MATH 118, Spring 2020, Linear Algebra
Condensed Lecture Notes

Taken in part from
Introduction to Linear Algebra, 4e,
Gilbert Strang
and *Linear Algebra and its Applications 5e,*
David C. Lay, Steven R. Lay, Judy J. McDonald

Notes compiled by Bobby McDonald
Yale University, 2020

NOTE: I will update these notes as often as I can with the topics and examples (which will be worked out by hand in a separate document) we cover in class.

Contents

Chapter 1

Introduction to Vectors

1.1 Vectors and linear combinations

1.1. Key Ideas

- A vector $\mathbf{v}$ in two-dimensional space has two components v_1 and v_2.
- $\mathbf{v}+\mathbf{w} = \langle v_1+w_1, v_2+w_2\rangle$ and $c\mathbf{v} = \langle cv_1, cv_2\rangle$ are found a component at a time.
- A linear combination of three vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ is $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$.
- In three dimensions, *all* linear combinations of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ typically fill a line, then a plane, then the whole space $\mathbb{R}^3$.

Definition 1.1.1 (Vectors in $\mathbb{R}^2$)**.** A array with only one column is called a **column vector**, or just a **vector**. Examples of vectors with two entries are

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} \sqrt{2} \\ \pi \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

where w_1, w_2 are real numbers. The set of all vectors with two entries is called $\mathbb{R}^2$. Two vectors are **equal** if and only if their corresponding entries are equal.

Definition 1.1.2. Given two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$, their **sum** is the vector $\mathbf{u}+\mathbf{v}$ obtained by adding the corresponding entries of $\mathbf{u}$ and $\mathbf{v}$. For example,

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

Given a vector $\mathbf{v}$ and a real number c, the **scalar multiple** of $\mathbf{u}$ is the vector $c\mathbf{u}$ obtained by multiplying each entry of $\mathbf{u}$ by c. For example if

$$c = 2 \text{ and } \mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \text{ then } c\mathbf{u} = 2\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

Example 1.1.3. Given vectors $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -3 \\ 4 \end{bmatrix}$, find $(-2)\mathbf{u}$, $(-2)\mathbf{v}$, and $\mathbf{u} + (-3)\mathbf{v}$.

Observation 1.1.4 (Vectors in $\mathbb{R}^2$)**.** We can identify the column vector $\left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ with *the point* (a, b) in the plain, so we can consider $\mathbb{R}^2$ as the set of all points in the plain. We usually visualize a vector by including an arrow from the origin.

Partners 1.1.5. Let $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Graph $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u} + \mathbf{v}$ on the plane.

Proposition 1.1.6 (Parallelogram Rule)**.** *Geometrically,* $\mathbf{u} + \mathbf{v}$ *is the last vertex of the parallelogram with vertices are* $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{0}$.

Partners 1.1.7. Let $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Graph $\mathbf{u}$, $(-2)\mathbf{u}$, and $3\mathbf{u}$. What's special about $c\mathbf{u}$ for any c?

Groups 1.1.8. If $\mathbf{u} = (a, b)$, can you find an equation for the line that contains all multiples of $\mathbf{u}$?

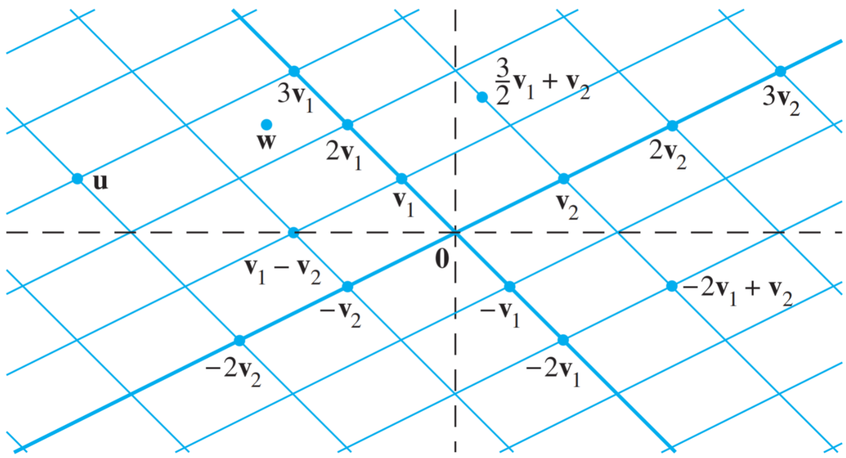
Question 1.1.9. What properties do addition of vectors and multiplication by scalars enjoy?

Definition 1.1.10. Given vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, and real numbers c, d. The vector

$$c\mathbf{u} + d\mathbf{v}$$

is called a **linear combination** of $\mathbf{u}$ and $\mathbf{v}$ with **weights** c and d.

Example 1.1.11. The figure below shows linear combinations of $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ where with integer weights. Estimate the linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$ that produce $\mathbf{u}$ and $\mathbf{w}$.



Definition 1.1.12. If $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are in $\mathbb{R}^3$, then the set of all linear combinations of is denoted by Span$\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ and is called the **span of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$**. It's the collection of all vectors that look like

$$a\mathbf{u} + b\mathbf{v} + c\mathbf{w}, \text{ with } a, b, c \text{ real numbers.}$$

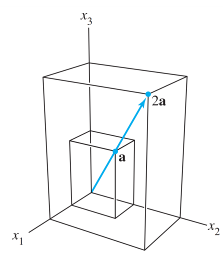
Remark 1.1.13. This definition can be extended to linear combinations of three or more vectors

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n.$$

Think, Pair, Share 1.1.14. Let $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. If we plotted all linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$ what would we get?

Observation 1.1.15 (Vectors in $\mathbb{R}^3$)**.** Vectors in $\mathbb{R}^3$ are 3×1 matrices. Like above, we can represent them geometrically in three-dimensional coordinate space. For example,

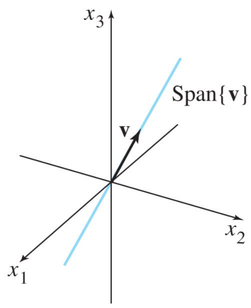
$$\mathbf{a} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

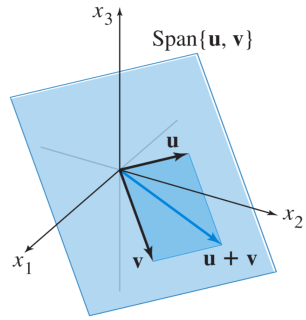


Question 1.1.16. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^3$. What do linear combinations of $\mathbf{u}$ and $\mathbf{v}$ look like?

Observation 1.1.17 (Geometric Descriptions of Span$\{\mathbf{u}\}$ and Span$\{\mathbf{u}, \mathbf{v}\}$)**.** Let $\mathbf{u}$ and $\mathbf{v}$ be nonzero vectors in $\mathbb{R}^3$, with $\mathbf{u}$ not a multiple of $\mathbf{v}$.

- Span$\{\mathbf{v}\}$ is the line through $\mathbf{0}$ and $\mathbf{v}$,
- Span$\{\mathbf{u}, \mathbf{v}\}$ is the plane containing $\mathbf{0}$, $\mathbf{u}$, and $\mathbf{v}$.





Definition 1.1.18 (Vectors in $\mathbb{R}^n$)**.** If n is a positive integer, $\mathbb{R}^n$ denotes the collection of ordered n-tuples of n real numbers, usually written as $n \times 1$ column matrices, such as

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix},$$

we we sometimes denote $\langle a_1, a_2, \ldots, a_n \rangle$. The **zero vector**, denoted $\mathbf{0}$ is the vector whose entries are all zero. We also denote $(-1)\mathbf{u} = -\mathbf{u}$.

