret what it tells us about the linear system that produced it. For example, if A is an augmented matrix for a linear system, then

- If RREF(A) produces a false statement, such as "$0 = 1$", then the system is inconsistent.
- If RREF(A) produces an trivial statement, such as "$0 = 0$", then this tells us nothing about the system.

Theorem

A linear system is consistent if and only if the rightmost column of the augmented matrix is *not* a pivot column. This corresponds to having *no* row of the form $\begin{bmatrix} 0 & \cdots & 0 & \blacksquare \end{bmatrix}$, where $\blacksquare \neq 0$. If a linear system is consistent, then the solution set contains either a unique solution (no free variables) or infinitely many solutions (free variable(s)).

To conclude this section, we want to use Gauss-Jordan Elimination to solve linear systems in this way:

1. Create the augmented matrix A for the linear system.
2. Find RREF(A).
3. Deduce the unique or parametric solution by turning RREF(A) back into a system.

3 Vector Equations

3.1 Vectors

We begin with a working definition. This is not a definition that will stick, but we will use it for the first half of this course.

Definition

A **vector** is an ordered list of numbers.

There are two kinds of vectors: column vectors $\begin{bmatrix} a \\ b \end{bmatrix}$ and row vectors $\begin{bmatrix} a & b \end{bmatrix}$. We will make a conventional decision to use column vectors by default.

Definition

In a vector $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, the numbers $a_1, a_2, \ldots, a_n$ are called the **entries** of the vector.

Now that we have this new mathematical object, let's explore how vectors work.

Recall $\mathbb{R}$ is the set of all real numbers.

3.1.1 R2

Definition

$\mathbb{R}^2$ is the set of vectors with 2 entries. That is, $\mathbb{R}^2 = \left\{ \begin{bmatrix} a \\ b \end{bmatrix} \mid a, b \in \mathbb{R} \right\}$.

Definition

Two vectors in $\mathbb{R}^2$ are **equal** iff their corresponding entries are equal.

For example, $\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$ implies $a = 1, b = 2$. Also, $\begin{bmatrix} 1 \\ 2 \end{bmatrix} \neq \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

Definition

Given $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$, the **sum** of $\mathbf{u}$ and $\mathbf{v}$ is the vector $\mathbf{u} + \mathbf{v}$ obtained by adding the corresponding entries of $\mathbf{u}$ and $\mathbf{v}$.

Example 1. Add $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 3 \\ -5 \end{bmatrix}$.

Example 2. Can we add $\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$?

Definition

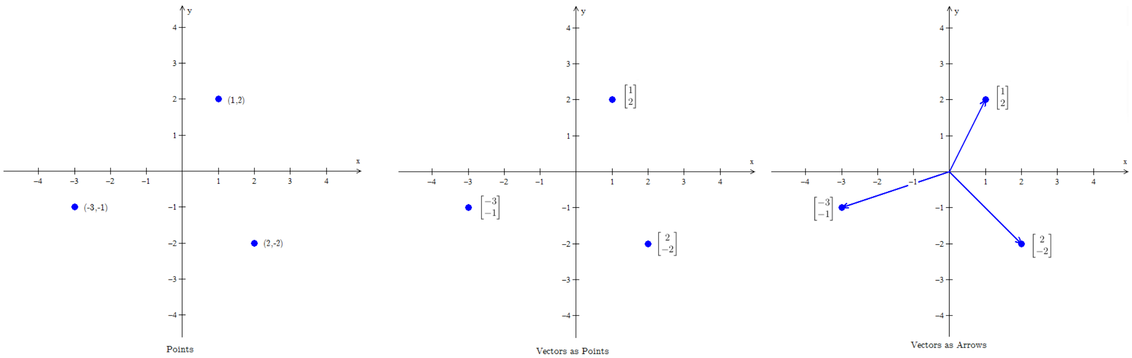
Given a vector $\mathbf{u}$ and a constant $c \in \mathbb{R}$, the **scalar multiple of u by** c is the vector $c\mathbf{u}$ obtained by multiplying each entry of $\mathbf{u}$ by c. The number c is called a **scalar**. Note that it is not bold.

Example 3. If $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, find $6\mathbf{u}$.

Because points in the Cartesian plane are made up of 2 ordered entries, we can *identify* a geometric point with a column vector.

$$(a, b) = \begin{bmatrix} a \\ b \end{bmatrix}$$

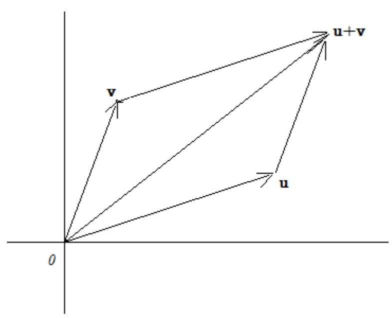
So $\mathbb{R}^2$ is the set of all points in the plane. But how do we represent them as images?



Convention: We again adopt a convention that we will use *vectors as arrows* to draw vectors. Moreover, all of our vectors will begin with the tail at the origin and the tip at (a, b).

Parallelogram Rule for Addition

If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$, then $\mathbf{u} + \mathbf{v}$ is the fourth vertex of the parallelogram with vertices at the origin, $\mathbf{u}$, and $\mathbf{v}$.



Example 4. Let $\mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$. Graph $\mathbf{u}, \mathbf{v}, \mathbf{u}+\mathbf{v}$. On a second set of axes, graph $\mathbf{v}, 2\mathbf{v}, -\frac{1}{3}\mathbf{v}$.

We have now defined equality, sum, and scalar multiplication for vectors in $\mathbb{R}^2$. To the difference of $\mathbf{u}$ and $\mathbf{v}$ is simply $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$. Moreover, notice that we have no definition for multiplication of two vectors (yet), nor do we have a definition for division.

3.1.2 Rn

Definition

We define $\mathbb{R}^n$ to be the collection of ordered n-tuples of real numbers, usually written $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$. Moreover, the **zero vector** is the vector whose entries are all 0, written $\mathbf{0}$.

It turns out, equality, sum, difference, and scalar multiplication are defined just as in $\mathbb{R}^2$.

Properties of $\mathbb{R}^n$

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

(ii) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(iv) $\mathbf{u} - \mathbf{u} = \mathbf{u} + (-1\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$

(v) $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$

(vi) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vii) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

The proofs are not too difficult, and I encourage them as a exercises.

3.2 Linear Combinations and Span

One of the more important terms of this course is the notion of a linear combination. This is really a combination of sum and scalar multiplication of vectors. The term comes from the fact that we are *combining* vectors in a *linear* way (remember the definition of a linear equation?).

Definition

The **linear combination** of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{R}^n$ with **weights** $c_1, c_2, \ldots, c_p \in \mathbb{R}$ is the vector $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$.

Example 5. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$. Identify which of the following are linear combinations of $\mathbf{u}$ and $\mathbf{v}$. In these cases, identify the weights in the linear combination.

- $\mathbf{v}_1 + \mathbf{v}_2$
- $\mathbf{v}_1 - \mathbf{v}_2$
- $2\mathbf{v}_1 + 3\mathbf{v}_2$
- $\mathbf{0}$
- 0
- $\mathbf{v}_1$
- $-\mathbf{v}_1 - \pi\mathbf{v}_2$
- $\sqrt{2}\mathbf{v}_1 + \mathbf{v}_2$
- $\sqrt{2\mathbf{v}_1} + \mathbf{v}_2$

There are a few examples that I consider to be *blueprint prompts*, and this next one is one of them. We will encounter prompts that are essentially this one throughout this course.

Example 6. If $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$, determine if $\mathbf{b}$ can be written as a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. (That is, does the equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$ have a solution?)

This workthrough provides a new way to express a matrix – a row of column vectors. Moreover, we notice a vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ has the same solution set as the linear system whose augmented matrix is $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}$. In particular, $\mathbf{b}$ can be generated by a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ iff there exists a solution to the linear system corresponding to A.

3.3 Span

Now that we have linear combinations, it is important to know what vectors may or may not be a linear combination of a particular set of vectors. This leads us to one of the biggest definitions of this entire course – the span of a set of vectors.

Definition

If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{R}^n$, then we define the **span** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ to be the set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$, denoted $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$.

Note $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} = \{c_1\mathbf{v}_1 + c_2\mathbf{v} + 2 + \cdots + c_p\mathbf{v}_p \mid c_1, c_2, \ldots, c_p \in \mathbb{R}\}$. It is also important to note that $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is a subset of $\mathbb{R}^n$.

To connect this back to the previous section, we have the following proposition.

Proposition

If $\mathbf{b} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, then $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots x_p\mathbf{v}_p = \mathbf{b}$ has a solution.

Example 7. Is $\mathbf{0} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$?

3.3.1 Envisioning Span

We finish this section with a few questions to ponder, and the answers are not obvious.

Consider $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$.

- What does $\operatorname{span}\{\mathbf{u}\}$ look like?
- What does $\operatorname{span}\{\mathbf{u}, \mathbf{v}\}$ look like?

4 Matrix Equations

4.1 Matrix Product

We now have linear combinations, so we will revisit previous topics to connect it all.

Definition

If A is an $m \times n$ matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and $\mathbf{x} \in \mathbb{R}^n$, then the **product** $A\mathbf{x}$ is the linear combination of the columns of A with weights being the corresponding entries of $\mathbf{x}$. That is,

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

Example 1. Multiply $\begin{bmatrix} 1 & -1 & 2 & 4 \\ 6 & -2 & 3 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 2 \\ 3 \end{bmatrix}$

Example 2. If $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \in \mathbb{R}^m$, write the linear combination $-\mathbf{v}_1 + \mathbf{v}_2 - 3\mathbf{v}_3 + 6\mathbf{v}_4$ as the product of a matrix and a vector.

4.2 Matrix Equation

In the first section, we saw a system of equations

$$\begin{array}{rcrcrcr} x_1 & + & 2x_2 & - & x_3 & = & 4 \\ & - & 5x_2 & + & 3x_3 & = & 1 \end{array}$$

Last section, we saw this was equivalent to the vector equation

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -5 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

Transforming the left side, this is now equivalent to the matrix equation

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$$

This last equation can be abbreviated in the form $A\mathbf{x} = \mathbf{b}$.

Definition

An equation of the form $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m$, is a **matrix equation**.

That is,

$$\text{Linear System} \longleftrightarrow x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots x_n\mathbf{v}_n = \mathbf{b} \longleftrightarrow A\mathbf{x} = \mathbf{b}$$

Notice A is a coefficient matrix for the linear system and *any* linear system can be written as a linear combination or matrix equation!

Theorem

If A is an $m \times n$ matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and $b \in \mathbb{R}^m$, then

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

has the same solution set as

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

which has the same solution set as the system whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}$$

Linear systems can now be viewed in three ways, all of which reduce to finding RREF(A)! This means that we solve matrix equations in the same way as vector equations, and those are the same as solving linear systems. Moreover, all of this connects to the concept of span and linear combinations.

Proposition

The matrix equation $A\mathbf{x} = \mathbf{b}$ has a solution iff $\mathbf{b}$ is a linear combination of the columns of A.

This helps us determine when $\mathbf{b} \in \operatorname{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\} \equiv A\mathbf{x} = \mathbf{b}$ is consistent.

Question: Is $A\mathbf{x} = \mathbf{b}$ *always* consistent?

Example 3. Let $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. Is $A\mathbf{x} = \mathbf{b}$ consistent $\forall b_1, b_2, b_3 \in \mathbb{R}$?

Hint: $\text{RREF}\left(\begin{bmatrix} 1 & 3 & 4 & b_1 \\ -4 & 2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 0 & 0 & b_1 - \frac{1}{2}b_2 + b_3 \end{bmatrix}$

Theorem

Let A be an $m \times n$ matrix. The following are equivalent.

(i) $A\mathbf{x} = \mathbf{b}$ is consistent $\forall \mathbf{b} \in \mathbb{R}^m$.

(ii) A has a pivot in every row.

(iii) Each $\mathbf{b} \in \mathbb{R}^m$ is a linear combination of the columns of A.

(iv) The columns of A span $\mathbb{R}^m$.

Note this theorem is about A, not the augmented matrix $\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$.

Proof: (i) $\Rightarrow$ (ii). Suppose (for a contradiction) that RREF A has a row of 0s. Augment on a column with a nonzero entry in the last entry. "Unwinding" the row operations back to A. Then the augmented column will not allow a consistent system. This contradicts the assumption that RREF A has a row of 0s.

(ii) $\Rightarrow$ (iii). A leading 1 in each row allows for a solution for any $\mathbf{b} \in \mathbb{R}^m$. That is, $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix}$ has solution $\mathbf{x} \in \mathbb{R}^n$. By the definition of a matrix equation, $A\mathbf{x} = \mathbf{b}$ leads to $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$. Hence, any $\mathbf{b}$ is a linear combination of the columns of A.

(iii) $\Rightarrow$ (iv). Trivial from the definition of span.

(iv) $\Rightarrow$ (i). Consider $\begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{bmatrix}\mathbf{x} = \mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^m$. Since the columns of A span $\mathbb{R}^m$, there esit $\mathbf{x} \in \mathbb{R}^n$ such that $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$. Hence, $A\mathbf{x} = \mathbf{b}$ is consistent. □

4.3 The Identity Matrix

Example 4. Multiply $\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}a\\ b\\ c\end{bmatrix}$

Definition

A square matrix with 1's on the main diagonal and 0's elsewhere is called an **identity matrix**, denoted I. Typically, I_n is the $n \times n$ identity matrix.

This matrix is an incredibly important matrix that will continue to appear throughout this course.

Proposition

For all$\mathbf{x} \in \mathbb{R}^n$, $I_n\mathbf{x} = \mathbf{x}$.

Theorem

If A is an $m \times n$ matrix, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n, c \in \mathbb{R}$, then

- $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$, and
- $A(c\mathbf{u}) = c(A\mathbf{u})$.

5 Solution Sets

5.1 Homogeneous Equations

Now that we have vectors, let's revisit linear systems.

Definition

A linear system is **homogeneous** if it can be written as $A\mathbf{x} = \mathbf{0}$, where A is $m \times n$, and $\mathbf{0} \in \mathbb{R}^m$ is the zero vector.

Last time, we found that $A\mathbf{x} = \mathbf{b}$ is not always consistent. On the other hand, $A\mathbf{x} = \mathbf{0}$ *always* has the solution $\mathbf{x} = \mathbf{0} \in \mathbb{R}^n$.

