Linear Algebra Notes

Grant Talbert

2024 Spring

Abstract

Linear algebra introduces vector spaces and linear transformations on finite-dimensional vector spaces. It introduces linear systems, matrices, determinants, inner product spaces, and eigenvalues. This course is taught by Professor Thau-Ni Luu at AACC.
One thing anyone other than myself will notice while reading these notes - my notation for various things evolves as I get further into the material.

Contents

Chapter 1

Systems of Linear Equations

1.1 Introduction to Systems of Linear Equations

In this section, we will

- Recognize a linear equation in n variables.
- Find a parametric representation of a solution set.
- Determine whether a system of linear equations is consistent or inconsistent.
- Use back-substitution and Gaussian elimination to solve a system of linear equations.

Definition 1.1 (Linear Equation in n Variables). A linear equation in n variables $x_1, x_2, \ldots, x_n$ has the form

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

with $a_1, \ldots, a_n, b \in \mathbb{R}$.

In 2.7, b is the **constant term** and $a_1, \ldots, a_n$ are the **coefficients** of the linear equation. The number a_1 is the **leading coefficient** and the variable x_1 is the **leading variable**. It also follows that any multivariate linear function may not have terms containing the product of any number of variables, trigonometric functions, exponential functions, logarithmic functions, and any powers of variables other than 1.

Definition 1.2 (Solutions and Solution Sets). The **solution** of a linear equation in n variables is a sequence of n real numbers $s_1, \ldots, s_n$ that satisfy the equation for $(x_1, \ldots, x_n) = (s_1, \ldots, s_n)$. The set of all solutions to a linear equation is known as its **solution set**, and is often given with a **parametric representation**.

An example of a parametric representation of the solution set of a linear function, as seen in 2.10, can be seen here.

Example. Consider the linear relationship

$$3x_1 - 2x_2 = 6$$

Parametrize and give the solution set of this relationship.

Solution.

$$\Rightarrow 3x_1 = 2x_2 + 6$$
$$\Rightarrow x_1 = \frac{2}{3}x_2 + 2$$
$$\text{Let } t = x_2, t \in \mathbb{R}$$
$$\Rightarrow x_1 - \frac{2}{3}t + 2$$
$$\therefore \left\{ (x_1, x_2) \mid x_1 = \frac{2}{3}t + 2, x_2 = t, t \in \mathbb{R} \right\}$$

A more general way of expressing this set is as:

$$(x_1, x_2) = \left(\frac{2}{3}t + 2, t \right)$$

□

Definition 1.3 (System of Linear Equations)**.** A system of m linear equations in n variables is a set of m equations, each of which is linear in the same n variables,

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = b_3 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = b_m \end{cases}$$

The solution set of an equation is a set of numbers $s_1, s_2, s_3, \ldots, s_n$ that satisfies all equations.

Notation (Solution Set)**.** I've decided to use $\mathcal{S}$ as a shorthand for the solution set of an arbitrary system of equations.

Remark (Number of Solutions of a System of Linear Equations)**.** For any system of equations, exactly one of the following holds.

1. The system has exactly one solution. (Consistent system) $n(\mathcal{S}) = 1$
2. The system has infinitely many solutions. (Consistent system) $n(\mathcal{S}) = \aleph_1$
3. The system has no solutions. (Inconsistent system) $n(\mathcal{S}) = 0$

Let $E : \mathcal{S} \longrightarrow \mathbb{R}^n$ be a system of equations in n variables $x_1, x_2, \ldots, x_n$ with solution set $\mathcal{S} \subseteq \mathbb{R}^n$. From the previous remark, we see that any consistent system is a system of linear equations will satisfy the statement

$$\exists (s_1, s_2, \ldots, s_n) \in \mathbb{R}^n ((s_1, s_2, \ldots, s_n) \in \mathcal{S})$$

By converse and also from , we see that any inconsistent system of linear equations satisfies

$$\nexists (s_1, s_2, \ldots, s_n) \in \mathbb{R}^n ((s_1, s_2, \ldots, s_n) \in \mathcal{S})$$

It is thus implied that for an inconsistent system, $\mathcal{S} = \varnothing$.