1.2 Lengths and dot products

1.2. Key Ideas

- The dot product $\mathbf{v} \cdot \mathbf{w}$ multiplies each component v_i by w_i and adds all $v_i w_i$.
- The length $\|\mathbf{v}\|$ of a vector $\mathbf{v}$ is the square root of $\mathbf{v} \cdot \mathbf{v}$.
- $\mathbf{u} = \mathbf{v}/\|\mathbf{v}\|$ is a **unit vector.** Its length is 1
- The dot product $\mathbf{v} \cdot \mathbf{w} = 0$ when the vectors $\mathbf{v}$ and $\mathbf{w}$ are perpendicular
- If θ is the angle between $\mathbf{v}$ and $\mathbf{w}$, then $\cos\theta = \dfrac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\|\|\mathbf{w}\|}$

Definition 1.2.1. If $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ are vectors in $\mathbb{R}^n$, then the **dot product** or **inner product** of $\mathbf{u}$ and $\mathbf{v}$ is $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2$

Question 1.2.2. What is $\mathbf{u} \cdot \mathbf{v}$? Is it a vector?

Example 1.2.3. Suppose we are buying and selling candy. Gum costs \$1.00 for a pack, chocolate is \$0.75 a bar, and hard candies are \$1.50 for a roll. If we sell 10 packs of gum and 20 chocolate bars, and buy 10 rolls of hard candy, what is our total income? Okay. We know how to do this. What does it have to do with dot products?

Groups 1.2.4. Compute $\mathbf{u} \cdot \mathbf{v}$ for $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}$.
Without any calculation, can you decide what $\mathbf{v} \cdot \mathbf{u}$ is?

Theorem 1.2.5. Let $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.

Definition 1.2.6. The **length** (or **norm**) of $\mathbf{v}$ is the square root of dotted with itself:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + v_3^2}, \quad \text{and} \quad \|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}$$

Example 1.2.7. Find the norm of $\mathbf{v} = (1, -2, 2)$.

Think, Pair, Share 1.2.8. Let $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. What does the norm of $\mathbf{v}$ represent geometrically?

Definition 1.2.9. A vector whose length is 1 is called a **unit vector**.

Example 1.2.10. Let $\mathbf{v} = (1, -2, 2)$. Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$.

Definition 1.2.11. If we divide a vector by its length, we obtain a unit vector *in the same direction as* $\mathbf{v}$. This is called **normalizing** $\mathbf{v}$.

Groups 1.2.12. Show that $\mathbf{u} = (2, 3)$ and $\mathbf{v} = (-3, 2)$ meet at right angles. Hint: we've already seen that (a, b) lives on the line $ay = bx$.

Definition 1.2.13. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (or perpendicular) if $\mathbf{u} \cdot \mathbf{v} = 0$.

Groups 1.2.14. Find a nonzero vector in $\mathbb{R}^3$ that is orthogonal to $\mathbf{u} = (1, 2, 3)$.

Theorem 1.2.15. If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors, then

$$\|\mathbf{u}\|\|\mathbf{v}\|\cos(\theta) = \mathbf{u} \cdot \mathbf{v}$$

Question 1.2.16. What does this tell us about the *sign* of the dot product $\mathbf{u} \cdot \mathbf{v}$?

Theorem 1.2.17 (Pythagorean Theorem)**.** Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

1.3 Matrices

1.3. Key Ideas

- linear equations and vector equations
- solving simple systems of equations
- matrices
- **Matrix times vector:** $A\mathbf{x}$ = linear combination of the columns of A with x_i as weights.

1.3.1 Linear equations

Example 1.3.1. Suppose we are buying and selling candy, again. Remember, gum costs \$1.00 for a pack, chocolate is \$0.75 a bar, and hard candies are \$1.50 for a roll. Suppose

- Monday, we sell 10 packs of gum and 20 chocolate bars and buy 10 rolls of hard candy,
- Tuesday, we buy 10 packs of gum, sell 10 chocolate bars, and buy/sell no hard candies,
- Wednesday, we buy/sell no packs of gum, buy 4 chocolate bars, and buy/sell no hard candies.

What is our net profit?

Observation 1.3.2. We can represent the previous example using three separate equations,

$$\begin{aligned} 1.00 \times 10 \ + \ 0.75 \times 20 \ + \ 1.50 \times 10 &= \text{Monday Profit} \\ 1.00 \times 10 \ + \ 0.75 \times 10 \ + \ 1.50 \times \ \ 0 &= \text{Tuesday Profit} \\ 1.00 \times 10 \ + \ 0.75 \times 20 \ + \ 1.50 \times 10 &= \text{Wednesday Profit} \end{aligned}$$

or as a **vector equation**

$$1.00 \times \begin{bmatrix} 10 \\ 10 \\ 0 \end{bmatrix} + 0.75 \times \begin{bmatrix} 20 \\ 10 \\ 4 \end{bmatrix} + 1.50 \times \begin{bmatrix} 10 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \text{Mon. P.} \\ \text{Tues. P.} \\ \text{Wed. P.} \end{bmatrix}$$

Definition 1.3.3. A **linear equation** in the variables $x_1, \dots, x_n$ is an equation that can be written in the form

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

where b and the **coefficients** $a_1, \dots, a_n$ are real or complex numbers.

Example 1.3.4. Which of the following are linear equations?

1. $4x_1 - 5x_2 + 2 = x_1$
2. $x_2 = 2(\sqrt{6} - x_1) + x_3$
3. $4x_1 - 5x_2 = x_1x_2$
4. $x_2 = 2\sqrt{x_1} - 6$

Definition 1.3.5. A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same variables, say $x_1, \dots, x_n$.

Example 1.3.6. Is $(5, 6.5, 3)$ in the solution set (the set of all solutions) of the system

$$\begin{aligned} 2x_1 - x_2 + 1.5x_3 &= 8 \\ x_1 \qquad\quad - \ \ 4x_3 &= -7 \end{aligned}$$

Definition 1.3.7. The collection of all solutions to a system is called the **solution set**. Two linear systems are called **equivalent** if they have the same solution set.

Remark 1.3.8. You already know how to find the solution set to a system of two linear equations in two unknowns! Just find the intersection of the two lines!

Example 1.3.9. What are the solution sets of the following systems?

(a) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 3x_2 &= 3 \end{aligned}$$

(b) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 3 \end{aligned}$$

(c) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ 2x_1 - 4x_2 &= -2 \end{aligned}$$

Example 1.3.10. What is the solution set of the following system? If we fix b_1, b_2, b_3, how many solutions will it have?

$$\begin{aligned} x_1 \qquad\qquad &= b_1 \\ -x_1 + x_2 \qquad &= b_2 \\ -x_2 + x_3 &= b_3 \end{aligned}$$

Example 1.3.11. How can we interpret solutions to systems of equations with three variables geometrically?

1.3.2 Matrices

Observation 1.3.12. We can also represent our first example more concisely using something called a *matrix* and multiplying it by a vector.

$$1.00 \times \begin{bmatrix} 10 \\ 10 \\ 0 \end{bmatrix} + 0.75 \times \begin{bmatrix} 20 \\ 10 \\ 4 \end{bmatrix} + 1.50 \times \begin{bmatrix} 10 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 & 20 & 10 \\ 10 & 10 & 0 \\ 0 & 4 & 0 \end{bmatrix} \begin{bmatrix} 1.00 \\ 0.75 \\ 1.50 \end{bmatrix}$$

Definition 1.3.13. A matrix is an array of numbers. If A is a matrix with m rows and n columns, then A is called an $m \times n$ matrix. The entry in the ith row and jth column of A is called the (i, j) entry of A.