Definition

The solution $\mathbf{x} = \mathbf{0}$ to $A\mathbf{x} = \mathbf{0}$ is called the **trivial solution**.

With $A\mathbf{x} = \mathbf{b}$, we always ask ourselves if this equation is consistent or inconsistent. If it is consistent, we ask if the solution is unique or not.

With $A\mathbf{x} = \mathbf{0}$, we have no need to ask about consistency. Since $A\mathbf{x} = \mathbf{0}$ always has the trivial solution, we ask if it has a nontrivial solution as well. It turns out, we can determine this by looking at the variables.

Recall Existence and Uniqueness: If a linear system is consistent, then the solution set contains either

(i) A unique solution when there are no free variables, or

(ii) Infinitely many solutions when there is at least one free variable.

Proposition

The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution iff the equation has a free variable.

Example 1. Determine if the homogeneous system has a nontrivial solution. Describe the solution set.

$$\begin{array}{llll} 3x_1 & +5x_2 & -4x_3 & =0 \\ -3x_1 & -2x_2 & +4x_3 & =0 \\ 6x_1 & +x_2 & -8x_3 & =0 \end{array}.$$

Hint: $\text{RREF}(A) = \begin{bmatrix} 1 & 0 & -\frac{4}{3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

5.2 Geometry and Algebra of Solutions to $A\mathbf{x} = \mathbf{0}$

Notice the solution to $A\mathbf{x} = \mathbf{0}$ could be expressed as span$\{\mathbf{v}\}$.

Generally, the solution to the homogeneous equation $A\mathbf{x} = \mathbf{0}$ is span$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ for some $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. Geometrically, this means...

- If there are no free variables, then the trivial solution is the only solution, and span$\{\mathbf{0}\} = \{\mathbf{0}\}$ is the solution set (just a point).
- If there is one free variable, then the solution set is span$\{\mathbf{v}_1\}$ and will be a line through the origin.
- If there are two free variables, then the solution set is span$\{\mathbf{v}_1, \mathbf{v}_2\}$ and will be a plane through the origin.
- Etc.

This is the geometry. What about the algebra?

Definition

A **parametric vector equation** is an equation of the form $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$, where $c_i \in \mathbb{R}$. When a parametric vector equation represents a solution set, it is in **parametric vector form**.

Example 2. Describe all solutions to $A\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 3 & -5 \\ 1 & 4 & -8 \\ -3 & -7 & 9 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 4 \\ 7 \\ -6 \end{bmatrix}$$

Hint: $\text{RREF}(\begin{bmatrix} A & \mathbf{b} \end{bmatrix}) = \begin{bmatrix} 1 & 0 & 4 & -5 \\ 0 & 1 & -3 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

So $\mathbf{x} = \mathbf{p} + t\mathbf{v}$, where $\mathbf{p} = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -4 \\ 3 \\ 1 \end{bmatrix}, t = x_3 \in \mathbb{R}$.

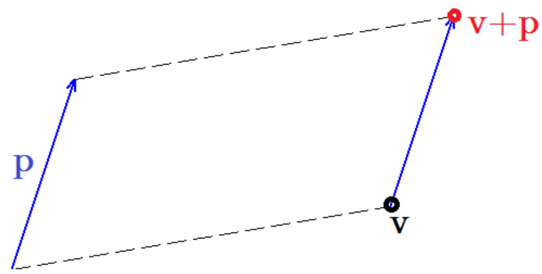
In the example, $\begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix}$ is a **particular solution**, and $\begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -4 \\ 3 \\ 1 \end{bmatrix}$ is the **general solution**, where each choice of the variable t produces a different particular solution.

5.3 Geometry of Solutions to $A\mathbf{x} = \mathbf{b}$

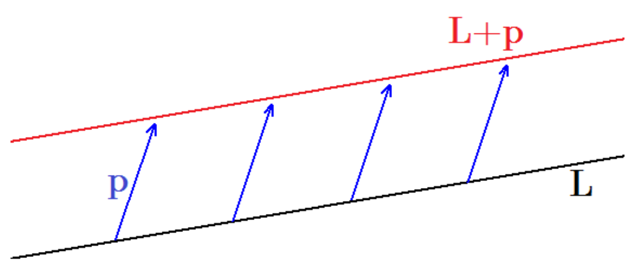
In $\mathbb{R}^2$, we graph $y = mx + b$ by graphing $y = mx$ and then translating that vertically by b. That is, the graph of $y = mx + b$ is the graph of $y = mx$ shifted a bit.

Because geometrically, vector addition is a translation, the solutions of $A\mathbf{x} = \mathbf{b}$ are the solutions of $A\mathbf{x} = \mathbf{0}$ shifted a bit.

For example, adding $\mathbf{p} + \mathbf{v}$ moves $\mathbf{v}$ in a direction parallel to the line through $\mathbf{p}$ and $\mathbf{0}$. We say $\mathbf{v}$ is **translated by p** to $\mathbf{v} + \mathbf{p}$.

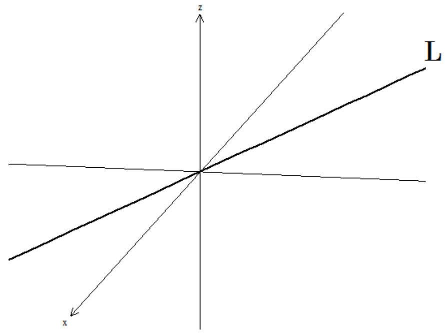


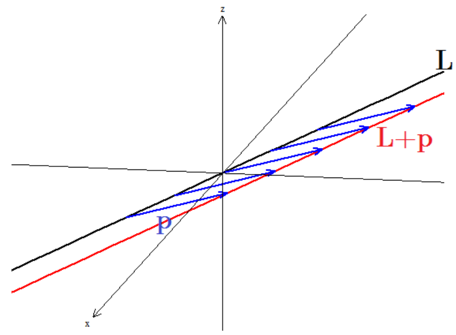
If L is the line through $\mathbf{0}$ and $\mathbf{v}$, adding $\mathbf{p}$ to each point on L translates every point to the new line $L + \mathbf{p}$.



In the previous example, $\mathbf{x} = t \begin{bmatrix} -4 \\ 3 \\ 1 \end{bmatrix}$ is the general solution to $A\mathbf{x} = \mathbf{0}$. Geometrically, this represents the line through $\mathbf{0}$ and $\begin{bmatrix} -4 \\ 3 \\ 1 \end{bmatrix}$.

On the other hand, $\mathbf{x} = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -4 \\ 3 \\ 1 \end{bmatrix}$ is that line translated by $\begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix}$.





This relationship is outlined in this theorem.

Theorem

Suppose $A\mathbf{x} = \mathbf{b}$ is consistent for some $\mathbf{b}$, and let $\mathbf{p}$ be the solution. If $\mathbf{v}_h$ is any solution to the homogeneous equation $A\mathbf{x} = \mathbf{0}$, then the solution to $A\mathbf{x} = \mathbf{b}$ is all vectors of the form $\mathbf{w} = \mathbf{p} + t\mathbf{v}_h$, where $t \in \mathbb{R}$.

Algorithm for Solving a Matrix Equation

To solve $A\mathbf{x} = \mathbf{b}$, we follow these steps.

1. Compute RREF($\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$).
2. Express basic variables in terms of the free variables.
3. Write $\mathbf{x}$ in parametric vector form.
4. Decompose $\mathbf{x}$ into a linear combination of vectors with weights as the free variables.

6 Linear Independence

6.1 Definition

Any homogeneous equation $A\mathbf{x} = \mathbf{0}$ can be turned into a vector equation. For example.

$$\begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \longrightarrow x_1 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 3 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Notice that the trivial solution always works, but is it unique?

Definition

An indexed set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, $\mathbf{v}_i \in \mathbb{R}^n$, is **linearly independent** if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution. If the trivial solution is not unique, we call the set **linearly dependent**.
If a set of indexed vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, $\mathbf{v}_i \in \mathbb{R}^n$, is linearly dependent, then there exists some weights $c_1, c_2, \ldots, c_p \in \mathbb{R}$ *not all zero* that satisfy the equation. The equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

is called a **linear dependence relation**.

Example 1. Are the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 3 \\ 5 \\ 3 \end{bmatrix}$ linearly independent? If not, find the linear dependence relation.

Hint: RREF $\left(\begin{bmatrix} 1 & 0 & 3 & 0 \\ 2 & 1 & 5 & 0 \\ 1 & 0 & 3 & 0 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

6.2 Linear Dependence Without a Linear Dependence Relation

As we explore linear dependence and linear independence, we can always solve the homogeneous equation and interpret our results. However, one of the recurring themes of this course is to identify conclusions without doing *all* of the computational work. So let's start exploring what linear dependence means without finding a linear dependence relation.

Example 2. Is $\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \right\}$ linearly independent or linearly dependent?

Linear Dependence Test for a Two-Vector Set

Consider $\{\mathbf{v}_1, \mathbf{v}_2\}$. If $\mathbf{v}_2 = c\mathbf{v}_1$ for some $c \in \mathbb{R}, c \neq 0$, then $\mathbf{v}_1, \mathbf{v}_2$ are linearly dependent.

Example 3. Is $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \end{bmatrix} \right\}$ linearly independent?

Linear Dependence Test for a Set of Several Vectors

A set of vectors with more vectors than there are entries in those vectors is linearly dependent.

Example 4. When is the set of a single vector independent?

Proposition

The set $\{\mathbf{0}\}$ is linearly dependent.

Corollary

Any set containing $\mathbf{0}$ is linearly dependent.

Characterization of Linearly Dependent Sets

An indexed set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, $p > 1$, is linearly dependent iff at least one vector in S is a linear combination of the others. Moreover, if S is linearly dependent and $\mathbf{v}_1 \neq \mathbf{0}$, then some $\mathbf{v}_i$, $i > 1$, is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{i-1}$.

Proof: First, assume some $\mathbf{v}_i \in S$ is a linear combination of the other vectors, then

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{i-1}\mathbf{v}_{i-1} + c_{i+1}\mathbf{v}_{i+1} + \cdots + c_p\mathbf{v}_p = c_i\mathbf{v}_i$$

where $c_1, c_2, \ldots, c_p$ are not all zero.

Then $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{i-1}\mathbf{v}_{i-1} - c_i\mathbf{v}_i + c_{i+1}\mathbf{v}_{i+1} + \cdots + c_p\mathbf{v}_p = \mathbf{0}$, so S is linearly dependent.

On the other hand, assume S is linearly dependent. If $\mathbf{v}_1 = \mathbf{0}$, then it is a trivial linear combination of the other vectors. Suppose $\mathbf{v}_1 \neq \mathbf{0}$. Then there exist $c_1, c_2, \ldots, c_p \in \mathbb{R}$ not all zero such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

Let i be the largest subscript such that $c_i\mathbf{v}_i \neq \mathbf{0}$. Since $i > 1$,

$$c_1\mathbf{v}_1 + \cdots + c_i\mathbf{v}_i + 0\mathbf{v}_{i+1} + \cdots + 0\mathbf{v}_p = \mathbf{0}$$

$$c_i\mathbf{v}_i = -c_1\mathbf{v}_1 - \cdots - c_{i-1}\mathbf{v}_{i-1}$$

$$\mathbf{v}_i = -\frac{c_1}{c_i}\mathbf{v}_1 - \cdots - \frac{c_{i-1}}{c_i}\mathbf{v}_{i-1}$$

$\square$

Example 5. Let $\mathbf{u} = \begin{bmatrix} 2 \\ 0 \\ 5 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$. Describe span$\{\mathbf{u}, \mathbf{v}\}$.

7 Linear Transformations

7.1 Transformations

Previously, we have taken a linear combination of vectors and turned it into a matrix product. However, a matrix product does not necessarily have to come from a linear combination of vectors. A matrix product, such as $A\mathbf{x}$, produces a vector, so a matrix can "act" on a vector by multiplication.

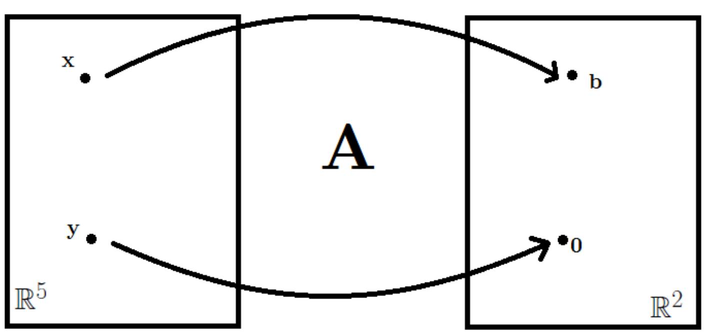
Consider $A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ -2 & -1 & 1 & 4 & 8 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 7 \\ -6 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Notice

$$A\mathbf{x} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ -2 & -1 & 1 & 4 & 8 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 35 \end{bmatrix} = \mathbf{b}$$

and

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

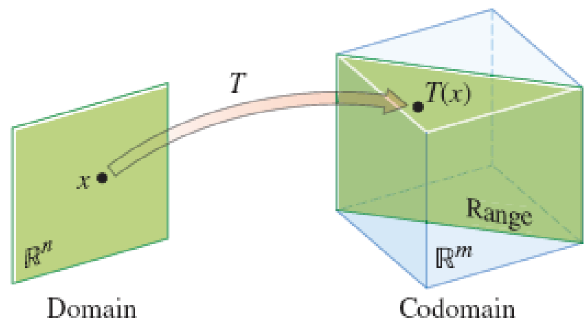
Multiplying a vector by a matrix fundamentally changes the vector. It may change both the entries of the vector as well as the size of the vector. So we say A *transforms* $\mathbf{x}$ into $\mathbf{b}$, and A transforms $\mathbf{y}$ into $\mathbf{0}$.



Solving $A\mathbf{x} = \mathbf{b}$ amounts to finding all $\mathbf{x} \in \mathbb{R}^5$ such that $A\mathbf{x} = \mathbf{b} \in \mathbb{R}^2$. But with this new language, we can instead ask, "Find all vectors in $\mathbb{R}^5$ that are transformed into $\mathbf{b} \in \mathbb{R}^2$ by the action of multiplying by A."