Theorem 1.4 (Gaussian Elimination). Two systems of linear equations may be considered equivalent when they have the same solution set, that is

$$E_1 : \mathcal{S} \to \mathbb{R}^n \wedge E_2 : \mathcal{S} \to \mathbb{R}^n \implies E_1 \equiv E_2$$

From this definition, we can perform the following operations on a system while maintaining an equivalent system:

1. Interchange two equations
2. Multiply an equation by a nonzero constant
3. Add a multiple of an equation to another equation

Proof.

Let $E : \mathcal{S} \to \mathbb{R}^n$ be the system of m linear equations of n variables with solution set $\mathcal{S}$ given by

$$E(f_1, f_2, \ldots, f_m) = \begin{cases} f_1(x_1, x_2, \ldots, x_n) \\ f_2(x_1, x_2, \ldots, x_n) \\ f_3(x_1, x_2, \ldots, x_n) \\ \vdots \\ f_m(x_1, x_2, \ldots, x_n) \end{cases}$$

Note: it is not implied that $n = m$, but it is also not implied that $n \neq m$.

Proof of condition 1:
The proof is obvious and is left as an exercise to the reader.

Proof of condition 2:

$$\text{Let } c \in \mathbb{R} \setminus \{0\}$$

$$c \cdot f_i \Rightarrow ca_{i1}x_1 + ca_{i2}x_2 + ca_{i3}x_3 + \cdots + ca_{in}x_n = cb_i$$

$$\Rightarrow c(a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + ca_{in}x_n) = cb_i$$

$$\Rightarrow a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + ca_{in}x_n = \frac{cb_i}{c}$$

$$\Rightarrow a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + ca_{in}x_n = b_i$$

$$\therefore \forall (c \in \mathbb{R}, f_i, i \in \setminus \{r | r \in \mathbb{N} \wedge r \leq n\})(c \cdot f_i \equiv f_i)$$

Proof of condition 3:

$$\forall (i, j)(f_i \equiv f_j) \because (E : \mathcal{S} \to \mathbb{R}^n \Rightarrow (f_i : X \to \mathbb{R} \Rightarrow X \subseteq \mathcal{S}))$$

$$\Rightarrow \forall (x_1, \ldots, x_n) \in \mathcal{S}((a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + a_{in}x_n = b_i) \equiv (a_{j1}x_1 + _{j2}x_2 + a_{j3}x_3 + \cdots + a_{jn}x_n = b_j))$$

$$\Rightarrow \forall (x_1, \ldots, x_n) \in \mathcal{S}(a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + a_{in}x_n + b_j = b_i + b_j)$$

$$\Rightarrow \forall (x_1, \ldots, x_n) \in \mathcal{S}(a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3 + \cdots + a_{in}x_n + a_{j1}x_1 + _{j2}x_2 + a_{j3}x_3 + \cdots + a_{jn}x_n = b_i + b_j)$$

$$\text{If } \exists (s_1, s_2, \ldots, s_n) \in \mathbb{R}^n)((s_1, s_2, \ldots, s_n) \in \mathcal{S}) \text{ then the above statements hold.}$$

$$\text{If } \exists (s_1, s_2, \ldots, s_n) \in \mathbb{R}^n)((s_1, s_2, \ldots, s_n) \in \mathcal{S}) \Rightarrow \mathcal{S} = \varnothing \therefore f_i \neq f_j$$

The process of Gaussian elimination necessitates the assumption that $\mathcal{S} \neq \varnothing$,and thus if there does not exist a solution to the system of equations, Gaussian elimination will inevitably lead to a contradiction, indicating that a solution does not exist. If a solution does exist, then the above statements prove that the process of elimination preserves the solution set by adding equal values to both sides of the equation, which happen to be expressed in terms of variables. ■

1.2 Gaussian Elimination and Gauss-Jordan Elimination

Definition 1.5 (Matrices). If $n, m \in \mathbb{Z}$, then an $m \times n$ matrix is a rectangular array

$$\underbrace{\left.\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_32 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}\right\} m}_{n}$$

in which each entry a_{ij} is located in the ith row and the jth column.
A matrix satisfying $m = n$ is a square matrix of order n, and the entries $a_{11}, a_{22}, \ldots, a_{nn}$ constitute the *main diagonal*.