Definition 1.3.14 (Matrix times a vector)**.** If A has columns $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, and $\mathbf{x}$ is the vector (c, d, e), then

$$A\mathbf{x} = \begin{bmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{bmatrix} \begin{bmatrix} c \\ d \\ e \end{bmatrix} = c\mathbf{u} + d\mathbf{v} + e\mathbf{w}.$$

Example 1.3.15. Compute the product $A\mathbf{x}$ where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Procedure 1.3.16. If $A\mathbf{x} = \mathbf{b}$, then the ith entry of $\mathbf{b}$ is the dot product of the ith row of A with $\mathbf{x}$.

1.3.3 Linear equations and matrices

Example 1.3.17. Compute the product $A\mathbf{x}$ where

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Example 1.3.18. What if $A\mathbf{x} = \mathbf{b}$ where A and $\mathbf{b}$ are given, but $\mathbf{x}$ is unknown? How could we find $\mathbf{x}$ if we're told

$$A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Example 1.3.19. What is $\mathbf{x}$ if

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

Chapter 2

Solving Linear Equations

2.1 Vectors and linear equations

2.1. Key Ideas

- systems of equations can have no, one, or many solutions
- a system of equations with at least one solution is called consistent
- systems can be solved using back substitution

Example 2.1.1. How many solutions do each of the following systems have?

(a)
$$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 3x_2 &= 3 \end{aligned}$$

(b)
$$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 3 \end{aligned}$$

(c)
$$\begin{aligned} x_1 - 2x_2 &= -1 \\ 2x_1 - 4x_2 &= -2 \end{aligned}$$

Proposition 2.1.2. *A system of linear equations has*

1. *no solution, or*
2. *exactly one solution, or*
3. *infinitely many solutions.*

A system is called **consistent** if it has either one or infinitely many solutions, and **inconsistent** if it has no solution.

Example 2.1.3. Are the following system consistent?

(a) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 3x_2 &= 3 \end{aligned}$$

(b) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 3 \end{aligned}$$

(c) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ 2x_1 - 4x_2 &= -2 \end{aligned}$$

Remark 2.1.4. We will be interested in two fundamental questions about linear systems:

1. Is the system consistent?
2. If a system is consistent, is the solution unique?

Example 2.1.5. Determine if the following system of equations is consistent.

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

2.2 The idea of elimination

2.2. Key Ideas

- A linear system becomes upper triangular after elimination.
- We subtract ℓ_{ij} times equation j from equation i to make the (i, j) entry zero, where

$$\ell_{ij} = \frac{(i, j) \text{ entry}}{\text{pivot in row } j}.$$

- The upper triangular system is solved by back substitution.

Example 2.2.1. In Section 1.3 we determined whether the following systems were consistent using geometric and substitution arguments. Is there an algebraic way to do this without substitution?

(a) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 3x_2 &= 3 \end{aligned}$$

(b) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 3 \end{aligned}$$

(c) $$\begin{aligned} x_1 - 2x_2 &= -1 \\ 2x_1 - 4x_2 &= -2 \end{aligned}$$

Remark 2.2.2 (The idea of elimination)**.** We can solve a system by replacing it with an equivalent system that's easier to solve. We can do this by replacing one equation by adding multiples of equations, interchanging equations, or multiplying an equation by a nonzero constant.

Example 2.2.3. Determine if the following system of equations is consistent without substitution

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

Example 2.2.4. Determine if the following system is consistent:

$$\begin{aligned} x_2 - 4x_3 &= 8 \\ 2x_1 - 3x_2 + 2x_3 &= 1 \\ 4x_1 - 6x_2 + 4x_3 &= 1 \end{aligned}$$

2.3 Elimination using matrices

2.3. Key Ideas

- we can record information about a system in a matrix
- we can use elementary row operations to reduce matrices
- row reduction algorithm and how to use it to solve a system of equations

Definition 2.3.1. The essential information in a linear system can be recorded into a rectangular array called a **matrix**. For example, given the system

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

with the coefficients of each variable aligned in columns, the matrix

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$$

is called the **coefficient matrix** of the system, and

$$\left[\begin{array}{ccc:c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{array}\right]$$

is called the **augmented matrix** of the system. The **size** of a matrix tells us how many rows and columns it has. An $\mathbf{m} \times \mathbf{n}$ **matrix** has m rows and n columns.

Remark 2.3.2. Matrices will make our lives much easier when solving systems of linear equations!

Example 2.3.3. Solve the system

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

using a matrix.

Definition 2.3.4 (Elementary Row Operations)**.**

1. (Replacement) replace one row by the sum of itself and a multiple of another row.
2. (Interchanging) Interchange two rows.
3. (Scaling) Multiply all entries in a row by a nonzero constant.

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

Remark 2.3.5. If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

Example 2.3.6. Determine if the following system is consistent:

$$\begin{aligned} x_2 - 4x_3 &= 8 \\ 2x_1 - 3x_2 + 2x_3 &= 1 \\ 4x_1 - 8x_2 + 12x_3 &= 1 \end{aligned}$$

Definition 2.3.7. The leftmost nonzero entry in a row is called the **leading entry**. A matrix is in **echelon form** (or **row echelon form**) if it has the following three properties:

1. all zero row are at the bottom
2. each leading entry of a row to the right of the leading entry above it
3. all entries in a column below a leading entry are zeros

If a matrix in echelon form satisfies the following addition conditions, then it is in **reduced echelon form** (or **reduced row echelon form**:

4. the leading entry in each nonzero row is 1.
5. each leading 1 is the only nonzero entry in its column.

Example 2.3.8. Which of the following is in echelon form? Reduced echelon form?

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 5/2 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & 0 & -5 \\ 0 & 1 & 2 & 12 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Remark 2.3.9. Any nonzero matrix can be **row reduced** into infinitely many matrices in echelon form. However, *reduced* echelon form for a matrix is unique.

Definition 2.3.10. A **pivot position** in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A **pivot column** is a column of A that contains a pivot position. A **pivot** is a nonzero entry in a pivot position.

Example 2.3.11. Label the pivot positions and pivot columns of the matrices above.

Example 2.3.12. Row reduce the matrix A to echelon form and locate pivot columns.

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix}$$

Procedure 2.3.13 (Row Reduction Algorithm)**.** To transform a matrix into echelon form:

1. begin with the leftmost nonzero column
2. interchange rows if necessary to move a nonzero entry in this column to the top
3. use row replacement to create zeros in all positions below the new pivot.
4. ignore the row containing the pivot positions and all rows, if any, above it and apply steps 1-3 to the submatrix that remains

If you want reduced echelon form, add one more step

5. Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. If a pivot is not 1, make it 1 by scaling.

Example 2.3.14. Apply elementary row operations to transform the following matrix into echelon form, and then reduced echelon form.

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

Definition 2.3.15. Steps 1-4 above are called the **forward phase** of the row reduction algorithm. Step 5 is called the **backward phase.**

Example 2.3.16. Find the general solution of a linear system whose augmented matrix can be reduced to the matrix below.

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

Definition 2.3.17. The variables corresponding to pivot columns of a matrix are called **basic variables**, the other variables are called **free variables**.

Remark 2.3.18. Whenever a system is consistent, the solution set can be described explicitly by solving the *reduced* system of equations for the basic variables in terms of the free variables.