Definition

A **transformation** (or **mapping** or **function**) $T : \mathbb{R}^n \to \mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x} \in \mathbb{R}^n$ a vector $T(\mathbf{x}) \in \mathbb{R}^m$. The set $\mathbb{R}^n$ is the **domain** of T, and $\mathbb{R}^m$ is the **codomain** of T. For each $\mathbf{x} \in \mathbb{R}^n$, the vector $T(\mathbf{x}) \in \mathbb{R}^m$ is the **image of x** (under the action of T). The set of all images of T is the **range** of T.



In this section, $T(\mathbf{x}) = A\mathbf{x}$, where A is an $m \times n$ matrix. Written another way, $\mathbf{x} \mapsto A\mathbf{x}$. The domain of T is $\mathbb{R}^n$ (where n is the number of columns of A), and the codomain is $\mathbb{R}^m$ (where m is the number of entries in each column). The range is the set of all linear combinations of the columns of A, since each image $T(\mathbf{x}) = A\mathbf{x}$ (linear combination of the columns of A).

Example 1. Let $A = \begin{bmatrix} 1 & 3 \\ 9 & -1 \\ -2 & 5 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ -29 \\ 16 \end{bmatrix}$. Define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$.

(a) Find the image of $\mathbf{u}$ under T.

(b) Find $\mathbf{x} \in \mathbb{R}^2$ whose image under T is $\mathbf{v}$. Is it unique?

Note that $\begin{bmatrix} 1 & 3 & 3 \\ 9 & -1 & -29 \\ -2 & 5 & 16 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$.

7.2 Describing a Transformation Given Algebraically

We are learning that a matrix can transform something, so how do we describe what that transformation is? In reality, we typically make an observation and try to describe that with words; we then try to take those words and describe that with formulas and algebra. Let's try to see if we can come up with some words based on these formulas.

Example 2. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. What does the transformation $\mathbf{x} \mapsto A\mathbf{x}$ do to points in $\mathbb{R}^3$?

Example 3. Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ and $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(\mathbf{x}) = A\mathbf{x}$. Consider an Instagram image in the x_1x_2-plane with vertices $\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \end{bmatrix}$. What does T do to this image?

This is type of transformation is called a **shear transform**.

7.3 Linear Transformations

Just like functions, special transformations have particular names. Previously, $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$ and $A(c\mathbf{u}) = c(A\mathbf{u})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Similarly,

sheep

sheared sheep

Figure 1: From Lay, Lay, McDonald's *Linear Algebra and its Applications, 5th Edition*

Definition

A transformation T is a **linear transformation** if

(i) For all $\mathbf{u}, \mathbf{v}$ in the domain of T, $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$, and

(ii) For all $c \in \mathbb{R}$ and $\forall \mathbf{u}$ in the domain of T, $T(c\mathbf{u}) = cT(\mathbf{u})$.

Proposition

Every matrix transformation is a linear transformation.

It turns out, with these properties, we can conclude the following.

Proposition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and $c, d \in \mathbb{R}$,

(i) $T(\mathbf{0}) = \mathbf{0}$, and

(ii) $T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$

Proof: Since T is linear, we can use the scalar properties of linear transformations to do the following

$$T(\mathbf{0}) = T(0\mathbf{u}) = 0T(\mathbf{u}) = \mathbf{0}$$

Moreover, we can use both properties of linear transformations to show that

$$T(c\mathbf{u} + d\mathbf{v}) = T(c\mathbf{u}) + T(d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$$

More efficiently, to show that a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, we only need to show that it satisfies $T(c\mathbf{u}+d\mathbf{v}) = cT(\mathbf{u})+dT(\mathbf{v})$ for all scalars c, d and vectors $\mathbf{u}, \mathbf{v}$ in the domain of T. That said, showing that $T(\mathbf{0}) = \mathbf{0}$ is *not* sufficient to conclude linearity.

Repeating this, we can generalize that linear transformations have the property that

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_pT(\mathbf{v}_p)$$

8 The Matrix of a Linear Transformation

8.1 The Columns of an Identity Matrix

Suppose that we have a transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. In the last section, we showed that if the rule to evaluate T is $T(\mathbf{x}) = A\mathbf{x}$, then T must be a linear transformation. In this section, we aim to show that if T is a linear transformation, then there must be some matrix A such that $T(\mathbf{x}) = A\mathbf{x}$.

So the question is, "How do we find A?" It turns out, A is uniquely determined by what T does to the columns of I_n.

Definition

Let I be the $n \times n$ identity matrix. We define the vector $\mathbf{e}_j$ as the jth column of I. That is, $\mathbf{e}_j$ is the $\mathbb{R}^n$ vector whose jth entry is 1 while all other entries are 0.

Example 1. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^4$ is linear such that

$$T(\mathbf{e}_1) = \begin{bmatrix} 2 \\ -4 \\ 5 \\ 7 \end{bmatrix} \quad \text{and} \quad T(\mathbf{e}_2) = \begin{bmatrix} -1 \\ 3 \\ 0 \\ -8 \end{bmatrix}.$$

Find a formula for $T(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^2$.

The only assumption we made was what $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ were. We determined A from only that information. Since $T(\mathbf{x})$ represents a linear combination of vectors, we can create a matrix product.

$$T(\mathbf{x}) = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\mathbf{x}$$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that for all $\mathbf{x} \in \mathbb{R}^n$, $T(\mathbf{x}) = A\mathbf{x}$. Moreover, A is the $m \times n$ matrix whose jth column is the vector $T(\mathbf{e}_j)$, where e_j is the jth column of I_n. That is, $A = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$.

Proof: Write $\mathbf{x} = I_n\mathbf{x} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix} \mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$.

By linearity,

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) \\ &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n) \\ &= \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x} \end{aligned}$$

So A exists.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation such that $T(\mathbf{x}) = B\mathbf{x}$ for some $m \times n$ matrix B. Let A_i be the ith column of A and B_i be the ith column of B.

Then $A_i = T(\mathbf{e}_i) = B\mathbf{e}_i = B_i$ for $1 \leq i \leq n$. It follows that each column of B is equivalent to the corresponding column of A, so $B = A$. Therefore A is unique. □

8.2 Standard Matrix for a Linear Transformation

The matrix that we have established in the previous theorem has a name.

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let A be the matrix such that $T(\mathbf{x}) = A\mathbf{x}$. Then

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

is known as the **Standard Matrix for the Linear Transformation** T.

Example 2. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates each point in $\mathbb{R}^2$ about the origin through an angle of θ (CCW is positive). Assuming the transformation is linear, find the standard matrix A for this transformation.

Example 3. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that

1. First rotates points by $\frac{\pi}{6}$ about the origin, then
2. Second reflects points about the line $x_2 = x_1$, and
3. Lastly dilates points by a factor of 8.

Find the standard matrix for the linear transformation T.

8.3 Onto and One-to-One

There are several other adjectives that belong to transformations. We will focus on two more – one-to-one and onto transformations.

Definition

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **onto** (**surjective**) $\mathbb{R}^m$ if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at least one $\mathbf{x} \in \mathbb{R}^n$.

Definition

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is **one-to-one** (**injective**) if each $\mathbf{b} \in \mathbb{R}^m$ is the image of at most one $\mathbf{x} \in \mathbb{R}^n$. That is, a mapping T is one-to-one if $T(\mathbf{x}_1) = T(\mathbf{x}_2)$ implies $\mathbf{x}_1 = \mathbf{x}_2$.

Notice how parallel each of these definitions is. Moreover, these definitions have everything to do with the existence and uniqueness of preimages.

- Existence: "Does each $\mathbf{b} \in \mathbb{R}^m$ have a **pre-image**?" If T is onto, then "Yes".
- Uniqueness: "Is each solution to $T(\mathbf{x}) = \mathbf{b}$ unique?" If T is one-to-one, then "Yes".

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is one-to-one iff $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

Proof: Since T is linear, $T(\mathbf{0}) = \mathbf{0}$. We will show that both statements are true or both are false.

If T is one-to-one, then $T(\mathbf{x}) = \mathbf{0}$ has only one solution. Since $T(\mathbf{0}) = \mathbf{0}$, the only solution is trivial.

If T is not one-to-one, then there is at least one $\mathbf{b} \in \mathbb{R}^m$ with two different pre-images in $\mathbb{R}^n$, say $\mathbf{u}, \mathbf{v}$. So $T(\mathbf{u}) = \mathbf{b}$ and $T(\mathbf{v}) = \mathbf{b}$. Since T is linear,

$$T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

Since $\mathbf{u} \neq \mathbf{v}$, $\mathbf{u} - \mathbf{v} \neq 0$. Thus, $T(\mathbf{x}) = \mathbf{0}$ has at least two solutions.

It follows that either both are true or both are false. □

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A. Then

- T is onto iff the columns of A span $\mathbb{R}^m$.
- T is one-to-one iff the columns of A are linearly independent.

Proof:

- The columns of A span $\mathbb{R}^m \iff \forall \mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ is consistent $\iff \forall \mathbf{b}$, $T(\mathbf{x}) = \mathbf{b}$ has at least one solution $\iff T$ is onto.
- T is one-to-one $\iff T(\mathbf{x}) = 0$ has only the trivial solution $\iff A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\iff$ the columns of A are linearly independent.

□

Example 4. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(x_1, x_2) = (3x_1 + x_2, 5x_1 + 7x_2, x_1 + 3x_2)$. Is T a linear transformation? Is it injective? Surjective?

9 Matrix Operations

9.1 Matrix Arithmetic

An $m \times n$ matrix can be represented in several ways, some more detailed than others. Here are four ways to represent a matrix, some old and some new.

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} = \begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

Notice that a_{ij} is the ith entry of $\mathbf{a}_j$. So $\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}$. We can use these to make some definitions.

Definition

The **diagonal entries** in an $m \times n$ matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ are $a_{11}, a_{22}, \ldots$, and they form the **main diagonal**.

Definition

A **diagonal matrix** is a square $n \times n$ matrix whose nondiagonal entries are 0. (Think I_n)

Definition

An $m \times n$ matrix whose entries are all zero is a **zero matrix**, written 0 (size from context).

The same definitions for equality, sum, difference, and scalar multiples from vectors apply here. There is a very meaningful reason for this that will be explored in the future.

Example 1. Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}$, and $C = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 5 & 8 & 13 & 21 \end{bmatrix}$. Find $A + B$, $A + C$, and $2A - 3C$.

Continued...

Properties of Matrix Arithmetic

Let A, B, C be matrices of the same size, r, s be scalars.

(i) $A + B = B + A$

(ii) $(A + B) + C = A + (B + C)$

(iii) $A + 0 = A$

(iv) $r(A + B) = rA + rB$

(v) $(r + s)A = rA + sA$

(vi) $r(sA) = (rs)A$

Proofs follow from verifying corresponding column equality. We will leave these to be thought about if necessary. Notice that we have omitted any notion of multiplying matrices with each other.

9.2 Matrix Multiplication

When a matrix B acts on a vector $\mathbf{x}$ by multiplication, it transforms $\mathbf{x} \mapsto B\mathbf{x}$. If a matrix A then acts on the resulting vector by multiplication, then we get $B\mathbf{x} \mapsto A(B\mathbf{x})$.

$A(B\mathbf{x})$ is produced from $\mathbf{x}$ by a composition of mappings. Hopefully, we can find a single matrix so that $A(B\mathbf{x}) = (AB)\mathbf{x}$.

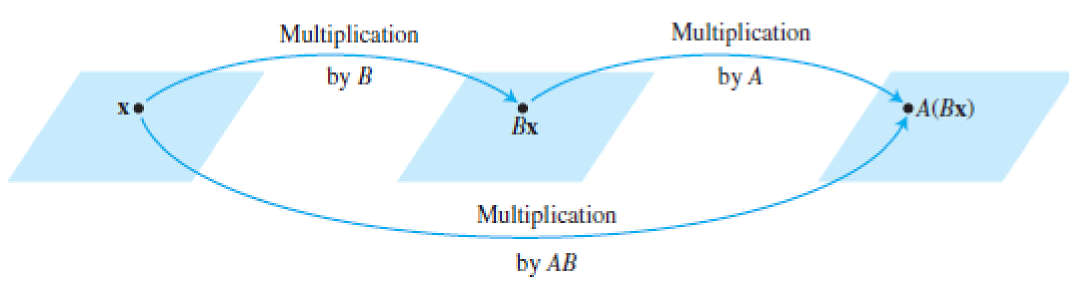


Figure 2: From Lay, Lay, McDonald's *Linear Algebra and its Applications, 5th Edition*

Suppose A is $m \times n$ while B is $n \times p$ with $\mathbf{x} \in \mathbb{R}^p$. Then by matrix-vector product, $B\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_p\mathbf{b}_p$. By the linearity of multiplication by A,

$$A(B\mathbf{x}) = A(x_1\mathbf{b}_1) + A(x_2\mathbf{b}_2) + \cdots + A(x_p\mathbf{b}_p) = x_1A\mathbf{b}_1 + x_2A\mathbf{b}_2 + \cdots + x_pA\mathbf{b}_p$$

So $A(B\mathbf{x})$ is a linear combination of the vectors $A\mathbf{b}_1, A\mathbf{b}_2, \ldots, A\mathbf{b}_p$ with $\mathbf{x}$ providing weights. Thus,

$$A(B\mathbf{x}) = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \end{bmatrix} \mathbf{x}$$

Matrix multiplication corresponds to a composition of linear transformations. Based on this exploration, we provide the following defintion.

Definition

If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then the **product** AB is the $m \times p$ matrix whose columns are $A\mathbf{b}_1, A\mathbf{b}_2, \ldots, A\mathbf{b}_p$. That is,

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

Example 2. Compute AB for $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}$.

Properties of Matrix Multiplication

Let A, B, C be matrices for which the indicated products are defined, r be a scalar.

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

(ii) $A(B+C) = AB+AC$ (left distributive law)

(iii) $(B + C)A = BA + CA$ (right distributive law)

(iv) $r(AB) = (rA)B = A(rB)$

(v) If A is $m \times n$, then $I_m A = A = AI_n$

Notice that this list of properties is similar to the previous list. The main omission is commutativity. With matrices, the order of multiplication matters.

Consider in the previous examples the size of A was 2×4 and B was 4×2, so what sizes are AB and BA?

Pitfalls: In general,

(i) $AB \neq BA$

(ii) Cancellation laws do not hold for matrix multiplication. ($AC = BC \not\Rightarrow B = C$)

(iii) Zero Product Property does not hold ($AB = 0 \not\Rightarrow A = 0$ or $B = 0$).