Corollary 1.6 (Matrix Representation of a Linear System of Equations). For a system of equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = b_3 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = b_m \end{cases}$$

we give the *augmented matrix* of the system as

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} & b_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} & b_m \end{bmatrix}$$

and we give the *coefficient matrix* of the system as

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

For example, we can consider the system:

$$\begin{gathered} x + y + z = 2 \\ -x + 3y + 2z = 8 \\ 4x + y + 0z = 4 \end{gathered}$$

For this system, we have the coefficient matrix:

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

and the augmented matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ -1 & 3 & 2 & 8 \\ 4 & 1 & 0 & 4 \end{bmatrix}$$

Theorem 1.7 (Elementary Row Operations). Gaussian elimination allows for *elementary row operations* to be taken on an augmented matrix. These operations include:

1. Interchange two rows.
2. Multiply a row by a nonzero constant.
3. Add a multiple of a row to another row.

For an example, we can consider the above augmented matrix.

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ -1 & 3 & 2 & 8 \\ 4 & 1 & 0 & 4 \end{bmatrix} \xrightarrow{R_1+R_2\to R_2} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 4 & 3 & 10 \\ 4 & 1 & 0 & 4 \end{bmatrix} \xrightarrow{-4R_1+R_3\to R_3} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 4 & 3 & 10 \\ 0 & -3 & -4 & -4 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{4}R_2\to R_2} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & \frac{3}{4} & \frac{5}{2} \\ 0 & -3 & -4 & -4 \end{bmatrix} \xrightarrow{3R_2+R_3\to R_3} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & \frac{3}{4} & \frac{5}{2} \\ 0 & 3 & -\frac{7}{4} & \frac{7}{2} \end{bmatrix} \xrightarrow{-\frac{7}{4}R_3\to R_3} \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & \frac{3}{4} & \frac{5}{2} \\ 0 & 3 & 1 & -2 \end{bmatrix} \leftrightarrow \begin{cases} x+y+z=2 \\ y+\frac{3}{4}z=\frac{5}{2} \\ z=-2 \end{cases}$$

This form of a matrix is known as row-echelon form.

Definition 1.8 (Row-Echelon and Reduced Row-Echelon Form). A matrix in *row-echelon* form has the properties:

1. Any row consisting entirely of zeros occurs at the bottom of the matrix.
2. For each ropw that does not consist entirely of zeros, the first nonzero entry is a 1 (known as a **leading 1**).
3. For two successive (nonzero) rows, the leading 1 in the higher row is further to the left than the leading 1 in the lower row (it can be any distance to the left, not just 1 space).

A matrix in row-echelon form is in *reduced row-echelon form* when every column that has a leading 1 has zeros in every spot above and below its leading 1 (not just in the rows directly above/below it, but ALL rows)

Remark. The procedure for using Gaussian elimination with back-substitution is summarized below.

1. Write the augmented matrix of the system of equations.
2. Use elementary row operations to rewrite the matrix in row-echelon form.
3. Write the system of linear equations corresponding to the matrix in row-echelon form, and use back-substitution to find the solution.

Remark. Gauss-Jordan uses the processes allowed under Gaussian elimination to put a matrix into reduced row-echelon form, rather than row-echelon form.

1.2.1 Homogeneous Systems of Linear Equations

Definition 1.9 (Homogeneous Systems of Linear Equations)**.** A system of equations is *homogeneous* if and only if a system of m equations in n variables has the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n = 0 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n = 0 \end{cases}$$

Simply, a homogeneous system of equations has all constant terms equal to 0.

Corollary 1.10 (Homogeneous Systems)**.** Every homogeneous system of equations in n variables has the trivial solution $\underbrace{(0, 0, 0, \ldots, 0)}_{n \text{ 0s}}$. For a system of m equations in n variables, any system with $m < n$ has infinitely many solutions with $m - n$ free variables.