Example 2.3.19. Find the general solution of a system whose augmented matrix is reduced to

$$\left[\begin{array}{ccccc:c} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{array}\right]$$

Example 2.3.20. Determine the existence and uniqueness of the solutions to the system

$$\begin{aligned} 3x_2 - 6x_3 + 6x_4 + 4x_5 &= -5 \\ 3x_1 - 7x_2 + 8x_3 - 5x_4 + 8x_5 &= 9 \\ 3x_1 - 9x_2 + 12x_3 - 9x_4 + 6x_5 &= 15 \end{aligned}$$

Theorem 2.3.21. *A linear system is consistent if and only if the rightmost column of the augmented matrix is* not *a pivot column. If a linear system is consistent, then the solution set contains either*

(i) a unique solution, where there are no free variables, or

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

Using the theorem, and the rest of this section, we have the following procedure to find and describe all the solutions of a linear system.

Procedure 2.3.22 (Using Row Reduction to Solve a Linear System)**.**

1. Write the augmented matrix of the system.
2. Use the row reduction algorithm to write the matrix in echelon form. If the system is inconsistent, stop, there are no solutions; otherwise, go to the next step.
3. Use the row reduction algorithm to write the matrix in reduced echelon form.
4. Write the system of equations corresponding to the reduced matrix.
5. Solve each basic variable in terms of any free variables.

2.4 Rules for matrix operations

2.4. Key Ideas

- The (i,j) entry of AB is the dot product of row i of A with column j of B.
- An $m \times n$ matrix times an $n \times p$ matrix gives an $m \times p$ matrix, and uses mnp separate multiplications.
- $A(BC) = (AB)C$, but $AB \neq BA$ in general

Definition 2.4.1. If A is an $m \times n$ matrix (m rows and n columns), then the entry in the ith row and jth column of A, typically denoted a_{ij}, is called the (i,j)**-entry** of A. We write $A = [a_{ij}]$ using this notation. Columns of A are vectors in $\mathbb{R}^m$, usually denoted $\mathbf{a}_1, \dots \mathbf{a}_n$. We often write:

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}.$$

The **main diagonal** of $A = [a_{ij}]$ is the entries $a_{11}, a_{22}, a_{33}, \dots$. A **zero matrix** is one whose entries are all zero. The **identity matrix** is a $n \times n$ square matrix with ones on the main diagonal and zeros everywhere else, usually denoted I_n

Definition 2.4.2. Two matrices are **equal** if they have the same size and their corresponding entires are equal. If A and B are matrices of the same size, then the **sum** $A + B$ is the matrix whose entries are the sums of the corresponding entries in A and B.

Example 2.4.3. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ -4 & 5 & -6 \end{bmatrix}$, $B = \begin{bmatrix} 4 & 5 & 6 \\ 7 & -8 & 9 \end{bmatrix}$, and $C = \begin{bmatrix} 1 & 3 \\ 5 & -6 \end{bmatrix}$.
Find $A + B$, $B + A$, and $A + C$.

Definition 2.4.4. If r is a scalar and A is a matrix, then the **scalar multiple** rA is the matrix whose entries are r times the corresponding entries of A. Notationally, $-A$ stands for $(-1)A$, and $A - B = A + (-1)B$.

Example 2.4.5. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ -4 & 5 & -6 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 5 & 6 \\ 7 & -8 & 9 \end{bmatrix}$. Find $2B$ and $A - 2B$.

Theorem 2.4.6. *Let A, B, and C be matrices of the same size, and r and s be scalars.*

a. $A + B = B + A$

b. $(A + B) + C = A + (B + C)$

c. $A + 0 = A$.

d. $r(A + B) = rA + rB$

e. $(r + s)A = rA + sA$

f. $r(sA) = (rs)A$.

Definition 2.4.7. If A is an $m \times n$ matrix, and B is an $n \times p$ matrix with columns $\mathbf{b}_1, \ldots, \mathbf{b}_p$, then the product AB is the $m \times p$ matrix whose columns are $A\mathbf{b}_1, \ldots, A\mathbf{b}_p$. That is

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

Remark 2.4.8. If the number of columns of A doesn't match the number of rows of B, then the product AB is *undefined.*

Example 2.4.9. Compute AB and BA, when $A = \begin{bmatrix} 2 & 1 \\ -3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 5 & 1 \\ 2 & -8 & 3 \end{bmatrix}$.

Procedure 2.4.10 (Row-Column Rule for AB)**.** If the product AB is defined, then the (i, j)-entry of AB is the sum of the products of corresponding entries from row i of A and column j of B. If $(AB)_{ij}$ denotes the (i, j)-entry in AB, and A is an $m \times n$ matrix, then

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}$$

Example 2.4.11. With A and B from Example 2.4.9, compute AB using the row-column rule.

Theorem 2.4.12. *Let A be an $m \times n$ matrix, and let B and C have the right sizes so that the following sums and products are defined.*

a. $A(BC) = (AB)C$

b. $A(B + C) = AB + AC$

c. $(B + C)A = BA + CA$.

d. $r(AB) = (rA)B = A(rB)$ *(for any scalar r)*

e. $I_m A = A = AI_n$

Example 2.4.13. Let $A = \begin{bmatrix} 2 & -3 \\ -4 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \end{bmatrix}$, $C = \begin{bmatrix} 5 & -2 \\ 3 & 1 \end{bmatrix}$, and $D = \begin{bmatrix} 3 & 9 \\ 2 & 6 \end{bmatrix}$.

(a) Find AB and BA.

(b) Find AC.

(c) Find AD.

Watchout! 2.4.14. Here are some important warnings for matrix multiplication:

1. In general, $AB \neq BA$.
2. Cancellation laws *do not hold* for multiplication; $CA = CB$ (or $AC = BC$) does not mean $A = B$.
3. If $AB = 0$, this *does not mean* $A = 0$ or $B = 0$.

Definition 2.4.15. If A is an $m \times n$ matrix, the **transpose** of A is the $n \times m$ matrix, denoted A^T, whose columns are formed from the corresponding rows of A.

Example 2.4.16. Let $A = \begin{bmatrix} a & b & d \end{bmatrix}$, $B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \\ 6 & 2 \end{bmatrix}$, and $C = \begin{bmatrix} 5 & -2 & 1 & 3 \\ 3 & 1 & 2 & -6 \end{bmatrix}$.

Find A^T, B^T, and C^T.

Theorem 2.4.17. *Let A and B be matrices who are the right size for the following operations.*

a. $(A^T)^T = A$

b. $(A+B)^T = A^T + B^T$

c. $(rA)^T = rA^T$ *(for any scalar r)*

d. $(AB)^T = B^T A^T$

2.5 Inverse matrices

2.5. Key Ideas

- The inverse matrix gives $AA^{-1} = I$ and $A^{-1}A = I$.
- A is invertible if and only if it has n pivots
- If $A\mathbf{x} = 0$ for a nonzero vector $\mathbf{x}$, then A has no inverse
- The inverse of AB is $B^{-1}A^{-1}$, and $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.
- Reducing $\begin{bmatrix} A & I \end{bmatrix}$ to reduced row echelon form gives $\begin{bmatrix} I & A^{-1} \end{bmatrix}$.

Definition 2.5.1. An $n \times n$ matrix A is **invertible** if there is an $n \times n$ matrix C such that

$$CA = I \text{ and } AC = I, \text{ where } I = I_n \text{ is the identity matrix.}$$

In this case, C is called the **inverse** of A. A matrix that is *not* invertible is called a **singular matrix**, and an invertible matrix is called a **non-singular matrix**.

Remark 2.5.2. Suppose B and C were both inverses of A. Then

$$B = BI = B(AC) = (BA)C = IC = C.$$

It turns out, that if A has an inverse, it's unique. We call this unique inverse A^{-1}.