These statements *can* be true, but in general, we cannot assume commutativity of matrix multiplication, cancellation laws, or the zero product property. This also means that we must be considerate of the *side* that we multiply on. Particularly, are we multiplying by a matrix A on the *right* of an expression, or are we multiplying on the *left* of an expression?

9.3 Powers and Transpose of a Matrix

Because of how sizes must work out, we notice that if we want to multiply a matrix by itself, then the matrix must have the same number of rows and columns. That is, the matrix must be square. So if A is $n \times n$, then AA is defined and is $n \times n$. So is AAA, etc. We can define $A^k = \underbrace{A \cdots A}_{k \text{ times}}$.

Definition

If $A \neq 0$, $\mathbf{x} \in \mathbb{R}^n$, and $k \in \mathbb{N}$, then $A^k\mathbf{x}$ is the vector produced by left-multiplying by A k times. If $k = 0$, then $A^0\mathbf{x} = \mathbf{x}$, so $A^0 = I_n$.

Example 3. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Evaluate A^2.

Notice that the entries of A^2 are not seemingly directly related to the entries of A. It turns out, the idea of finding a power of a matrix directly is difficult to do, but it is something we will explore in the future.

Definition

Given an $m \times n$ matrix A, the **transpose** of A is the $n \times m$ matrix A^T whose columns are formed from the corresponding rows of A.

Example 4. If $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}$, find $A^T, B^T, (AB)^T$.

Transpose Properties

Let A, B be matrices for which the indicated products are defined, r be a scalar.

(i) $(A^T)^T = A$

(ii) $(A + B)^T = A^T + B^T$

(iii) $(rA)^T = rA^T$

(iv) $(AB)^T = B^T A^T$

10 The Inverse of a Matrix

10.1 Invertible and Singular Matrices

Throughout mathematics, we have continued to learn to do things, and then we learn how to undo them.

Do	**Undo**
Addition	Subtraction
Multiplication	Division
Powers	Roots (sort of)
Exponential	Logarithm
Differentiation	Integration
Matrix Transformation	?

Recall that a multiplicative inverse of a nonzero number $c \in \mathbb{R}$ is found by $c \cdot c^{-1} = 1$ and $c^{-1} \cdot c = 1$. We use this as the baseline for our definition of an invertible matrix.

Definition

An $n \times n$ matrix A is **invertible** if there is another $n \times n$ matrix C such that $CA = I$ and $AC = I$. We call C the **inverse** of A and denote it A^{-1}.

Proposition

The inverse of an invertible matrix A is unique.

Proof: Suppose an invertible matrix A has two inverses, B and C. Then

$$\begin{aligned} B &= BI \\ &= B(AC) \\ &= (BA)C \\ &= IC \\ &= C \end{aligned}$$

Therefore, $B = C$, and the inverse of A is unique. □

It turns out, depending on who you ask, someone may be more concerned with a matrix being invertible, and another person may be more concerned with a matrix that is not invertible. For this reason, we have a name for a matrix that is decidedly not invertible.

Definition

An $n \times n$ matrix that is not invertible is called **singular**.

Theorem

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and $A^{-1} = \dfrac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. If $ad - bc = 0$, then A is singular.

Definition

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. The quantity $ad - bc$ occurs often and is called the **determinant** of the 2×2 matrix A. We write $\det A = |A| = ad - bc$.

Example 1. Find $\det A$ and A^{-1} where $A = \begin{bmatrix} 2 & 2 \\ 3 & 5 \end{bmatrix}$

Note that the words "invertible" and "singular" apply only to square matrices. In this way, a square matrix may be called invertible, noninvertible, singular, or nonsingular, where

"Nonsingular" means "Invertible", and
"Noninvertible" means "Singular".

On the other hand, none of these adjectives apply to matrices that are nonsquare.

10.2 Some Theory Involving Invertible Matrices

Previously, we posed the question, "Is $A\mathbf{x} = \mathbf{b}$ always consistent?" We found the answer is no; however, with a bit of tinkering, we can actually get a result of yes.

Theorem

If A is a nonsingular $n \times n$ matrix, then for each $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof: If $\mathbf{b} \in \mathbb{R}^n$, then $A\mathbf{x} = \mathbf{b}$ has a solution in $\mathbf{x} = A^{-1}\mathbf{b}$ because

$$A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I_n\mathbf{b} = \mathbf{b}$$

If $\mathbf{u}$ is any other solution, then

$$A\mathbf{u} = \mathbf{b} \Rightarrow A^{-1}A\mathbf{u} = A^{-1}\mathbf{b} \Rightarrow I_n\mathbf{u} = A^{-1}\mathbf{b} \Rightarrow \mathbf{u} = A^{-1}\mathbf{b}$$

Thus, the solution is unique. □

Though $A^{-1}\mathbf{b}$ is a solution to $A\mathbf{x} = \mathbf{b}$ (if A is invertible), few use this formula to solve, for RREF($\begin{bmatrix} A & \mathbf{b} \end{bmatrix}$) is almost always faster than finding A^{-1} (exception being 2×2.

The power of inverses is much deeper, which we must discover.

Theorem

If A, B are invertible matrices, then

(i) A^{-1} is invertible, and $(A^{-1})^{-1} = A$.

(ii) AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.

(iii) A^T is invertible, and $(A^T)^{-1} = (A^{-1})^T$.

Generalizing the second item,

Theorem

The product of $n \times n$ invertible matrices is invertible, and the inverse is the product of the inverses in reverse order.

This is going to be an incredibly powerful theorem in order to actually find the inverse of a matrix that is not 2×2. First, we will need to introduce a particular sort of matrix.

10.3 Elementary Matrices

Elementary matrices are matrices that describe our row-reduction steps. Particularly, scaling, interchange, and replacement can all be described by matrices.

Definition

An **elementary matrix** is a matrix obtained by performing a single elementary row operation on an identity matrix. Describe what each of these elementary matrices does.

Example 2. Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $E_3 = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$.
Compute E_1A, E_2A, and E_3A.

Proposition

If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.

Recall that row operations are reversible. This was an important property of the row operations. Because elementary matrices represent the EROs, this means that E is invertible.

Proposition

Each elementary matrix E is invertible and is the elementary matrix of the same type that transforms E back into I.

Example 3. Find the inverse of $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$.

10.4 Finding the Inverse of an Invertible Matrix

Elementary matrices are the cornerstone of finding whether A is invertible as well as how to find the inverse. Moreover, these can happen at the same time!

Theorem

An $n \times n$ matrix A is invertible iff A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

Thus, A is invertible iff $\text{RREF}(A) = I$.

Proof: A is invertible $\Rightarrow A\mathbf{x} = \mathbf{b}$ is consistent $\forall \mathbf{b} \Rightarrow A$ has a pivot in every row. Since A is square, A has a pivot in every column, and the pivots must be on the diagonal. Thus, $\text{RREF}(A) = I_n$.

On the other hand, $\text{RREF}(A) = I_n \Rightarrow \exists$ elementary matrices such that

$$A \;\; E_1 A \;\; E_2(E_1 A) \;\; \cdots \;\; E_p(E_{p-1} \cdots E_2 E_1 A) = I_n$$

So $E_p \cdots E_1 A = I_n$. Since The product of $E_p \cdots E_1$ is invertible,

$$(E_p \cdots E_1)A = I_n \Rightarrow (E_p \cdots E_1)^{-1}(E_p \cdots E_1)A = (E_p \cdots E_1)^{-1} I_n \Rightarrow A = (E_p \cdots E_1)^{-1}$$

Thus, A is equal to an invertible matrix, so A is invertible. Moreover,

$$A^{-1} = ((E_p \cdots E_1)^{-1})^{-1} = E_p \cdots E_1$$

So the sequence that reduces A to I_n also transforms I_n into A^{-1}. □

Algorithm for Finding the Inverse of a Matrix

For any square matrix A,

1. Form $\begin{bmatrix} A & I \end{bmatrix}$
2. Row reduce this matrix. If we get $\begin{bmatrix} I & A^{-1} \end{bmatrix}$, then A is invertible, and A^{-1} is the second half of that matrix.
3. Write down A^{-1}.

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

11 The Invertible Matrix Theorem

11.1 The (Small) Invertible Matrix Theorem

There are a lot of consistencies throughout what we've learned. Each new piece of information has led back to some other piece of information, and we proceeded by relating back to each previous topic. This theorem seeks to unite all of those together.

The Invertible Matrix Theorem (Version 1)

Let A be an $n \times n$ matrix. The following are equivalent:

(a) A is invertible.

(b) A is row equivalent to I_n.

(c) A has n pivot positions.

(d) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(e) The columns of A are linearly independent.

(f) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.

(g) $A\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b} \in \mathbb{R}^n$.

(h) The columns of A span $\mathbb{R}^n$.

(i) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

(j) There is an $n \times n$ matrix C such that $CA = I$.

(k) There is an $n \times n$ matrix D such that $AD = I$.

(l) A^T is invertible.

This theorem is a big theorem that ties together nearly each of the big ideas so far. The IMT operates in a way that all items are simultaneously true, or they are all simultaneously false. Keep in mind, though, that this theorem *only* applies to square matrices. If a matrix is not square, then the IMT simply doesn't say anything about the matrix.

Now, because this theorem is so big, it will be a mainstay, and we will add to it throughout the remainder of the term. In fact, we will begin by altering it right now.

Recall that if A is invertible, then for all $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ has a unique solution ($\mathbf{x} = A^{-1}\mathbf{b}$). Thus, (g) can actually be replaced with "$A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$". We will update the IMT at the end of this section.

This theorem is incredibly useful for mining information and making conclusions with minimal effort. The next exercise demonstrates the power of the IMT.

Example 1. Let $A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 8 & -2 & 1 & 0 \\ 9 & 12 & -3 & -3 \end{bmatrix}$. Let $S = \left\{ \begin{bmatrix} 2 \\ 1 \\ 8 \\ 9 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -2 \\ 12 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -3 \end{bmatrix} \right\}$. Answer the following questions.

a. Is S linearly independent or linearly dependent?

b. What does $\operatorname{span} S$ look like?

c. Is $A\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ consistent or inconsistent? If it is consistent, how many solutions does it have?

d. Is A invertible or singular?

e. Let $T : \mathbb{R}^4 \to \mathbb{R}^4$ by $T(\mathbf{x}) = A\mathbf{x}$. Is T linear? Is T one-to-one? Is T onto $\mathbb{R}^4$?

11.2 Invertible Transformations

Everything so far with inverses has been about matrices. Previously, we've learned about invertible functions. Since transformations are functions, we would hope that there is a link between invertible matrices and invertible transformations. It turns out, the relationship is exactly what we would hope that it is.

Definition

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is **invertible** if there exists a function $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$S(T(\mathbf{x})) = \mathbf{x} \quad \text{for all} \mathbf{x} \in \mathbb{R}^n$$
$$T(S(\mathbf{x})) = \mathbf{x} \quad \text{for all} \mathbf{x} \in \mathbb{R}^n$$

Theorem

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be linear with standard matrix A. Then T is invertible iff A is an invertible matrix. In this case, the linear transformation S given by $S(\mathbf{x}) = A^{-1}\mathbf{x}$ is the unique function satisfying the invertible definition for T and is called the **inverse** of T.

Proof: Suppose T is invertible. Then $T(S(\mathbf{x})) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, so if $\mathbf{b}$ is any $\mathbb{R}^n$ vector and $\mathbf{x} = S(\mathbf{b})$, then $T(\mathbf{x}) = T(S(\mathbf{b})) = \mathbf{b}$, and T is onto $\mathbb{R}^n$. Thus, $\mathbf{b}$ is in the range of T. By the IMT, A is invertible (i).

Suppose A is invertible. Let $S(\mathbf{x}) = A^{-1}\mathbf{x}$. Then S is a linear transformation since it is a matrix transformation. Clearly, S satisfies the invertible definition of T. Thus, T is invertible. □

Example 2. Consider an injective linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$. Determine if T is onto $\mathbb{R}^n$.

The Invertible Matrix Theorem (Version 2)

Let A be an $n \times n$ matrix. The following are equivalent:

(a) A is invertible.

(b) A is row equivalent to I_n.

(c) A has n pivot positions.

(d) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(e) The columns of A are linearly independent.

(f) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.

(g) $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$.

(h) The columns of A span $\mathbb{R}^n$.

(i) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

(j) There is an $n \times n$ matrix C such that $CA = I$.

(k) There is an $n \times n$ matrix D such that $AD = I$.

(l) A^T is invertible.

12 Subspaces of Rn

12.1 Vector Spaces and Subspaces

This section will be a quick tour of some really big topics that are imperative to the establishment of linear algebra. Instead of the in-depth dive, we will take what we need while highlighting some of the big ideas we will be skipping past.

One of the big topics we are skipping is that of a vector space. A **vector space** is a set with established rules as to how to add and scale.

There are a few vector spaces that we are specifically interested in. We've already seen one of these: $\mathbb{R}^n$ is a vector space. The other is going to be introduced here for notational purposes.

Definition

Let $M_{m\times n}$ be the set of all matrices of size $m \times n$. This set is a vector space under the previous definitions of how to add and scale matrices.

Some of these vector space rules of addition and scaling include commutativity, associativity, distributivity, the existence of an additive identity, and the existence of additive inverses. The other necessities include additive closure, scalar multiplicative closure, and nonemptiness. These last three are included in the following definition that we *will* focus on.

Definition

A **subspace** of $\mathbb{R}^n$ is any set H in $\mathbb{R}^n$ such that

- $\mathbf{0} \in H$,
- H is closed under addition (that is, for each $\mathbf{u}, \mathbf{v} \in H$, $\mathbf{u} + \mathbf{v} \in H$), and
- H is closed under scalar multiplication (that is, for each $c \in \mathbb{R}$ and $\mathbf{u} \in H$, $c\mathbf{u} \in H$).

There are a number of subspaces that we are specifically interested in. The first is introduced in this example.

Example 1. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$. Let $H = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Show that H is a subspace of $\mathbb{R}^n$.

Continued...

This actually generalizes to any number of vectors, not just two.

Definition

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p \in \mathbb{R}^n$. Then $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ is called the **subspace spanned (or generated) by** $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$. Moreover, if H is a subspace of $\mathbb{R}^n$, and $H = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$, then we call $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ a **spanning (or generating) set for** H.

Another important subspace follows immediately from this definition.

Definition

The subspace of $\mathbb{R}^n$ generated by $\{\mathbf{0}\}$ is called the **zero subspace**.