1.3 Applications of Systems of Linear Equations

Polynomial Curve Fitting

$$\text{Let } \exists((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \in \mathbb{R}^2)((x_i, y_i) \in \mathcal{D} \land (i \neq j \Rightarrow (x_i, y_i) \neq (x_j, y_j)))$$

Let this set $\mathcal{D}$ represent a collection of data points. It will follow that there exists precisely 1 polynomial function $p(x)$ of degree $n - 1$ such that

$$p : \mathbb{R} \to X \land \mathcal{D} \subseteq X$$

where p is surjective, that is, $\forall(x \in \mathbb{R})(p(x) \in X) \land \forall(y)(\exists(x \in \mathbb{R})(p(x) = y) \Rightarrow y \in X)$. The function p can be given as

$$p(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_{n-1}x^{n-1}$$

Definition 1.11 (Polynomial Curve Fitting)**.** For an arbitrary data set $\mathcal{D}$, the process by which a function $p : X \to Y$ such that $\mathcal{D} \subseteq X \times Y$ is derived is known as *polynomial curve fitting.*

Remark. We will notice that a polynomial of infinite degree

$$p : \mathbb{R} \to Y$$

$$x \mapsto \sum_{n=0}^{\infty} a_i(bx)^n$$

will converge for a subset $x \in \left[-\frac{1}{b}, \frac{1}{b}\right]$. Allow us to define a data set $\mathcal{D}$ of data points. We need to define a few constraints on $\mathcal{D}$:

$$\forall((x_i, y_i), (x_j, y_j) \in \mathcal{D})(i \neq j \Rightarrow x_i \neq x_j)$$

$$n(\mathcal{D}) = \aleph_0$$
$$\exists(a, b \in \mathbb{R})(\forall((x_k, y_k) \in \mathcal{D})((a > x_k) \wedge (b < x_k)))$$

This final constraint is necessary since for some $I \subsetneq \mathbb{R}$ defining the interval of convergence for our polynomial of infinite degree, the same must hold for I.
It should follow that
$$\exists(p : \mathbb{R} \to Y)(\mathcal{D} \subset \mathbb{R} \times Y)$$

I have no idea why this is true but the proof seems to follow from the original proof of polynomial curve fitting.

Example: consider the polynomial that passes through the points $(-1, 3)$, $(0, 0)$, $(1, 1)$, and $(4, 58)$. We have a function $p(x) = ax^3 + bx^2 + cx + d$. We plug in our point values and set up the system:

$$\begin{cases} a(-1)^3 + b(-1)^2 + c(-1) + d = 3 \\ d = 0 \\ a + b + c + d = 1 \\ a(4)^3 + b(4)^2 + c(4) + d = 58 \end{cases}$$

We give the matrix representation

$$\begin{bmatrix} -1 & 1 & -1 & 1 & 3 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 64 & 16 & 4 & 1 & 58 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -\frac{3}{2} \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\therefore p(x) = \frac{1}{2}x^3 + 2x^2 - \frac{3}{2}x$$

Matrix models of systems of equations can be used to solve partial fraction decomposition.

Exercise. Use a matrix to complete partial fraction decomposition on the expression

$$\frac{12}{(x+1)(4x+1)}$$

Solution.

$$\frac{12}{(x+1)(4x+1)} = \frac{A}{x+1} + \frac{B}{4x+1}$$
$$\implies 12 = (4x+1)A + (x+1)B$$
$$\implies \begin{matrix} 4Ax + Bx = 0 \\ A1 + B1 = 12 \end{matrix} \sim \begin{bmatrix} 4 & 1 & 0 \\ 1 & 1 & 12 \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 16 \end{bmatrix}$$
$$\therefore A = -4 \wedge B = 16$$

□

Chapter 2

Introduction to Matrices

Note: $\mathbb{F}$ depicts an arbitrary scalar field herein. This class was taught with $\mathbb{F} := \mathbb{R}$ in mind, so these properties do not necessarily hold for any arbitrary $\mathbb{F}$.

2.1 Operations with Matrices

Definition 2.1 (Equality of Matrices). Two $m \times n$ matrices are equal if and only if $a_{ij} = b_{ij} \forall i, j$

Definition 2.2 (Matrix Addition). For $A = [a_{ij}]$ and $B = [b_{ij}]$ both being $m \times n$ matrices, we have the matrix $A + B$ as the matrix with each (i, j) entry equal to $a_{ij} + b_{ij}$.

Definition 2.3 (Scalar Multiplication). For $A = [a_{ij}]$ and $c \in \mathbb{F}$, then $cA = [c \cdot a_{ij}]$.