Example 2.5.3. Let $A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}$ and $C = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}$. Show that $C = A^{-1}$

Theorem 2.5.4. *Invertible matrices have the following three properties.*

1. *If A is an invertible matrix, then A^{-1} is invertible, and $(A^{-1})^{-1} = A$.*
2. *If A and B are $n \times n$ invertible matrices, then so is AB, and $(AB)^{-1} = B^{-1}A^{-1}$.*
3. *If A is an invertible matrix, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.*

Theorem 2.5.5. *Let* $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. *If* $ad - bc \neq 0$, *then* A *is invertible and*

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

If $ad - bc = 0$, *then* A *is not invertible.*

Example 2.5.6. Find the inverse of $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Theorem 2.5.7. *If* A *is an invertible* $n \times n$ *matrix, then for each* $\mathbf{b}$ *in* $\mathbb{R}^n$, *the equation* $A\mathbf{x} = \mathbf{b}$ *has the unique solution* $\mathbf{x} = A^{-1}\mathbf{b}$.

Example 2.5.8. Solve the system

$$\begin{aligned} x_1 + 2x_2 &= 1 \\ 3x_1 + 4x_2 &= 2 \end{aligned}$$

Theorem 2.5.9. *An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n. In this case, any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.*

Procedure 2.5.10. To find A^{-1}, row reduce the augmented matrix $\begin{bmatrix} A & I \end{bmatrix}$. If A is row equivalent to I, then $\begin{bmatrix} A & I \end{bmatrix}$ is row equivalent to $\begin{bmatrix} I & A^{-1} \end{bmatrix}$. Otherwise, A does not have an inverse.

Example 2.5.11. Find the inverse of $A = \begin{bmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix}$.

Example 2.5.12. Poké Balls are on sale everywhere! At the new pokéstore in Pallet Town, they're selling Poké Ball bundles of 1 Poké Ball, 2 Great Balls and 2 Ultra Balls, in Veridian, the bundles consist of 2 Poké Ball, 3 Great Balls and 3 Ultra Balls, and in Pewter City, bundles come with only 5 Great Balls. By the time he gets to Mount Moon, Ash has purchased a total of 5 Poké Ball, 15 Great Balls and 10 Ultra Balls. How many bundles did he buy in each town?

(a) Find a matrix equation $A\mathbf{x} = \mathbf{b}$ whose solution, $\mathbf{x}$, represents bundles in each town.

(b) Find $\mathbf{b}$ by using the inverse of A.

Chapter 3

Vector Spaces and Subspaces

3.1 Spaces of vectors

3.1. Key Ideas

- $\mathbb{R}^n$ contains all column vectors with n real components.
- $M_{2\times 2}$, $\mathbf{F}$, and $\{\mathbf{0}\}$ are vector spaces.
- A subspace containing $\mathbf{v}$ and $\mathbf{w}$ must contain all linear combinations $c\mathbf{v} + d\mathbf{w}$.
- The combinations of the columns of A form the column space $\mathrm{Col}(A)$. The column space is "spanned" by the columns.

Remember, $\mathbb{R}^n$ is the set of all vectors $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ with n entries. $\mathbb{R}^n$ has lots of nice properties

(i) The sum of $\mathbf{u}$ and $\mathbf{v}$, denoted $\mathbf{u} + \mathbf{v}$, is in $\mathbb{R}^n$.

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

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

(iv) There is a **zero** vector, $\mathbf{0}$ in $\mathbb{R}^n$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.

(v) For each $\mathbf{u}$ in $\mathbb{R}^n$, there is a vector $-\mathbf{u}$ in $\mathbb{R}^n$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

(vi) The scalar multiple of $\mathbf{u}$ by c, denoted $c\mathbf{u}$, is in $\mathbb{R}^n$.

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

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

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

(x) $1\mathbf{u} = \mathbf{u}$.

Today, you'll play with some other sets that enjoy a lot of the same properties.

Example 3.1.1. Let $\mathcal{M}_{m\times n}$ be the set of all $m \times n$ matrices with entries that are real numbers. Let's focus on $\mathcal{M}_{3\times 2}$, the set of all matrices with three rows and two columns.
Let A, B, and C be matrices in $\mathcal{M}_{3\times 2}$, and c and d be real numbers.

(a) True or false: $A + B$ is another matrix in $\mathcal{M}_{3\times 2}$.

(b) We know $A+B = B+A$ because the entries are real numbers. For example, in the $(1, 1)$-entry

$$A+B = \begin{bmatrix} a_{11} & * \\ * & * \\ * & * \end{bmatrix} + \begin{bmatrix} b_{11} & * \\ * & * \\ * & * \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & * \\ * & * \\ * & * \end{bmatrix} = \begin{bmatrix} b_{11}+a_{11} & * \\ * & * \\ * & * \end{bmatrix} = \begin{bmatrix} b_{11} & * \\ * & * \\ * & * \end{bmatrix} + \begin{bmatrix} a_{11} & * \\ * & * \\ * & * \end{bmatrix} = B+A.$$

Use the same logic to explain how we know $(A + B) + C = A + (B + C)$.

(c) How could it be possible for $A + B = A$?

(d) Let Z be the 3×2 zero matrix. What can you say about the entries of A and B if $A+B = Z$?

(e) By the same logic of part (b), explain how we know $c(A + B) = cA + cB$ and $c(dA) = (cd)A$.

(f) Is it possible for $cA = A$?

Example 3.1.2. Let $\mathbb{P}_n$ be the set of all polynomials of degree *up to* n. Focus on $\mathbb{P}_3$, the set of all polynomials of degree 1, 2, and 3, and all constants. Let f, g, and h be polynomials in $\mathbb{P}_3$, and c and d be real numbers.

(a) Could $f + g$ ever increase in degree? Could it decrease? What does this mean?

(b) We know $f+g = g+f$ because the coefficients are real numbers. For example, in the $(1,1)$-entry

$$f+g = (a_3x^3+a_2x^2+a_1x+a_0)+(b_3x^3+b_2x^2+b_1x+b_0) = \big((a_3+b_3)x^3+\cdots\big) = \big((b_3+a_3)x^3+\cdots\big) = g+f$$

Use the same logic to explain how we know $(f + g) + h = f + (g + h)$.

(c) How could it be possible for $f + g = h$?

(d) Let $z = 0$ be the "zero polynomial." What can you say about the entries of f and g if $f+g = z$?

(e) By the same logic of part (b), explain how we know $c(f + g) = cf + cg$ and $c(df) = (cd)f$.

(f) Is it possible for $cf = f$?

A lot of the theory in Chapters 1 and 2 used simple and obvious algebraic properties of $\mathbb{R}^n$, which we discussed in Section 1.3. Many other mathematical systems have the same properties. The properties we are interested in are listed in the following definition.

Definition 3.1.3. A **vector space** is a nonempty set V of objects, called *vectors*, on which two operations are defined: *addition* and *multiplication by scalars* (real numbers), subject to the ten axioms below. The axioms must hold for all $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ in V, and all scalars c and d.

1. The sum of $\mathbf{u}$ and $\mathbf{v}$, denoted $\mathbf{u}+\mathbf{v}$, is in V.
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.
3. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.
4. There is a **zero** vector, $\mathbf{0}$ in V such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is a vector $-\mathbf{u}$ in V such that $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$.
6. The scalar multiple of $\mathbf{u}$ by c, denoted $c\mathbf{u}$, is in V.
7. $c(\mathbf{u}+\mathbf{v})=c\mathbf{u}+c\mathbf{v}$.
8. $(c+d)\mathbf{u}=c\mathbf{u}+d\mathbf{u}$.
9. $c(d\mathbf{u})=(cd)\mathbf{u}$.
10. $1\mathbf{u}=\mathbf{u}$.