12.2 The Four Fundamental Subspaces

Let $A \in M_{m\times n}$. From this matrix, we can produce four incredibly important subspaces: the Column, Row, Null, and Left Null spaces.

Definition

Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \in M_{m\times n}$. The **column space** of A, written $\operatorname{Col} A$, is $\operatorname{Col} A = \operatorname{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$. Thus, $\operatorname{Col} A$ is a subspace of $\mathbb{R}^m$ spanned by the columns of A.

Definition

The **null space** of an $m \times n$ matrix A, written $\operatorname{Nul} A$ is the set of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. That is, $\operatorname{Nul} A = \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, A\mathbf{x} = \mathbf{0}\}$.

Definition

Let $A \in M_{m\times n}$. The set of all linear combinations of the *row* vectors is the **row space** of A, denoted $\operatorname{Row} A$. Since rows have n entries, $\operatorname{Row} A$ is a subspace of $\mathbb{R}^n$. Moreover, the rows of A are exactly the columns of A^T, so $\operatorname{Row} A = \operatorname{Col} A^T$.

Definition

Let $A \in M_{m\times n}$. The **left-null space** of A is the set of all vectors $\mathbf{y} \in \mathbb{R}^m$ such that $\mathbf{y}^T A = \mathbf{0}^T$ (where $\mathbf{0} \in \mathbb{R}^n$ – note $\mathbf{y}^T$, $\mathbf{0}^T$ are row vectors). Note that if we transpose, we get $A^T\mathbf{y} = \mathbf{0}$. Denote the left null space of A, $\operatorname{LNul} A$. Note that $\operatorname{LNul} A = \operatorname{Nul} A^T$.

Example 2. Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$.

(a) Determine if $\begin{bmatrix} 3 \\ 4 \\ 11 \end{bmatrix} \in \operatorname{Col} A$.

(b) Determine if $\begin{bmatrix} 2 \\ 2 \\ 0 \\ 1 \\ 1 \end{bmatrix} \in \operatorname{Nul} A$.

Continued...

12.3 Bases

As it turns out, each of $\operatorname{Col} A, \operatorname{Nul} A, \operatorname{Row} A, \operatorname{LNul} A$ are each subspaces generated by a set of vectors. Each of these spaces can be expressed by a span of some vectors. For example, in the previous example, $\operatorname{Col} A = \operatorname{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}, \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix} \right\}$. But also,

$$\begin{aligned} \operatorname{Col} A &= \operatorname{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}, \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix} \right\} \\ &= \operatorname{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}, \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix} \right\} \\ &= \operatorname{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix} \right\} \\ &= \operatorname{span}\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} \right\} \end{aligned}$$

In each line, we see that $\operatorname{Col} A$ is spanned by a smaller and smaller set of vectors. So even though the column space of A is the set of all linear combinations of *all* of the columns of

A, we don't actually need all of the columns of A to produce every vector in $\operatorname{Col} A$. In fact, we only need two of them. In this case, we have a special name for this set of two vectors.

Definition

A **basis** for a subspace H of $\mathbb{R}^n$ is a linearly independent spanning set for H.

So we would say that $\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} \right\}$ is a basis for $\operatorname{Col} A$.

Definition

The set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is called the **standard basis for** $\mathbb{R}^n$.

Example 3. Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$. Find a basis for $\operatorname{Nul} A$.

Hint: $A \sim \begin{bmatrix} 1 & -2 & 0 & -1 & 3 \\ 0 & 0 & 1 & 2 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

A Basis for Each of the Four Fundamental Subspaces

Let $A \in M_{m\times n}$.

Subspace	**A Basis**
$\operatorname{Col} A$	The set of pivot columns of the original matrix A.
$\operatorname{Nul} A$	The set of vectors in a parametric vector solution for $A\mathbf{x} = \mathbf{0}$.
$\operatorname{Row} A$	The set of pivot columns of the transposed matrix A^T.
$\operatorname{LNul} A$	The set of vectors in a parametric vector solution for $A^T\mathbf{x} = \mathbf{0}$.

Example 4. Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$. Find a basis for each of the four fundamental subspaces for A.

13 Dimension and Rank

13.1 Unique Representation

One of the main advantages of having a spanning set for a subspace H of $\mathbb{R}^n$ is that each vector $\mathbf{x} \in H$ can be expressed as a linear combination of the vectors in the spanning set.

For example, if $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ 11 \end{bmatrix}$, then

$$\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}, \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix} \right\}$$

is a spanning set for $\operatorname{Col} A$. This means that because $\mathbf{x} \in \operatorname{Col} A$, $\mathbf{x}$ can be written as a linear combination of $\begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix}, \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix}$. Notice that

$$\begin{aligned}
\begin{bmatrix} 3 \\ 4 \\ 11 \end{bmatrix} &= 1 \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix} + 1 \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix} + 1 \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix} \\
&= 3 \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix} + 1 \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix} + 1 \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix} \\
&= -2 \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix} + 0 \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} + 0 \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix} + 0 \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix} \\
&= -2 \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix} + 1 \begin{bmatrix} 6 \\ -2 \\ -4 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} + 1 \begin{bmatrix} 1 \\ 3 \\ 8 \end{bmatrix} + 1 \begin{bmatrix} -7 \\ -1 \\ -4 \end{bmatrix}
\end{aligned}$$

This is often a quandary, because $\mathbf{x}$ can be written as a linear combination of those vectors, but *how* it can be written as a linear combination provides an infinite number of responses.

The Unique Representation Theorem

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ be a basis for a subspace H of $\mathbb{R}^n$. Then for each $\mathbf{x} \in H$, there exist unique scalars such that $\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n$.

Proof. Since $\mathcal{B}$ is a basis, $\operatorname{span} \mathcal{B} = V$, so the scalars exist. Suppose $\mathbf{x} = d_1\mathbf{b}_1 + d_2\mathbf{b}_2 + \cdots + d_n\mathbf{b}_n$. Subtracting the two expressions for $\mathbf{x}$,

$$\begin{aligned}\mathbf{0} &= \mathbf{x} - \mathbf{x} \\ &= (c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_n\mathbf{b}_n) - (d_1\mathbf{b}_1 + d_2\mathbf{b}_2 + \cdots + d_n\mathbf{b}_n) \\ &= (c_1 - d_1)\mathbf{b}_1 + (c_2 - d_2)\mathbf{b}_2 + \cdots + (c_n - d_n)\mathbf{b}_n\end{aligned}$$

Since $\mathcal{B}$ is linearly independent, all of the weights must be 0, so $c_i = d_i$ for all $1 \leq i \leq n$. □

One of the main advantages of having a basis for a subspace H of $\mathbb{R}^n$ is that each vector $\mathbf{x} \in H$ can be expressed *uniquely* as a linear combination of the vectors in the basis.

For example, if $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 3 \\ 4 \\ 11 \end{bmatrix}$, then

$$\left\{ \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix} \right\}$$

is a basis for $\operatorname{Col} A$. This means that because $\mathbf{x} \in \operatorname{Col} A$, $\mathbf{x}$ can be written *uniquely* as a linear combination of $\begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$. Notice that

$$\begin{bmatrix} 3 \\ 4 \\ 11 \end{bmatrix} = -2 \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix} + 3 \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$$

13.2 Coordinate Systems

Definition

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\}$ be a basis for a subspace H of $\mathbb{R}^n$. For each $\mathbf{x} \in H$, the **coordinates of x relative to $\mathcal{B}$** are the weights $c_1, \ldots, c_p$ such that

$$\mathbf{x} = c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p$$

The $\mathbb{R}^p$ vector of those weights, written $[\mathbf{x}]_{\mathcal{B}}$, is given by

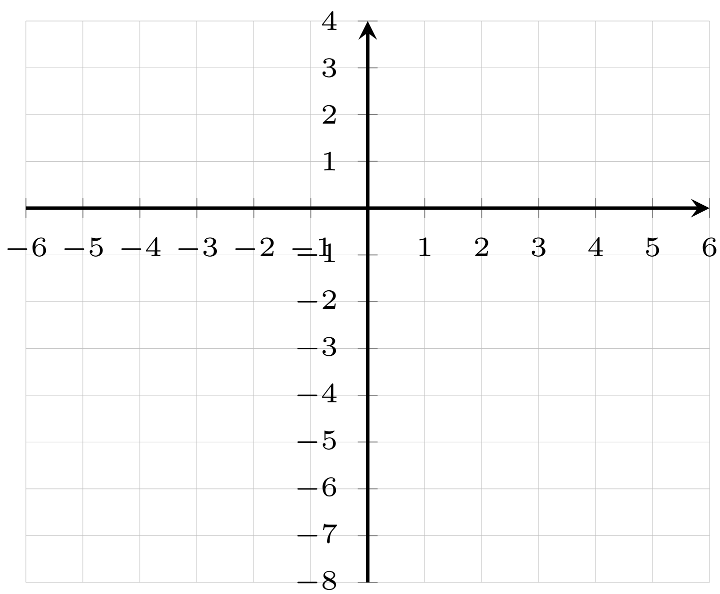
$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

and is called the **coordinate vector of x (relative to $\mathcal{B}$)**, or the **$\mathcal{B}$-coordinate vector of x**.

Example 1. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ be a basis for $\mathbb{R}^2$, where $\mathbf{b}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

(a) Suppose $\mathbf{x} \in \mathbb{R}^2$ has the coordinate vector $[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$, find $\mathbf{x}$. Draw a representation of this on the axes below.

(b) If $\mathbf{y} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$, find the $\mathcal{B}$-coordinate vector $[\mathbf{y}]_{\mathcal{B}}$ of $\mathbf{y}$.



13.3 Dimension and Rank

Theorem

If a subspace H of $\mathbb{R}^n$ has a basis of p vectors, then every basis of H must consist of exactly p vectors.

This gives us enough to define the dimension of a vector space.

Definition

If H is a subspace of $\mathbb{R}^n$ spanned by a finite set, then H is **finite-dimensional**, and the **dimension** of H, written $\dim H$, is the number of vectors in the basis for H. The dimension of $\{\mathbf{0}\}$ is 0. If H is not spanned by any finite set, then H is **infinite-dimensional**.

Example 2. Find the dimension of each of these spaces.

(a) $\mathbb{R}^3$

(b) $\mathbb{R}^5$

(c) $\mathbb{R}^n$

(d) A plane in $\mathbb{R}^3$ through the origin.

(e) A line in $\mathbb{R}^3$ through the origin.

(f) $\{\mathbf{0}\}$.

Example 3. Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$. Find a the dimension of each of the four fundamental subspaces for A.

Definition

Let $A \in M_{m\times n}$. The **rank** of A, written $\operatorname{rank} A$, is $\dim \operatorname{Col} A$.

Example 4. Let $A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$. Find $\operatorname{rank} A$.

The Rank Theorem

Let $A \in M_{m\times n}$. Then $\dim \operatorname{Col} A = \dim \operatorname{Row} A$. Moreover,

- $\operatorname{rank} A + \dim \operatorname{Nul} A = n$, and
- $\operatorname{rank} A + \dim \operatorname{LNul} A = m$.

Proof. Suppose A has r pivot columns. Since the pivot columns of A form a basis for $\operatorname{Col} A$, $\dim \operatorname{Col} A = r$, and so $\operatorname{rank} A = r$. The pivots correspond to both rows and columns, so $\dim \operatorname{Row} A = r$, as well. Thus, $\dim \operatorname{Col} A = r = \dim \operatorname{Row} A$.

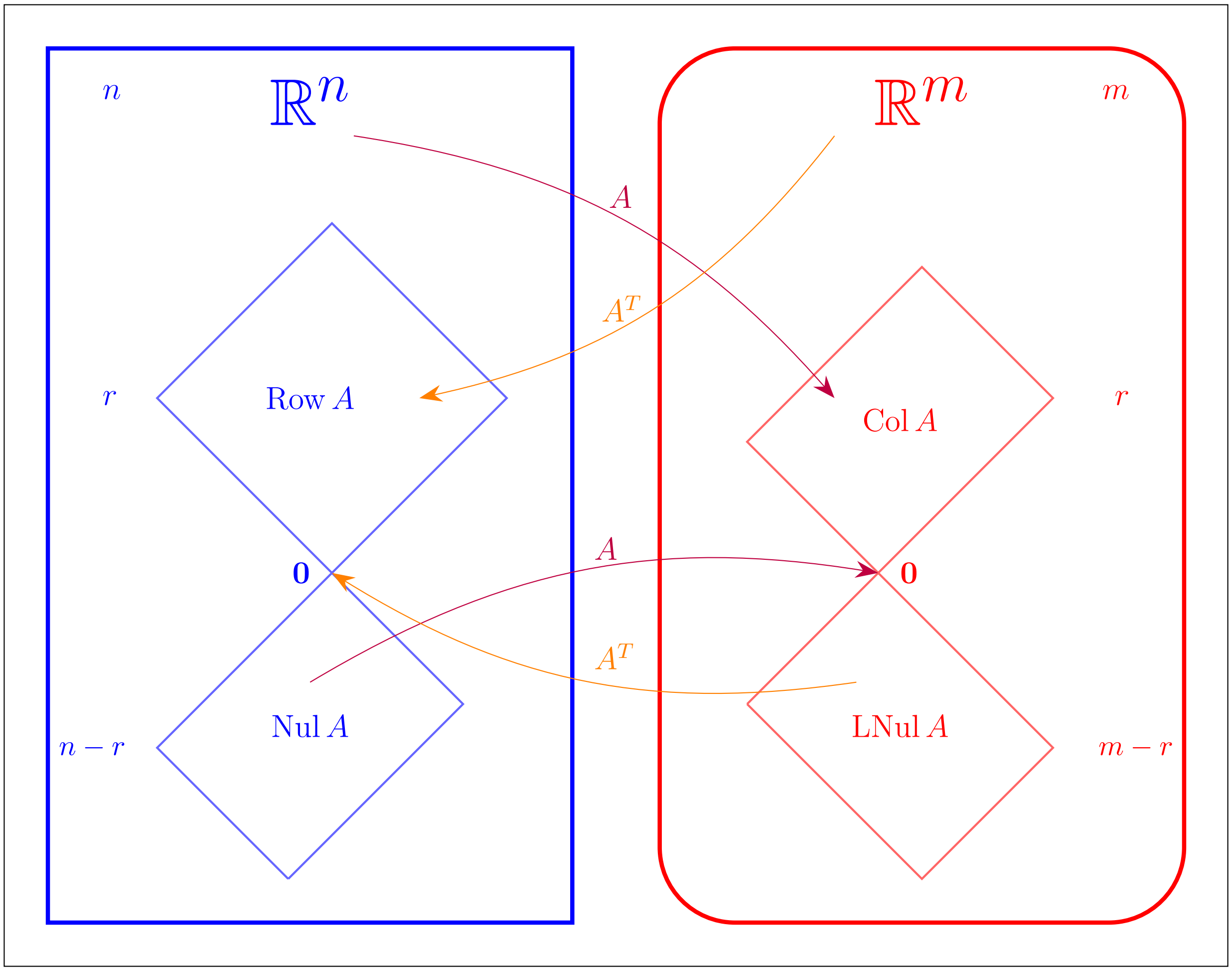
Notice $\dim \operatorname{Nul} A$ is the number of free variables of $A\mathbf{x} = \mathbf{0}$. The free variables correspond to columns that are *not* pivot columns. That is, $\dim \operatorname{Nul} A = n - \dim \operatorname{Col} A$, so $\operatorname{rank} A + \dim \operatorname{Nul} A = n$.