Definition 2.4 (Matrix Multiplication). For For $A = [a_{ij}]$ as an $m \times n$ matrix and $B = [b_{ij}]$ as an $n \times p$ matrix, the product AB is the $m \times p$ matrix $AB = [c_{ij}]$, where we give

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

$$= a_{i1} b_{1j} + \cdots + a_{in} + b_{nj}$$

We can give systems of linear equations as $A\vec{x} = \vec{b}$, where A is the coefficient matrix, $\vec{x}$ is the column vector with entries x_i representing variables, and $\vec{b}$ is the column vector of constant terms. We find that

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

gives a valid representation of systems of equations in a matrix algebra. This is due to the fact that we can partition matrices into a linear combination:

$$A\vec{x} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \vec{b}$$

Definition 2.5 (Linear Combination). If $a_i \in \mathbb{F}$ and $\vec{x}_i$ is a column vector, than a linear combination of n vectors is of the form

$$\sum_{i=1}^{n} a_i \vec{x}_i$$

2.2 Properties of Matrix Operations

Theorem 2.6 (Properties of Matrix Addition and Scalar Multiplication). If A, B, and C are $m \times n$ matrices, and $c, d \in \mathbb{F}$, the following properties are true.

$$A + B = B + A$$

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

$$(cd)A = c(dA)$$

$$1A = A$$

$$c(A + B) = cA + cB$$

$$(c + d)A = cA + dA$$

Definition 2.7 (Zero Matrix). O_{mn} denotes a matrix of dimension $m \times n$ with all entries equaling 0.

Theorem 2.8 (Properties of Zero Matrices). If A is an $m \times n$ matrix and $c \in \mathbb{F}$, we have

$$A + O_{mn} = A$$

$$A + (-A) = O_{mn}$$

$$cA = O_{mn} \Leftrightarrow c = 0 \lor A = O_{mn}$$

Theorem 2.9 (Properties of Matrix Multiplication). For A, B, C being matrices with dimensions such that multiplication operations are defined between them, and $c \in \mathbb{F}$, we have

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

$$A(B+C) = AB + AC$$
$$(A+B)C = AC + BC$$
$$c(AB) = (cA)B = A(cB)$$

Remark.

$$AB = C \not\Rightarrow BA = C$$

Definition 2.10 (Identity Matrix). An $n \times n$ matrix with all entries 1 along the main diagonal and other entries equal to 0 is known as the Identity matrix of size n, denoted I_n.

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

Theorem 2.11 (Properties of the Identity Matrix). For A as a matrix of size $m \times n$, we have

$$AI_n = A$$

$$I_m A = A$$

Definition 2.12 (The Transpose of a Matrix). The transpose of a matrix A, given as A^T, is formed by giving the rows of A as columns, and vice versa.

Example:

$$A = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \Longrightarrow A^T = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$$

Theorem 2.13 (Properties of Transposes). If A and B are matrices of dimensions such that operations are defined, and $c \in \mathbb{F}$, we have

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

$$(A+B)^T = A^T + B^T$$
$$(cA^T) = c(A^T)$$
$$(AB)^T = B^T A^T$$

2.3 The Inverse of a Matrix

Definition 2.14 (Inverse Matrices). An $n \times n$ matrix A is invertible if and only if

$$\exists B \in \mathbb{R}^{n\times n}(AB = BA = I_n)$$

where $\mathbb{R}^{n\times n}$ denotes the set of all $n \times n$ matrices with real entries. It follows from the above proposition that the matrix B is the multiplicative *inverse* of A.
A matrix such that $\nexists B \in \mathbb{R}^{n\times n}(AB = BA = I_n)$ is known as noninvertible or singular.

Proposition 2.15.

$$\nexists(A \in \mathbb{R}^{m\times n} \wedge m \neq n)(\exists B \in \mathbb{R}^{n\times m}(AB = BA = I))$$

Theorem 2.16 (Uniqueness of an Inverse). If A is an inverse matrix, then its inverse is unique:

$$\exists! B \in \mathbb{R}^{n\times n}(AB = BA = I_n)$$

We denote the inverse as A^{-1}.