Example 3.1.4. We've just seen $\mathbb{R}^n$, the "matrix space" $\mathcal{M}_{m\times n}$, and the "polynomial space" $\mathbb{P}_n$ are vector spaces!

Definition 3.1.5. A **subspace** of a vector space is a subset H of V that has the properties:

(a) The zero vector of V is in H.

(b) H is closed under vector addition: for every $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u}+\mathbf{v}$ is in H.

(c) H is closed under scalar multiplication: for all $\mathbf{u}$ in H and scalar c, the vector $c\mathbf{u}$ is in H.

Example 3.1.6. The set $\{\mathbf{0}\}$ is a subspace of any vector space, called the **zero subspace**.

Example 3.1.7. Let $H = \left\{ \begin{bmatrix} a & a+b \\ 0 & b \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$. Show that H is a subspace of $M_{2\times 2}$.

Example 3.1.8. Given $\mathbf{v}_1$ and $\mathbf{v}_2$ in a vector space V. Show that $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ is a subspace of V.

Theorem 3.1.9. *If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are in a vector space V, then Span$\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is a subspace of V.*

Definition 3.1.10. We call $\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ the **subspace spanned** (or **generated**) by $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$. For any subspace H of V, a **spanning set** (or **generating set**) for H is a set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ such that $H = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$.

Definition 3.1.11. The **column space** of an $m \times n$ matrix A, written as $\operatorname{Col} A$, is the set of all linear combinations of the columns of A. If $A = [\ \mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n\ ]$, then

$$\operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}.$$

Remark 3.1.12. A typical vector in $\operatorname{Col} A$ can be written as $A\mathbf{x}$ for some $\mathbf{x}$, since the notation $A\mathbf{x}$ stands for a linear combination of the columns of A. In other words

$$\operatorname{Col} A = \{\mathbf{b} : \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \text{ in } \mathbb{R}^n\}$$

Example 3.1.13. Find $\operatorname{Col} A$ if

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

Example 3.1.14. Find a matrix A such that $W = \operatorname{Col} A$.

$$W = \left\{ \begin{bmatrix} 6a - b \\ a + b \\ -7a \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$$

Theorem 3.1.15. *The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.*

Theorem 3.1.16. *The column space of an $m \times n$ matrix A is all of $\mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^m$.*

3.1.1 Extra Problems

Example 3.1.17. Show $H = \{at^2 + at + a : a \text{ in } \mathbb{R}\}$ is a subspace of $\mathbb{P}_3$.

Example 3.1.18. Show that $\mathbb{D} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x^2 + y^2 < 1 \right\}$ is not a subspace of $\mathbb{R}^2$.

3.2 The null space, solutions to $A\mathbf{x} = \mathbf{0}$

3.2. Key Ideas

- The null space $\mathrm{Nul}(A)$ is a subspace of $\mathbb{R}^n$. It contains all solutions to $A\mathbf{x} = \mathbf{0}$.
- If $n > m$, then A has at least one column without pivots, giving a special solution. So there are nonzero vectors in $\mathrm{Nul}(A)$.

Definition 3.2.1. A system of linear equations is said to be **homogeneous** if it can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix, and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$.

Remark 3.2.2. The equation $A\mathbf{x} = \mathbf{0}$ *always* has at least one solution, namely $\mathbf{x} = \mathbf{0}$, called the **trivial solution**. We will be interested in finding **non-trivial solutions**, where $\mathbf{x} \neq \mathbf{0}$.

Example 3.2.3. Determine if the following homogeneous system has a nontrivial solution, and describe the solution set.

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

Proposition 3.2.4. *The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one free variable.*

Example 3.2.5. Describe all solutions to the homogeneous system

$$10x_1 - 3x_2 - 2x_3 = 0.$$

Example 3.2.6. What do our solutions to the previous two examples look like geometrically? In general, if A is a matrix with three columns, what could the solution set possibly look like?

Definition 3.2.7. The **null space** of an $m \times n$ matrix A, written as $\operatorname{Nul} A$, is the set of all solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$. In set notation,

$$\operatorname{Nul} A = \{\mathbf{x} : \mathbf{x} \text{ is in } \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\}.$$

Example 3.2.8. Let $A = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix}$, and $\mathbf{u} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$. Show that $\mathbf{u}$ is in $\operatorname{Nul} A$.

Example 3.2.9. For a matrix with 3 columns, what does $\operatorname{Nul}(A)$ look like geometrically?

Theorem 3.2.10. *The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$.*

Example 3.2.11. Find a spanning set for the null space of the matrix

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

Question 3.2.12. If a matrix A has more columns than rows, what can you say about $\text{Nul}(A)$?

How are the null space and column space of a matrix related? In the next example, we'll see that the two spaces are very different.

Example 3.2.13. Consider the following matrix.

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

(a) If the column space of A is a subspace of $\mathbb{R}^k$, what is k?

(b) If the null space of A is a subspace of $\mathbb{R}^k$, what is k?

(c) Find a nonzero vector in $\operatorname{Col} A$, and a nonzero vector in $\operatorname{Nul} A$.

(d) Is $\mathbf{u} = (3, -2, -1, 0)$ in $\operatorname{Nul} A$? Could it be in $\operatorname{Col} A$?

(e) Is $\mathbf{v} = (3, -1, 3)$ in $\operatorname{Col} A$? Could it be in $\operatorname{Nul} A$?

Contrast Between Nul A and Col A for an $m \times n$ Matrix A

Nul A	Col A
1. Nul A is a subspace of $\mathbb{R}^n$.	**1**. Col A is a subspace of $\mathbb{R}^m$.
2. Nul A is implicitly defined; that is, you are given only a condition ($A\mathbf{x} = \mathbf{0}$) that vectors in Nul A must satisfy.	**2**. Col A is explicitly defined; that is, you are told how to build vectors in Col A.
3. It takes time to find vectors in Nul A. Row operations on $[\, A \quad \mathbf{0} \,]$ are required.	**3**. It is easy to find vectors in Col A. The columns of A are displayed; others are formed from them.
4. There is no obvious relation between Nul A and the entries in A.	**4**. There is an obvious relation between Col A and the entries in A, since each column of A is in Col A.
5. A typical vector $\mathbf{v}$ in Nul A has the property that $A\mathbf{v} = \mathbf{0}$.	**5**. A typical vector $\mathbf{v}$ in Col A has the property that the equation $A\mathbf{x} = \mathbf{v}$ is consistent.
6. Given a specific vector $\mathbf{v}$, it is easy to tell if $\mathbf{v}$ is in Nul A. Just compute $A\mathbf{v}$.	**6**. Given a specific vector $\mathbf{v}$, it may take time to tell if $\mathbf{v}$ is in Col A. Row operations on $[\, A \quad \mathbf{v} \,]$ are required.
7. Nul $A = \{\mathbf{0}\}$ if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	**7**. Col $A = \mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^m$.
8. Nul $A = \{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	**8**. Col $A = \mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

3.4 The complete solution to $A\mathbf{x} = \mathbf{b}$

3.4. Key Ideas

- One particular solution $\mathbf{x}_p$ has all free variables equal to zero.
- The pivot variables are determined after the free variables are chosen.
- The full solution to $A\mathbf{x} = \mathbf{b}$ is the solution set to $A\mathbf{x} = \mathbf{0}$ shifted by x_p.