Since $\dim \operatorname{Col} A + \dim \operatorname{Nul} A = n$, replacing A with A^T replaces n with m, and $\dim \operatorname{Col} A^T + \dim \operatorname{Nul} A^T = m$. Since $\dim \operatorname{Col} A^T = \dim \operatorname{Row} A = \operatorname{rank} A$ and $\operatorname{Nul} A^T = \operatorname{LNul} A$, we have $\operatorname{rank} A + \dim \operatorname{LNul} A = m$. □

Example 5. If $A \in M_{12\times 14}$ has a 4-dimensional null space, what is the rank of A? What is the dimension of the left null space?

Example 6. Could a 6×9 matrix have a two-dimensional null space?

13.4 The Fundamental Theorem of Linear Algebra



Proposition

Let $A \in M_{m\times n}$. Let $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $S : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ by $T(\mathbf{x}) = A\mathbf{x}$ and $S(\mathbf{y}) = A^T\mathbf{y}$. Then

$$\begin{aligned} \operatorname{ran} T &= \operatorname{Row} A \\ \operatorname{ran} S &= \operatorname{Col} A \\ \ker T &= \operatorname{Nul} A \\ \ker S &= \operatorname{LNul} A \end{aligned}$$

This proposition is part of a larger concept called the Fundamental Theorem of Linear Algebra. This is a unifying concept for bases, dimension, vector spaces, matrices, and transformations.

13.5 The IMT Revisited

With this, we return to the IMT.

The Invertible Matrix Theorem (Version 3)

Let A be an $n \times n$ matrix. The following are equivalent:

(a) A is invertible.

(b) A is row equivalent to I_n.

(c) A has n pivot positions.

(d) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(e) The columns of A are linearly independent.

(f) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.

(g) $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$.

(h) The columns of A span $\mathbb{R}^n$.

(i) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

(j) There is an $n \times n$ matrix C such that $CA = I$.

(k) There is an $n \times n$ matrix D such that $AD = I$.

(l) A^T is invertible.

(m) The columns of A form a basis for $\mathbb{R}^n$.

(n) $\operatorname{Col} A = \mathbb{R}^n$

(o) $\dim \operatorname{Col} A = n$

(p) $\operatorname{rank} A = n$

(q) $\operatorname{Nul} A = \{\mathbf{0}\}$

(r) $\dim \operatorname{Nul} A = 0$

14 Determinants

14.1 Defining the Determinant

The determinant of a matrix is going to be a number meant to represent an entire matrix. This number has a lot of interpretations and is very flexible, though we will focus only on a single application of the determinant. We will proceed by defining the determinant of a square matrix by using the Principle of Mathematical Induction on the size of the matrix.

Definition

For the uninteresting 1×1 matrix, we define $\det A = \det \begin{bmatrix} a_{11} \end{bmatrix} = a_{11}$.

We know how to compute $\det A$ when A is 2×2. Recall that if $A = \begin{bmatrix} a_{ij} \end{bmatrix}$, then $\det A = |A| = a_{11}a_{22} - a_{12}a_{21}$.

Consider a 3×3 matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ with $a_{11} \neq 0$. Then

$$A \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & 0 & a_{11}\Delta \end{bmatrix}$$

where $\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$.

By the IMT, A is invertible iff A has 3 pivots. Notice that since $a_{11} \neq 0$, Δ determines whether A will be invertible (if $\Delta \neq 0$) or not (if $\Delta = 0$).

Definition

For a 3×3 matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}$,

$$\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

is the **determinant** of A.

Definition

Define A_{ij} to be the **submatrix** obtained from A by deleting the ith row and jth column of A.

Example 1. For the 3×3 matrix $A = \left[a_{ij}\right]$ find A_{11}, A_{12}, and A_{13}. Then find $|A_{11}|, |A_{12}|, |A_{13}|$.

This gives us a strategy for finding $\det A$ using smaller matrices.

Definition

For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}$ is the sum of n terms of the form $\pm a_{1j} \det A_{1j}$, with $\pm$ alternating. That is,

$$\begin{aligned} \det A &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} \\ &= +a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} - + \cdots + (-1)^{1+n} a_{1n} \det A_{1n} \end{aligned}$$

Example 2. Let $A = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$. Find $\det A$.

Even though we can now compute a determinant, we will explore some alternative methods of computation to provide more flexibility.

Definition

Given $A = \begin{bmatrix} a_{ij} \end{bmatrix}$, the (i, j)**-cofactor** of A is the number $C_{ij} = (-1)^{i+j} \det A_{ij}$. Thus, $\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} = \sum_{j=1}^{n} a_{1j} C_{1j}$. This formula is the **cofactor expansion across the first row** of A.

Theorem

If A is $n \times n$, then $\det A$ can be computed by a cofactor expansion across any row or down any column. Thus,

$$\begin{aligned}\det A &= \sum_{j=1}^{n} a_{ij}C_{ij} = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \\ &= \sum_{i=1}^{n} a_{ij}C_{ij} = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}\end{aligned}$$

Notice C_{ij} has an alternating sign dependent upon this checkerboard:

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \\ + & - & + & - & \\ \vdots & & & & \ddots \end{bmatrix}$$

Figure 3: The sign of C_{ij}

This theorem is particularly powerful when there is a lot of zeros in the matrix.

Example 3. Let $A = \begin{bmatrix} 2 & 7 & 1 & 8 & 2 & 8 \\ 0 & 3 & 1 & 4 & 1 & 5 \\ 0 & 0 & -1 & 6 & 4 & 1 \\ 0 & 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 2 & 4 & -1 \\ 0 & 0 & 0 & 0 & -2 & 0 \end{bmatrix}$. Find $\det A$.

Continued...

The was easy because of all of the zeros. Matrices like this are called triangular.

Definition

An $n \times n$ matrix A is **upper-triangular** if all entries below the main diagonal are zero. It is **lower-triangular** if all entries above the main diagonal are zero. It is **triangular** if it is either lower- or upper-triangular.

Theorem

If A is a triangular matrix, then $\det A$ is the product of the entries on its main diagonal.

The proof is repeated use of cofactor expansion.

Example 4. If $A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 & 2 & 2 & 2 \\ 0 & 0 & 3 & 3 & 3 & 3 & 3 \\ 0 & 0 & 0 & 4 & 4 & 4 & 4 \\ 0 & 0 & 0 & 0 & 5 & 5 & 5 \\ 0 & 0 & 0 & 0 & 0 & 6 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 \end{bmatrix}$, find $\det A$.

15 Properties of the Determinant

15.1 Determinants and Row Reduction

We haven't looked at this result specifically, but intuitively, when we row reduce to find REF of a square matrix, we end up with a triangular matrix. Now suppose we compute $\det A$ and $\det \text{REF}\, A$. Which do we anticipate is easier? How are they related?

Theorem

Let A be a square matrix, and let B be obtained from A by an ERO. Then $\det B =$...

ERO	Effect	$\det B$
Replacement	Invariant	$\det B = \det A$
Interchange	Opposite	$\det B = -\det A$
Scaling by k	Scale by k	$\det B = k \det A$

The most complicated of the EROs when computing a determinant is scaling. A common strategy is to "factor out" common multiples of a row. That is,

$$\det \begin{bmatrix} * & * & * \\ 3 & -9 & 6 \\ * & * & * \end{bmatrix} = 3 \det \begin{bmatrix} * & * & * \\ 1 & -3 & 2 \\ * & * & * \end{bmatrix}$$

Example 1. Let $A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$. Find $\det A$.

15.2 Some Theory Related to the Determinant

Suppose A is a square matrix, and $U = \text{REF}(A)$. Since U is obtained from A by interchanges and replacements (it does not require scaling), $\det A = \pm \det U$. Since U is triangular, $\det U$ is the product of the pivots. Thus, $\det A$ must be either the product of the pivots of U or its opposite.

Recall that if A is singular, then A does not have a full set of pivots, so at least one diagonal entry of U must be 0. Thus...

Theorem

A square matrix A is invertible iff $\det A \neq 0$.

We can now include this in the IMT.

Example 2. Let $A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$. Is A invertible or singular?

Theorem

If A is square, then $\det A = \det A^T$.

Theorem

If A, B are $n \times n$ matrices, then $\det AB = (\det A)(\det B)$.

Note: $\det(A + B) \neq \det A + \det B$

The Invertible Matrix Theorem (Version 4)

Let A be an $n \times n$ matrix. The following are equivalent:

(a) A is invertible.

(b) A is row equivalent to I_n.

(c) A has n pivot positions.

(d) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(e) The columns of A are linearly independent.

(f) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.

(g) $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$.

(h) The columns of A span $\mathbb{R}^n$.

(i) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

(j) There is an $n \times n$ matrix C such that $CA = I$.

(k) There is an $n \times n$ matrix D such that $AD = I$.

(l) A^T is invertible.

(m) The columns of A form a basis for $\mathbb{R}^n$.

(n) $\operatorname{Col} A = \mathbb{R}^n$

(o) $\dim \operatorname{Col} A = n$

(p) $\operatorname{rank} A = n$

(q) $\operatorname{Nul} A = \{\mathbf{0}\}$

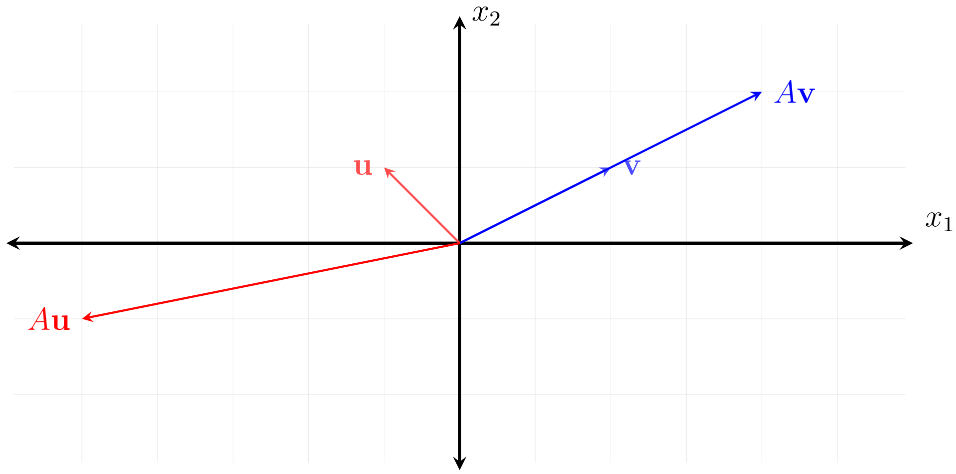
(r) $\dim \operatorname{Nul} A = 0$

(s) $\det A \neq 0$.

16 Introduction to Eigenvalues & Eigenvectors

16.1 What Are Eigenvalues and Eigenvectors?

Suppose $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Consider the images of $\mathbf{u}$ and $\mathbf{v}$ under $\mathbf{x} \mapsto A\mathbf{x}$.



Note that $A\mathbf{u}$ is a complicated transformation, but $A\mathbf{v}$ is quite simple – in fact, $A\mathbf{v} = 2\mathbf{v}$.

We will study equations such as $A\mathbf{x} = 2\mathbf{x}$ or $A\mathbf{x} = -7\mathbf{x}$. That is, we will explore when complicated transformations (matrix multiplication) can be computed in significantly simpler ways (scalar multiplication).

Definition

Let $A \in M_{n\times n}$. An **eigenvector** of A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$ – such an $\mathbf{x}$ is called an **eigenvector corresponding to** λ.

Notice that eigenvectors must not be $\mathbf{0}$, since it must correspond to a nontrivial solution.

Example 1. Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$.

(a) Determine if $\mathbf{u}, \mathbf{v}$ are eigenvectors of A.

(b) Show that 7 is an eigenvalue of A.

Continued...

Proposition

The scalar λ is an eigenvalue of $A \in M_{n\times n}$ iff $(A-\lambda I)\mathbf{x}=\mathbf{0}$ has a nontrivial solution.

Definition

The set of all solutions to $(A-\lambda I)\mathbf{x}=\mathbf{0}$ is precisely $\text{Nul}(A-\lambda I)$. So this set is a subspace of $\mathbb{R}^n$. We call this the **eigenspace** of A corresponding to λ, and this space consists of all of the eigenvectors of A corresponding to λ in addition to $\mathbf{0}$.