We can find the inverse of a matrix via Gauss-Jordan elimination. If we take A to be a square matrix of order n, then we construct the matrix $\begin{pmatrix} A & I_n \end{pmatrix}$. We perform Gauss-Jordan elimination until we have the resulting matrix $\begin{pmatrix} I_n & A^{-1} \end{pmatrix}$.

Theorem 2.17 (Inverse of a 2×2 Matrix). For some matrix $A \in \mathbb{R}^{2\times 2}$, then $\exists A^{-1} \Leftrightarrow ad - bc \neq 0$. We also have

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

Theorem 2.18 (Properties of Inverses). If A is an invertible matrix, $k \in \mathbb{Z}^+$, and $c \in \mathbb{F} \setminus \{0\}$, then A^{-1}, A^k, cA, A^T are invertible and the following statements are true:

$$(A^{-1})^{-1} = A$$

$$(A^k)^{-1} = \underbrace{A^{-1}A^{-1}\cdots A^{-1}}_{\text{k factors}} = (A^{-1})^k$$

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

$$(A^T)^{-1} = (A^{-1})^T$$

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

Defining multiplicative inverses over this matrix algebra allows us to define cancellation properties, akin to how we would see in the real numbers.

Theorem 2.19 (Cancellation Properties). If C is an invertible matrix, the following are true.

$$AC = BC \implies A = B$$

$$CA = CB \implies A = B$$

Theorem 2.20 (Square Systems of Linear Equations). If A is invertible, then the system of linear equations $A\vec{x} = \vec{b}$ has the unique solution

$$\vec{x} = A^{-1}\vec{b}$$

2.4 Elementary Matrices

Definition 2.21 (Elementary Matrix). An $n \times n$ matrix is an **elementary matrix** when it can be obtained from the identity matrix I_n with a single elementary row operation.

Theorem 2.22 (Representing Elementary Row Operations). Let E be the elementary matrix given by performing an elementary row operation on I_n. If that same operation is performed on some $m \times n$ matrix A, then the resulting matrix is equivalent to the product EA.

Remark. This is how we formalize the methods of gaussian elimination, however it's generally better to compute via gaussian elimination than with multiplication by elementary matrices.

Definition 2.23 (Row Equivalence). Two $n \times n$ matrices A, B are row equivalent when there exists a finite number of elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$B = E_k E_{k-1} \cdots E_2 E_1 A$$

$$B = \left(\prod_{i=1}^{k} E_i \right) A$$

Theorem 2.24 (Elementary Matrices are Invertible). If E is an elementary matrix, then $\exists E^{-1}$ and is *also* an elementary matrix.
Furthermore, for any elementary matrix which differs from I_n by an interchange of rows, it is its own inverse.
For any matrix which differs from I_n by a single number on the main diagonal not equaling 1, it's inverse is the same matrix except with the multiplicative inverse of that component.
For any matrix which differs from I_n by an extra nonzero component, its inverse is the same matrix but with the extra component's additive inverse.

Theorem 2.25 (Propert of Inverse Matrices). A square matrix A is invertble if and only if it can be written as the product of elementary matrices.

Theorem 2.26 (Conditions Equivalent to Invertibility). If A is an $n \times n$ matrix, the following statements are equivalent:

$$A \text{ is invertible.}$$

$$A\vec{x} = \vec{b} \text{ has a unique solution for every } n \times 1 \text{ column vector } \vec{b}.$$

$$A\vec{x} = \vec{0} \text{ has only the trivial solution.}$$

$$A \text{ is row-equivalent to } I_n.$$

$$A \text{ can be written as the product of elementary matrices.}$$

Definition 2.27 (LU-Factorization). If the $n \times n$ matrix A can be written as the product of a lower triangular matrix L and an upper triangular matrix U, then $A = LU$ is an LU-Factorization of A.

Remark. The LU-Factorization, if it exists, is not unique.

The method of LU factorization is as follows:

1. Find a set of elementary matrices $E_1, E_2, \ldots, E_k$ where there are no row interchanges, such that
$$E_k \cdots E_2 E_1 A = U$$
where A is the original matrix and U is upper-triangular.
2. Solve for A:
$$A = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1} U = A$$
3. set $E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1} = L$

We now have $A = LU$.