In this section, we will use vector notation to give explicit and geometric descriptions of solution sets of linear systems.

Example 3.4.1. Determine if the following homogeneous system has a nontrivial solution, and describe the solution set.

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

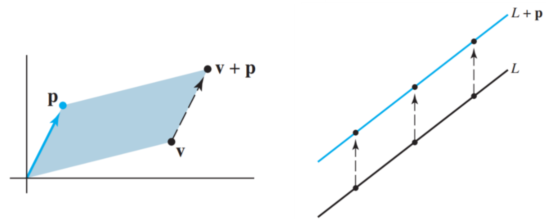
Proposition 3.4.2. *The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one free variable.*

Example 3.4.3. Describe all solutions of

$$\begin{aligned} 3x_1 + 5x_2 - 4x_3 &= 7 \\ -3x_1 - 2x_2 + 4x_3 &= -1 \\ 6x_1 + x_2 - 8x_3 &= -4 \end{aligned}$$

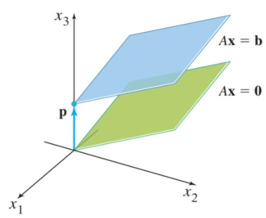
Example 3.4.4. Describe the relationship between the solutions to $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{x} = \mathbf{b}$ in the last two examples.

Definition 3.4.5. We can think of vector addition as *translation*. Given $\mathbf{p}$ and $\mathbf{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, the effect of adding $\mathbf{p}$ to $\mathbf{v}$ is to *move* v in a direction parallel to the line through $\mathbf{p}$ and $\mathbf{0}$. We say that $\mathbf{v}$ is **translated by** $\mathbf{p}$ to $\mathbf{v}+\mathbf{p}$. If each point on a line L is translated by a vector $\mathbf{p}$, the result is a line parallel to L.



For $t \in \mathbb{R}$, we call $\mathbf{p} + t\mathbf{v}$ **the equation of the line parallel to v through p.**

Theorem 3.4.6. *Suppose the equation* $A\mathbf{x} = \mathbf{b}$ *is consistent for some given* $\mathbf{b}$, *and let* $\mathbf{x}_p$ *be a solution. Then the solution set of* $A\mathbf{x} = \mathbf{b}$ *is the set of all vectors of the form* $\mathbf{w} = \mathbf{x}_p + \mathbf{v}_h$, *where* $\mathbf{v}_h$ *is any solution of the homogeneous equation* $A\mathbf{x} = \mathbf{0}$.



Procedure 3.4.7. To write a solution set in parametric vector form

1. Row reduce the augmented matrix to RREF
2. Express each basic variable in terms of any free variables
3. Write $\mathbf{x}$ as a vector whose entries depend on the free variables (if there are any)
4. Decompose $\mathbf{x}$ into a linear combination of vectors using free variables as parameters

Example 3.4.8. Describe and compare the solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$ if

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

3.5 Independence, basis, and dimension

3.5. Key Ideas

- The columns of A are independint if $\mathbf{x} = 0$ is the only solution to $A\mathbf{x} = \mathbf{0}$.
- The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ span a space if their combinations fill that space.
- A basis consists of linearly independent vectors that span the space, and every vector is a unique combination of vectors in that basis.
- All bases for a space have the same number of vectors, this number is called the dimension.
- The pivot columns are one basis for the column space. The dimension is $\mathbf{r}$.

Question 3.5.1 (From homework)**.** When does $A\mathbf{x} = \mathbf{b}$ have a unique solution for all $\mathbf{b}$?

Definition 3.5.2. An indexed set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be **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. S is **linearly dependent** if for some $c_1, \ldots, c_p$ not all zero

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

Example 3.5.3. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix}$.

Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly independent? If not, find a linear dependence relation.

Example 3.5.4. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$.

Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly independent? If not, find a linear dependence relation.

Definition 3.5.5. The columns of a matrix A are linearly independent if and only if every column of A is a pivot column.

Example 3.5.6. Determine if the columns of $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \\ 0 & 0 & 3 \end{bmatrix}$ are linearly independent.

Example 3.5.7. Determine if the following sets of vectors are linearly independent.

(a) $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$ (b) $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$

Proposition 3.5.8 (Sets of two vectors)**.** *A set of two vectors* $\{\mathbf{v}_1, \mathbf{v}_2\}$ *is linearly independent if and only if neither of the vectors is a multiple of the other.*

Question 3.5.9. What can you say about linear dependence/independence of a set of p vectors in $\mathbb{R}^n$ if $p > n$?

Theorem 3.5.10 (Too many vectors)**.** *If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set* $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ *in* $\mathbb{R}^n$ *is linearly dependent if* $p > n$.

Example 3.5.11. Determine by inspection (without matrices) if given sets are linearly dependent.

(a) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}$ (b) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$ (c) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix}$

Definition 3.5.12. An indexed set of vectors $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ in V is a **basis** for V if

(a) $\mathcal{B}$ is a linearly independent set, and

(b) $\mathcal{B}$ spans all of V; that is,

$$V = \operatorname{Span}(\mathcal{B}) = \operatorname{Span}\{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$$

The **dimension** of a vector space is the number of vectors in any basis for the space.

Example 3.5.13. Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ -6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 7 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}$. Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ a basis for $\mathbb{R}^3$?

Question 3.5.14. What is the dimension of $\mathbb{R}^n$?

Example 3.5.15. Which of the following is a basis for $\mathbb{R}^3$?

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \right\} \quad \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \right\} \quad \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix} \right\}$$

Remark 3.5.16. In one sense, a basis for V is a spanning set of V that is as small as possible. In another sense, a basis for V is a linearly independent set that is as large as possible.

Example 3.5.17. Find a basis for $\operatorname{Col} U$, where $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

Example 3.5.18. Below, A is row equivalent to U from the last example. Find a basis for $\operatorname{Col} A$.

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix}.$$

What is the dimension of $\operatorname{Col}(A)$?

Theorem 3.5.19. *The pivot columns of a matrix A form basis for* $\operatorname{Col} A$. *The dimension of* $\operatorname{Col}(A)$ *is the number of pivots.*

Example 3.5.20. Find a basis for $\operatorname{Nul} A$, where A is the same as the previous example:

$$A = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix}.$$

What is the dimension of $\operatorname{Nul}(A)$?

Theorem 3.5.21. *The dimension of* $\operatorname{Nul}(A)$ *is the number of free variables of* A*. In other words, the dimension of* $\operatorname{Nul}(A)$ *is the number of columns minus the number of pivots.*

Chapter 4

Orthogonality

4.1 Orthogonality

4.1. Key Ideas

- Subspaces V and W are orthogonal if every $\mathbf{v}$ in V is orthogonal to every $\mathbf{w}$ in W.

Question 4.1.1. How do we know when two vectors are *orthogonal* to each other?

Definition 4.1.2. Two vectors in $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$.

Definition 4.1.3. A vector $\mathbf{u}$ in $\mathbb{R}^n$ is orthogonal to a subspace V of $\mathbb{R}^n$ if and only if it's orthogonal to every vector in V.
In English...

Example 4.1.4. Show that $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ is orthogonal to the column space of

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

Definition 4.1.5. Two subspaces V and W of $\mathbb{R}^n$ are orthogonal if and only if every vector in V is orthogonal to every vector in W.
In English...

Example 4.1.6.

Definition 4.1.7. If A is an $m \times n$ matrix, the **transpose** of A is the $n \times m$ matrix, denoted A^T, whose columns are formed from the corresponding rows of A.