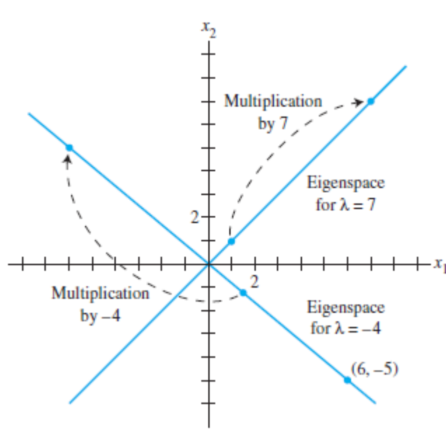


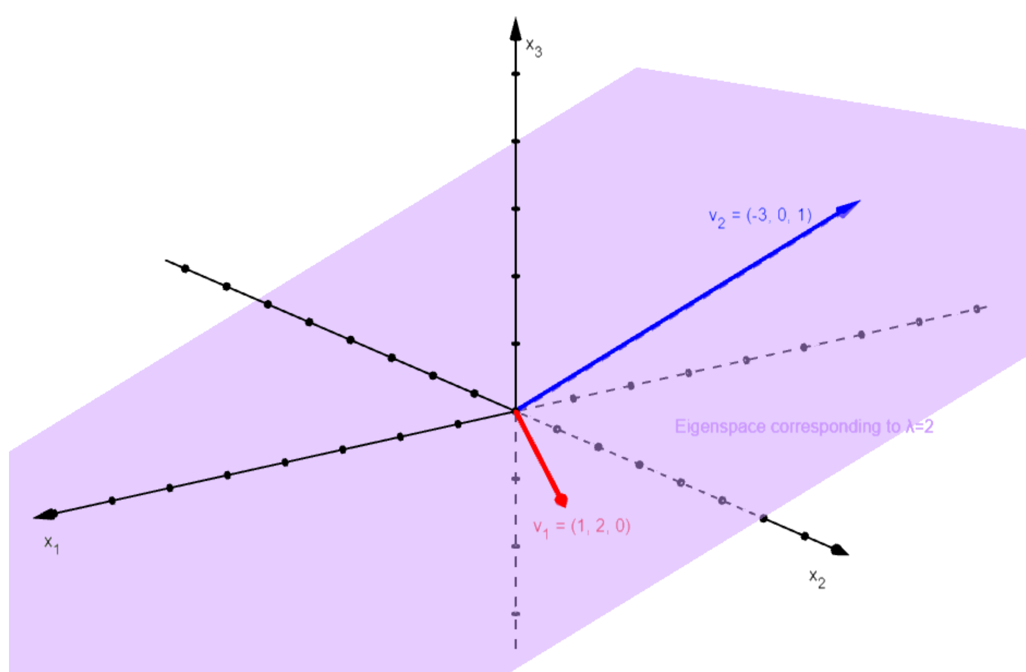
Figure 4: From Lay, Lay, McDonald's *Linear Algebra and its Applications, 5th Edition*

The action of multiplying by A on eigenspace vectors is scaling by the eigenvalue.

16.2 Finding a Basis for an Eigenspace

Example 2. Suppose we know that an eigenvalue of the matrix $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$ is 2. Find a basis for the corresponding eigenspace.

In the previous example, what is the action of multiplying by A on the eigenspace? (Dilation by a factor of 2 – doubles each vector).



Theorem

The eigenvalues of a triangular matrix are the entries on its main diagonal.

Example 3. What are the eigenvalues of $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} -5 & 0 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & -1 \end{bmatrix}$?

Proposition

A square matrix A is invertible iff 0 is not an eigenvalue of A.

Theorem

If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent.

The Invertible Matrix Theorem (Version 5)

Let A be an $n \times n$ matrix. The following are equivalent:

(a) A is invertible.

(b) A is row equivalent to I_n.

(c) A has n pivot positions.

(d) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

(e) The columns of A are linearly independent.

(f) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.

(g) $A\mathbf{x} = \mathbf{b}$ has a unique solution for each $\mathbf{b} \in \mathbb{R}^n$.

(h) The columns of A span $\mathbb{R}^n$.

(i) The linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ onto $\mathbb{R}^n$.

(j) There is an $n \times n$ matrix C such that $CA = I$.

(k) There is an $n \times n$ matrix D such that $AD = I$.

(l) A^T is invertible.

(m) $\det A \neq 0$.

(n) The columns of A form a basis for $\mathbb{R}^n$.

(o) $\operatorname{Col} A = \mathbb{R}^n$

(p) $\dim \operatorname{Col} A = n$

(q) $\operatorname{rank} A = n$

(r) $\operatorname{Nul} A = \{\mathbf{0}\}$

(s) $\dim \operatorname{Nul} A = 0$

(t) 0 is not an eigenvalue of A

17 The Characteristic Equation

17.1 Finding the Eigenvalues of a Matrix

We just learned that λ is an eigenvalue of A iff $A - \lambda I$ is singular. By the IMT, the determinant of a singular matrix must be 0. It follows that

$$\begin{gathered} \lambda \text{ is an eigenvalue of } A \\ \text{iff} \\ A - \lambda I \text{ is singular} \\ \text{iff} \\ \det(A - \lambda I) = 0. \end{gathered}$$

Definition

The equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A. A scalar λ is an eigenvalue of $A \in M_{n\times n}$ iff λ is a solution to the characteristic equation of A. The characteristic equation will always be a polynomial equations, and $\det(A - \lambda I)$ is called the **characteristic polynomial**.

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

17.2 A Shortcut for 2 by 2 Matrices

Matrices of size 2×2 are incredibly common, and that's why we have some shortcuts for these small matrices. Here's one for eigenvalues.

Definition

The **trace** of a square matrix A is the sum of the entries on its main diagonal.

Example 2. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

(a) Find $\det A$.

(b) Find $\operatorname{tr} A$.

(c) Find $\operatorname{char} A$.

Proposition

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $\operatorname{char} A = \lambda^2 - \operatorname{tr} A\lambda + \det A$.

Example 3. Find the eigenvalues of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

17.3 Theory and Similarity

Example 4. A matrix A is 6×6 and has characteristic polynomial $\lambda^6 - 4\lambda^5 - 12\lambda^4$. Find the eigenvalues (and their multiplicities) of A. Is A singular or invertible?

Finding the eigenvalues of an $n \times n$ matrix results in solving a polynomial equation of degree n – this is almost always extremely difficult, so we leave this to computers, except in the 2×2 case, which is not so hard.

Similarity is a direct application of the characteristic equation, and we will see similarity a bit in the near future.

Definition

Let $A, B \in M_{n\times n}$. Then A is **similar** to B if there is an invertible matrix P such that $P^{-1}AP = B$ or $A = PBP^{-1}$ (If $Q = P^{-1}$, then $B = QAQ^{-1}$, so B is similar to A). The transformation $A \mapsto P^{-1}AP$ is called a **similarity transform**.

Theorem

If $A, B \in M_{n\times n}$ are similar, then A and B have the same characteristic polynomial (and hence eigenvalues).

Proof. If $B = P^{-1}AP$, then

$$\begin{aligned}\det(B - \lambda I) &= \det(P^{-1}AP - \lambda P^{-1}IP) = \det(P^{-1}(A - \lambda I)P) = \det P^{-1} \det(A - \lambda I) \det P \\ &= \det P^{-1} \det P \det(A - \lambda I) = \det(P^{-1}P) \det(A - \lambda I) = \det(A - \lambda I)\end{aligned}$$

□

18 Diagonalization

18.1 Powers of a Matrix

It is often the case that we write expressions in different ways in order to see different pieces of information. Here are two examples from the past.

Equation of a Line	**Name**	**Advantage**
$Ax + By = C$	Standard Form	Easy to generalize
$y - y_1 = m(x - x_1)$	Point-Slope Form	We can see the slope and a point
$y = mx + b$	Slope-Intercept Form	We can see the slope and the y-intercept; easy to write as a function

Equation of a Parabola	**Name**	**Advantage**
$y = ax^2 + bx + c$	Standard Form	Easy to generalize; we can see the y-intercept
$ax^2 + bx + cy + d = 0$	Conic Section Form	Recognize a parabola as a conic
$y = a(x - h)^2 + k$	Vertex Form	We can see the vertex
$y = a(x - r_1)(x - r_2)$	Factored Form	We can see the x-intercepts

Each of these tables shows an object written in several different forms, and each form offers different information about the expression *just by looking at it.*

For us, we are going to start *factoring a matrix.* This particular factorization will be beneficial for us when we try to take a power of a matrix, which we have found notoriously difficult.

Example 1. Let $D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$. Compute D^2, D^3, D^k.

Example 2. Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Verify that $A = PDP^{-1}$, where $P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$, and find a formula for A^k.

18.2 The Diagonalization Theorem

Definition

A square matrix A is **diagonalizable** if A is similar to a diagonal matrix – $A = PDP^{-1}$ for some invertible P.

The Diagonalization Theorem

$A \in M_{n\times n}$ is diagonalizable iff A has n linearly independent eigenvectors.
Moreover, $A = PDP^{-1}$ where D is diagonal iff the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.
That is, A is diagonalizable iff there are enough eigenvectors of A to form a basis for $\mathbb{R}^n$, and we call such a basis an **eigenvector basis** of $\mathbb{R}^n$.

Proof. Write $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ and $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$. Recall that

$$AP = A\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{bmatrix}$$

Consider $PD = P\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}$.

Suppose A is diagonalizable and $A = PDP^{-1}$. Right-multiplying by P, we have $AP = PD$. Then

$$\begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}$$

Equating columns,

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1 \qquad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2 \qquad \cdots \qquad A\mathbf{v}_n = \lambda_n\mathbf{v}_n$$

Since P is invertible, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent and nonzero. Thus, $\lambda_1, \ldots, \lambda_n$ are eigenvalues with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. This proves Diagonalizable $\Rightarrow$ eigenvectors.

On the other hand, suppose $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent eigenvectors corresponding to $\lambda_1, \ldots, \lambda_n$, and construct P and D as before. Then $AP = PD$, as before. Since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent, P is invertible, and $A = PDP^{-1}$, so A is diagonalizable. □

Algorithm for Diagonalizing a Square Matrix

1) Find the eigenvalues of A.
2) Find n linearly independent eigenvectors. If they cannot be found, then A is not diagonalizable.
3) Construct $P = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$
4) Construct D, the diagonal matrix whose diagonal entries are the corresponding eigenvalues of A.
5) $A = PDP^{-1}$.

Example 3. Diagonalize $A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$. *It suffices to find a diagonal matrix* D *and an invertible matrix* P *such that* $A = PDP^{-1}$.

19 Inner Products, Lengths, and Orthogonality

19.1 Inner Products

It is often the case that vectors are introduced as *quantities that have both size and direction.* This is not how we introduced it in this course, but vectors do nevertheless have size and direction. We will explore that in this section.

Definition

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be considered as matrices. Then $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, which is a scalar, and is called the **inner product of u and v**, written $\mathbf{u} \cdot \mathbf{v}$. This is sometimes known as the **dot product**.

Example 1. Let $\mathbf{u} = \begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -1 \\ -2 \\ 6 \end{bmatrix}$. Compute $\mathbf{u} \cdot \mathbf{v}$ and $\mathbf{v} \cdot \mathbf{u}$.

Inner Product Properties

Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, c be a scalar. Then

- $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
- $(c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$
- $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ iff $\mathbf{u} = \mathbf{0}$

The last property is exceedingly important and is called the **positive definite** property of the inner product. Since $\mathbf{u} \cdot \mathbf{u} \geq 0$, we can take the square root of it.

19.2 Length and Unit Vectors

Definition

The **magnitude** (or **norm** or **length**) of $\mathbf{v}$ is the nonnegative scalar $|\mathbf{v}|$ is defined by $|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$ and $|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}$.

This allows us to find the size of any vector in $\mathbb{R}^n$!

Proposition

For any $\mathbf{v} \in \mathbb{R}^n$ and scalar c, the length of $c\mathbf{v}$ is $|c|$ times the length of $\mathbf{v}$; that is $|c\mathbf{v}| = |c||\mathbf{v}|$.

Proof.

$$|c\mathbf{v}|^2 = (c\mathbf{v}) \cdot (c\mathbf{v}) = c^2(\mathbf{v} \cdot \mathbf{v}) = c^2|\mathbf{v}|^2$$

Now, $|c\mathbf{v}|^2 = c^2|\mathbf{v}|^2$, so finding the square root of both sides, $|c\mathbf{v}| = |c||\mathbf{v}|$. $\square$

We will study vectors of length 1 quite a bit. Such vectors can be created by scaling a nonzero vector by the reciprocal of its length. This process is called normalization.

Definition

A vector with length 1 is a **unit vector**. If $\mathbf{v} \neq \mathbf{0}$, then $\mathbf{u} = \frac{1}{|\mathbf{v}|}\mathbf{v}$ is a unit vector in the same direction as $\mathbf{v}$. The process of creating $\mathbf{u}$ from $\mathbf{v}$ is called **normalizing v**.

Example 2. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ -2 \\ 0 \end{bmatrix}$. Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$.

19.3 Distance Between Vectors

In $\mathbb{R}$, the distance between two numbers is $|a - b|$. In higher dimensions, the same idea persists.

Definition

If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the **distance between u and v**, written $\text{dist}(\mathbf{u}, \mathbf{v})$ is the length of $\mathbf{u} - \mathbf{v}$. That is, $\text{dist}(\mathbf{u}, \mathbf{v}) = |\mathbf{u} - \mathbf{v}|$.

Example 3. Verify that this distance formula matches the distance formula in $\mathbb{R}^2$.

19.4 Orthogonal Vectors

Definition

Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

Example 4. Let $\mathbf{v}, \mathbf{0} \in \mathbb{R}^n$. Show $\mathbf{v}$ is orthogonal to $\mathbf{0}$.

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal iff $|\mathbf{u} + \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2$.

Definition

Let W be a subspace of $\mathbb{R}^n$. If $\mathbf{z}$ is orthogonal to every vector in W, then $\mathbf{z}$ is **orthogonal to** W. The set of all $\mathbf{z}$ that are orthogonal to W is the **orthogonal complement of** W, denoted $W^\perp$.

Suppse W is a plane through $\mathbf{0}$ in $\mathbb{R}^3$ and L is the line through $\mathbf{0}$ perpendicular to W (this is called a **normal line** to W). If $\mathbf{u} \in W$ and $\mathbf{v} \in L$, then $\overleftrightarrow{\mathbf{0u}} \perp \overleftrightarrow{\mathbf{0v}}$, so each vector on L is perpendicular to each vector on W. In fact, these are the *only* vectors. That is, $L = W^\perp$ and $W = L^\perp$.

Proposition

Let W be a subspace of $\mathbb{R}^n$. Then $\mathbf{x} \in W^\perp$ iff $\mathbf{x}$ is orthogonal to every vector in a spanning set for W. Moreover, $W^\perp$ is a subspace of $\mathbb{R}^n$.