Example 4.1.8. Let $A = \begin{bmatrix} a & b & d \end{bmatrix}$, $B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \\ 6 & 2 \end{bmatrix}$, and $C = \begin{bmatrix} 5 & -2 & 1 & 3 \\ 3 & 1 & 2 & -6 \end{bmatrix}$.

Find A^T, B^T, and C^T.

Observation 4.1.9. Using this notation, sometimes it's convenient to write the dot product as matrix multiplication

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$$

4.2 Projections

4.2. Key Ideas

- projection of a point onto a line
- projection of a point onto a space

Question 4.2.1. Given a point P and a line ℓ, how can we find the closest point on ℓ to P?

Question 4.2.2. Given two vectors $\mathbf{b}$ and $\mathbf{a}$ in $\mathbb{R}^n$, suppose we're interested in the closest point in $\text{Span}(\mathbf{a})$ to $\mathbf{b}$. How is this like our line and point example above?

Definition 4.2.3. The **orthogonal projection** of $\mathbf{b}$ onto the span of $\mathbf{a}$ is

$$\mathbf{p} = \operatorname{proj}_{\mathbf{a}} \mathbf{b} = \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}} \mathbf{a} = \frac{\mathbf{a}^T \mathbf{b}}{\mathbf{a}^T \mathbf{a}} \mathbf{a} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}} \mathbf{b}.$$

The **error** is the vector $\mathbf{e} = \mathbf{b} - \mathbf{p}$. It's the vector that's perpendicular to $\mathbf{a}$ and has length equal to the distance from $\mathbf{b}$ to $\mathbf{p}$.

$$\text{The matrix } P = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T \mathbf{a}} \text{ is called the } \textbf{projection matrix.}$$

Example 4.2.4. Find the closest point on the line $y = x$ to $(1, 3)$.

Example 4.2.5. Find the projection matrix P onto the line through $\mathbf{a} = (1, 2, 2)$, and use this to find the point on the line closest to $(1, 1, 1)$.

Question 4.2.6. Given a collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_m$, and $\mathbf{b}$ in $\mathbb{R}^n$, how could we find the closest point to $\mathbf{b}$ in $\text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_m)$?

Definition 4.2.7. The combination $\mathbf{p} = \widehat{x}_1\mathbf{a}_1 + \cdots + \widehat{x}_m\mathbf{a}_m = A\widehat{\mathbf{x}}$ closest to $\mathbf{b}$ comes from

$$A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}.$$

The **projection** of $\mathbf{b}$ onto the subspace spanned by the $\mathbf{a}_i$ is

$$\mathbf{p} = A\widehat{\mathbf{x}} = A(A^T A)^{-1}A^T\mathbf{b}.$$

Again, $P = A(A^T A)^{-1}A^T$ is called the **projection matrix**. The **error** is the vector $\mathbf{e} = \mathbf{b} - \mathbf{p}$. Its length is equal to the distance from $\mathbf{b}$ to $\mathbf{p}$.

Example 4.2.8. Find the closest point to $\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ in the plane spanned by $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$.

Theorem 4.2.9 (Best Approximation)**.** Let V be a subspace of $\mathbb{R}^n$, $\mathbf{b}$ be any vector in $\mathbb{R}^n$, and $\mathbf{p} = \text{proj}_V(\mathbf{b})$. Then $\mathbf{p}$ is the closest point in V to $\mathbf{u}$, in the sense that

$$\|\mathbf{b} - \mathbf{p}\| < \|\mathbf{b} - \mathbf{v}\|$$

for all other $\mathbf{v}$ in V.

Example 4.2.10. Let $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ and find the closest point in W to $\mathbf{y}$ where

$$\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \qquad \text{and } \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

(see `https://www.geogebra.org/m/hybndwvh`)

Example 4.2.11. The distance from a point $\mathbf{u}$ in $\mathbb{R}^n$ to a subspace V is defined as the distance from $\mathbf{u}$ to the nearest point in V. Find the distance from $\mathbf{u}$ to $V = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ where

$$\mathbf{u} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \qquad \mathbf{v}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \qquad \text{and } \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

(see `https://www.geogebra.org/m/p9q2n95j`)

4.3 Least squares

4.3. Key Ideas

- The least squares solution $\widehat{\mathbf{x}}$ minimizes $E = \|A\mathbf{x} - \mathbf{b}\|^2$. This is the sum of squares of the errors in the m equations ($m > n$).
- The best $\widehat{\mathbf{x}}$ comes from the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$.
- To fit m points by a line $b = C + Dt$, the normal equations give C and D.
- The heights of the best line are $\mathbf{p} = (p_1, \ldots, p_m)$. The vertical distances to the data points are the errors $\mathbf{e} = (e_1, \ldots, e_m)$.
- If we try to fit m points by a combination of $n < m$ functions, the m equations $A\mathbf{x} = \mathbf{b}$ are generally unsolvable. The n equations $A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}$ give the least squares solution – the combination with the smallest mean square error.

Example 4.3.1. Find the closest line $y = a + bx$ to the points $(0, 6)$, $(1, 0)$ and $(2, 0)$.
(see `https://www.geogebra.org/m/aqtkdpbm`)

Observation 4.3.2. When $A\mathbf{x} = \mathbf{b}$ has no solution, solve $A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}$ instead. In this example, the heights of the best line are $\mathbf{p} = (p_1, \ldots, p_m)$. The vertical distances to the data points are the errors $\mathbf{e} = (e_1, \ldots, e_m)$.

Example 4.3.3. Find the parabola $y = a + bx + x^2$ through the points $(0, 6)$, $(1, 0)$ and $(2, 0)$. (see `https://www.geogebra.org/m/aqtkdpbm`)

Example 4.3.4. Find the plane through the points $(1, 1, 1)$, $(1, 1, 3)$ and $(1, 2, 3)$. Use the fact that a plane in $\mathbb{R}^3$ has the equation

$$x = ay + bz + c.$$

(see `https://www.geogebra.org/m/rgespmyf`)

Example 4.3.5. Find a plane that best fits $(1, 1, 1)$, $(2, 2, 1)$, $(1, 1, 3)$ and $(1, 2, 3)$. Use the fact that a plane in $\mathbb{R}^3$ has the equation

$$x = ay + bz + c.$$

(see `https://www.geogebra.org/m/rgespmyf`)

Definition 4.3.6. For an $m \times n$ matrix A and a vector $\mathbf{b}$, we cannot always get the error $\mathbf{e} = \mathbf{b} - A\mathbf{x}$ down to zero. When $\mathbf{e}$ is zero, $\mathbf{x}$ is an exact solution to $A\mathbf{x} = \mathbf{0}$. *When the length of* $\mathbf{e}$ *is as small as possible,* $\widehat{\mathbf{x}}$ *is a* **least squares solution**. This happens when $\widehat{\mathbf{x}}$ is a solution to $A^T A\widehat{\mathbf{x}} = A^T\mathbf{b}$. T

In English...

Example 4.3.7. You're taking average temperatures each month for the winter and spring months of 2020. You find the following data

Month	Average Temp C°
Jan	3
Feb	4
Mar	4
Apr	5
May	5

After sketching a scatterplot, you guess that a curve of the form $y = a + b\sqrt{x}$ will fit the data best. Find an equation of this type that best fits the data. How well does it fit?
(see `https://www.geogebra.org/m/maf7nqrv`)

Observation 4.3.8. The least squares solution $\widehat{\mathbf{x}}$ minimizes $E = \frac{1}{n}\|\mathbf{e}\|^2 = \frac{1}{n}\|A\mathbf{x} - \mathbf{b}\|^2$. This is the sum of squares of the errors in the m equations ($m > n$).