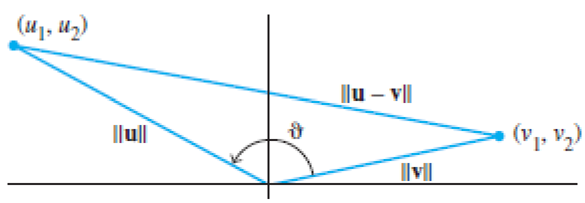
Theorem

Let $A \in M_{m\times n}$. The Row and Null spaces are orthogonal complements, and the Column and Left Null spaces are orthogonal complements. That is, $(\operatorname{Row} A)^\perp = \operatorname{Nul} A$ and $(\operatorname{Col} A)^\perp = \operatorname{LNul} A$. Moreover, if T and S are the linear transformations defined by $T(\mathbf{x}) = A\mathbf{x}$ and $S(\mathbf{x}) = A^T\mathbf{x}$, then $(\operatorname{ran} S)^\perp = \ker T$ and $(\operatorname{ran} T)^\perp = \ker S$.

Proposition

If $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then $\mathbf{u} \cdot \mathbf{v} = |\mathbf{u}||\mathbf{v}| \cos\theta$.

Proof. By the law of cosines, $|\mathbf{u} - \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 - 2|\mathbf{u}||\mathbf{v}|\cos\theta$.



$$
\begin{aligned}
2|\mathbf{u}||\mathbf{v}|\cos\theta &= |\mathbf{u}|^2 + |\mathbf{v}|^2 - |\mathbf{u} - \mathbf{v}|^2 \\
|\mathbf{u}||\mathbf{v}|\cos\theta &= \frac{1}{2}(u_1^2 + u_2^2 + v_1^2 + v_2^2 - (u_1 - v_1)^2 - (u_2 - v_2)^2) \\
&= u_1v_1 + u_2v_2 = \mathbf{u} \cdot \mathbf{v}
\end{aligned}
$$

This proves for $\mathbb{R}^2$. Similar proof for $\mathbb{R}^3$. In higher dimensions, we can use this to *define* the angle between two $\mathbb{R}^n$ vectors. □

20 Orthogonal Sets

20.1 What is an Orthogonal Set?

Definition

A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is an **orthogonal set** if each pair of distinct vectors from this set is orthogonal. That is, $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

Theorem

If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and is hence a basis for the subspace spanned by S.

Proof. Suppose S is orthogonal. Let $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \ldots, c_p$. We need to show S is linearly independent, and so we must show $c_1 = \cdots = c_p = 0$. Consider

$$\begin{aligned} 0 &= \mathbf{0} \cdot \mathbf{u}_1 \\ &= (c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\ &= (c_1\mathbf{u}_1) \cdot \mathbf{u}_1 + (c_2\mathbf{u}_2) \cdot \mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\ &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p \cdot \mathbf{u}_1) \\ &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) \end{aligned}$$

Since $\mathbf{u}_1 \neq \mathbf{0}$, $\mathbf{u}_1 \cdot \mathbf{u}_1 \neq 0$. Thus $c_1 = 0$. Doing this process with $\mathbf{u}_2, \ldots, \mathbf{u}_p$ shows $c_2, \ldots, c_p = 0$. Thus S is linearly independent. □

Definition

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

Theorem

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y} \in W$,the weights in the linear combination $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ are given by $c_j = \dfrac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j}$, where $j = 1, \ldots, p$.

This theorem states that given an orthogonal basis, the coefficients in any linear combination are easily computed.

Proof. Assume $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is orthogonal and let $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$. Consider

$$\mathbf{y} \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 = c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)$$

as in the previous proof. Since $\mathbf{u}_1 \cdot \mathbf{u}_1 \neq 0$, we can find $c_1 = \dfrac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}$. Similar for the others. $\square$

Example 1. Let $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$. Show that $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^3$ and find $\left[\mathbf{y}\right]_S$.

20.2 Orthogonal Projections

Given a nonzero vector $\mathbf{u} \in \mathbb{R}^n$, we sometimes want to decompose $\mathbf{y} \in \mathbb{R}^n$ as the sum of two vectors – one parallel to $\mathbf{u}$, and one orthogonal to $\mathbf{u}$.

We wish to write

$$\begin{aligned} \mathbf{y} &= \hat{\mathbf{y}} + \mathbf{z} \\ &= \underbrace{(\alpha \mathbf{u})}_{\text{parallel to } \mathbf{u}} + \underbrace{(\mathbf{y} - \alpha \mathbf{u})}_{\text{orthogonal to } \mathbf{u}} \end{aligned}$$

We want $\mathbf{z}$ to be a vector orthogonal to $\mathbf{u}$. Notice that

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$$

iff

$\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\mathbf{u}$

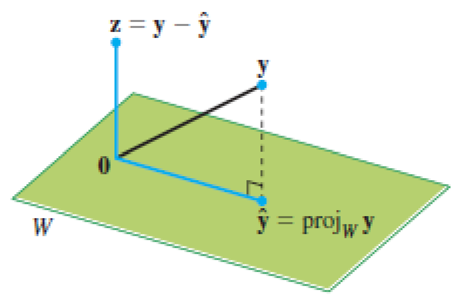
iff

$$\begin{aligned} 0 &= (\mathbf{y} - \alpha\mathbf{u}) \cdot \mathbf{u} \\ &= \mathbf{y} \cdot \mathbf{u} - (\alpha\mathbf{u}) \cdot \mathbf{u} \\ &= \mathbf{y} \cdot \mathbf{u} - \alpha(\mathbf{u} \cdot \mathbf{u}) \end{aligned}$$

Thus, $\alpha(\mathbf{u} \cdot \mathbf{u}) = \mathbf{y} \cdot \mathbf{u}$, and $\alpha = \dfrac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}$. Moreover, $\hat{\mathbf{y}} = \dfrac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}$.

Definition

The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$**, written $\hat{\mathbf{y}} = \text{proj}_{\mathbf{u}}\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. The vector $z = \mathbf{y} - \text{proj}_{\mathbf{u}}\,\mathbf{y}$ is called the **component of $\mathbf{y}$ orthogonal to $\mathbf{u}$**.



Example 2. Let $\mathbf{y} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$. Write $\mathbf{y}$ as a sum of two vectors – one in span $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$.

20.3 Orthonormal Bases

Definition

A set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If $W = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an **orthonormal basis** for W.

Example 3. Is the standard basis for $\mathbb{R}^n$ an orthonormal basis?

Example 4. Let $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}$. normalize $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$. Determine if these three normalized vectors form an orthonormal basis for $\mathbb{R}^3$.

Matrices whose columns form an orthonormal set are very important in applications and have wonderful properties. They simplify computations incredibly.

Theorem

$U \in M_{m \times n}$ has orthonormal columns iff $U^T U = I$.

Proof. We will prove for 2×2 matrices – general case is parallel.

Let $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$. Then

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \mathbf{u}_1^T \mathbf{u}_2 \\ \mathbf{u}_2^T \mathbf{u}_1 & \mathbf{u}_2^T \mathbf{u}_2 \end{bmatrix}$$

The entries are inner products using transpose notation. Thus, the columns are orthogonal iff $\mathbf{u}_1^T \mathbf{u}_2 = \mathbf{u}_2^T \mathbf{u}_1 = 0$. The columns have unit length iff $|\mathbf{u}_1|^2 = \mathbf{u}_1^T \mathbf{u}_1 = 1$ and $|\mathbf{u}_2|^2 = \mathbf{u}_2^T \mathbf{u}_2 = 1$. □

Example 5. If U has orthonormal columns, is U invertible? If so, what is U^{-1}?

Theorem

A square matrix U is **orthogonal** if $U^T U = I$.

Theorem

Let $U \in M_{m \times n}$ be orthogonal, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

a. $|U\mathbf{x}| = |\mathbf{x}|$

b. $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$

c. $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ iff $\mathbf{x} \cdot \mathbf{y} = 0$

This theorem essentially says the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality which is often mandatory for computer algorithms.

21 Orthogonal Projections

Let W be a subspace of $\mathbb{R}^n$. If $\mathbf{y} \in \mathbb{R}^n$, it is often useful to be able to write $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$, where $\mathbf{z}_1 \in W$ and $\mathbf{z}_2 \in W^\perp$, especially given orthogonal bases.

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y} \in \mathbb{R}^n$ can be written uniquely in the form $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$, where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$.

Proof. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be any orthogonal basis for W, and define

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p$$

Then $\hat{\mathbf{y}} \in W = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. Let $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$. Since $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$. It follows that

$$\mathbf{z} \cdot \mathbf{u}_1 = (y - \hat{\mathbf{y}}) \cdot \mathbf{u}_1 = \mathbf{y} \cdot \mathbf{u}_1 - \left(\frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 \cdot \mathbf{u}_1 - 0 - 0 - \cdots - 0 = \mathbf{y} \cdot \mathbf{u}_1 - \mathbf{y} \cdot \mathbf{u}_1 = 0$$

So $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$. Similarly, $\mathbf{z}$ is orthogonal to each $\mathbf{u}_j$ in the basis for W. So $\mathbf{z} \in W^\perp$.

For uniqueness, suppose $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$, where $\hat{\mathbf{y}}_1 \in W$ and $\mathbf{z}_1 \in W^\perp$. Then $\hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$, so $\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$. Since $\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 \in W$, $\mathbf{z}_1 - \mathbf{z} \in W^\perp$, and the two are equal, they must each be $\mathbf{0}$. Thus, $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1$ and $\mathbf{z}_1 = \mathbf{z}$. □

Example 1. Let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5\}$ be an orthogonal basis for $\mathbb{R}^5$ and W be the subspace of $\mathbb{R}^5$ generated by $\mathbf{u}_1, \mathbf{u}_2$. Let $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_5\mathbf{u}_5$. Write $\mathbf{y}$ as the sum of a vector $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$.

The uniqueness of this decomposition shows that the orthogonal projection $\hat{\mathbf{y}}$ depends only on W and not on the basis.

Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^n$, and $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$.
The vector $\hat{\mathbf{y}}$ is called **the best approximation to $\mathbf{y}$ by elements of W**. The distance between $\mathbf{y}$ and a vector $\mathbf{v} \in W$ used to approximate $\mathbf{y}$ can be thought of as the *error* of using $\mathbf{v}$ instead of $\mathbf{y}$. The Best Approximation Theorem states this error is minimized when $\mathbf{v} = \hat{\mathbf{y}}$.

Proof. Let $\mathbf{v} \in W$ be distinct from $\hat{\mathbf{y}}$. Then $\hat{\mathbf{y}} - \mathbf{v} \in W$. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}} \in W^{\perp}$. In particular, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$ ($\in W$).

Since $\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$, the Pythagorean Theorem says $|\mathbf{y} - \mathbf{v}|^2 = |\mathbf{y} - \hat{\mathbf{y}}| + |\hat{\mathbf{y}} - \mathbf{v}|^2$. Thus

$$|\mathbf{y} - \hat{\mathbf{y}}|^2 = |\mathbf{y} - \mathbf{v}|^2 - |\hat{\mathbf{y}} - \mathbf{v}|^2$$

Since $\mathbf{v}$ and $\hat{\mathbf{y}}$ are distinct, $|\hat{\mathbf{y}} - \mathbf{v}|^2 > 0$, and so $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$. □

Example 2. If $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $W = \text{span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Find the distance from $\mathbf{y}$ to W.

If the basis for W happens to be orthonormal, then computations are simplified greatly.

Theorem

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p$$

If $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p \end{bmatrix}$, then $\text{proj}_W \mathbf{y} = UU^T\mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^n$.

Proof. From the Orthogonal Decomposition Theorem, the denominators of the weights of each $\mathbf{u}_j$ is $\mathbf{u}_j \cdot \mathbf{u}_j = |\mathbf{u}_j|^2 = 1^2 = 1$. □

22 The Gram-Schmidt Process

22.1 Gram-Schmidt

Clearly, orthogonal bases are nice. They have unique representation of vectors with predictable weights. Projecting vectors onto spaces generated by orthogonal bases also have very nice expressions.

The Gram-Schmidt Process is an algorithm for producing an orthogonal or orthogonal basis for any nonzero subspace of $\mathbb{R}^n$.

The Gram-Schmidt Process: An Algorithm for Producing an Orthogonal Basis

Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a nonzero subspace W or $\mathbb{R}^n$, let

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \operatorname{proj}_{\mathbf{v}_1} \mathbf{x}_2 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \operatorname{proj}_{\mathbf{v}_1} \mathbf{x}_3 - \operatorname{proj}_{\mathbf{v}_2} \mathbf{x}_3 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \operatorname{proj}_{\mathbf{v}_1} \mathbf{x}_p - \operatorname{proj}_{\mathbf{v}_2} \mathbf{x}_p - \cdots - \operatorname{proj}_{\mathbf{v}_{p-1}} \mathbf{x}_p
\end{aligned}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W.

Example 1. Let $\mathbf{x}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$. Then $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthonormal basis for W.

Continued...

22.2 QR Factorization

Suppose $A \in M_{m\times n}$ has linearly independent columns. Then applying Gram-Schmidt on A will result in a factorization of A called a QR-factorization, used often in applications for solving and finding eigenvalues.

QR Factorization

If $A \in M_{m\times n}$ has linearly independent columns, then A can be factored as $A = QR$, where $Q \in M_{m\times n}$ consists of columns that form an orthogonal basis for $\operatorname{Col} A$, and $R \in M_{n\times n}$ is an upper-triangular matrix with positive diagonal entries.

Proof. The columns of A form a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ for $\operatorname{Col} A$. Use Gram-Schmidt (or another method) to form an orthogonal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for $\operatorname{Col} A$. Let $Q = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$.

For $k = 1, \ldots, n$, $\mathbf{x}_k \in \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_n\} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$. Thus,

$$\mathbf{x}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_n$$

We may assume $r_{kk} > 0$. Thus, $\mathbf{x}_k$ is a linear combination of the columns of Q using weights from

$$\mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Hence, $\mathbf{x}_k = Q\mathbf{r}_k$ for $k = 1, \ldots, n$. Let $R = \begin{bmatrix} \mathbf{r}_1 & \cdots & \mathbf{r}_n \end{bmatrix}$. It follows that

$$A = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} Q\mathbf{r}_1 & \cdots & Q\mathbf{r}_n \end{bmatrix} = QR$$

Since R is square, upper triangular, and its main diagonal entries are nonzero, $\det R \neq 0$. Thus, R is invertible. □

Proposition

If $A \in M_{m\times n}$ has linearly independent columns and QR factorization $A = QR$, then $R = Q^T A$.

Proof. $Q^T A = Q^T(QR) = (Q^T Q)R = IR = R$. □

Example 2. Find a QR factorization for $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.