Theorem 2.28 (Solving Systems of Equations with LU-Factorization). Let $\vec{y}$ be a column vector with new variables $y_1, y_2, \ldots, y_n$ be defined. Set

$$L\vec{y} = \vec{b}$$

where L is the lower triangular matrix of the coefficient matrix of some system, and $\vec{b}$ is the vector representing constants. Solve for $\vec{y}$. It will then follow that

$$\vec{y} = U\vec{x}$$

Solve for $\vec{x}$ to solve the system.

2.5 Linear Regressions

Consider a set of data $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. We can denote the linear regression line as $f(x) = a_0 + a_1 x$. It follows that we can use our data points to define a system of n equations:

$$\begin{cases} y_1 &= a_0 + a_1 x_1 + e_1 \\ y_2 &= a_0 + a_1 x_2 + e_2 \\ & \vdots \\ y_n &= a_0 + a_1 x_n + e_n \end{cases}$$

where e_i denotes the error for data point i. We define the following matrices:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad A = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \quad E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

The solution to this system is given as $A = (X^T X)^{-1} X^T Y$. The error is $E^T E$. Use a fucking calculator to find this.

Chapter 3

Determinants

3.1 The Determinant of a Matrix

Definition 3.1 (Determinant of a 2×2). The determinant of a 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is given as

$$\det(A) = |A| = a_{11}a_{22} - a_{12}a_{21}$$

Definition 3.2 (Minors and Cofactors). If A is a square matrix, then the minor M_{ij} of the entry a_{ij} is the determinant of the matrix obtained by deleting the ith row and jth column of A. The cofactor C_{ij} of the entry a_{ij} is $C_{ij} = (-1)^{i+j}M_{ij}$.

Definition 3.3 (Determinant of an $n \times n$ Matrix). IF A is a square matrix of order $n \geq 2$, then the **determinant** of A is the sum of the entries in the first row multiplied by their respective cofactors. That is,

$$\det(A) = |A| = \sum_{j=1}^{n} a_{1j}C_{1j}$$

Theorem 3.4 (Expansion by Cofactors). Let A be a square matrix of order n. Let $i, j \in [1, n]$. This theorem extends the determinant of A to be any of the following, for any i or j:

$$\det(A) = |A| = \sum_{j=1}^{n} a_{ij}C_{ij}$$

(Expansion across ith row)

$$\det A = |A| = \sum_{i=1}^{n} a_{ij}C_{ij}$$

(Expansion across jth column)

We prefer to choose the row or column that will make the computation the easiest, that is the one that has the most coefficients of the cofactors equal to 0.

Consider this triangle matrix as an example:

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

$$\det(A) = 6 \begin{vmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 2 & 0 & 3 \end{vmatrix} = 6 \left(5 \begin{vmatrix} 4 & 0 \\ 0 & 3 \end{vmatrix} \right) = 6 \cdot 5 \cdot 4 \cdot 3 = 360$$

We chose to compute the determinant from the top rows since they only had one nonzero term, thus we only had to consider one cofactor. This leads us into the next theorem.

Theorem 3.5 (Determinant of a Triangular Matrix)**.** If A is any triangular matrix of degree n, then its determinant is the product of its entries on its main diagonal.

$$\det(A) = |A| = a_{11}a_{22} \cdots a_{nn}$$

3.2 Determinants and Elementary Row Operations

Let $A \coloneqq \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Consider the following:

- Obtain B by swapping the rows of A:

$$B = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$$

$$\det(B) = cb - ad = -(ad - bc) = -\det(A)$$

- Obtain B by adding k times the first row to the second row.

$$B = \begin{bmatrix} a & b \\ ka + c & kb + d \end{bmatrix}$$

$$\det(B) = a(kb + d) - b(ka + c) = akb + ad - bka - bc = ad - bc = \det(A)$$

- Obtain B by multiplying the second row of A by a nonzero constant:

$$B = \begin{bmatrix} a & b \\ kc & kd \end{bmatrix}$$

$$\det(B) = akd - bkc = k(ad - bc) = k \det(A)$$

Theorem 3.6 (Elementary Row Operations on Determinants)**.** Let A and B be square matrices of order n.
When B is obtained from adding A by interchanging two rows of A:

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

When B is obtained from A by adding a multiple of a row of A to another row of A:

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

When B is obtained from A by multiplying a row of A by a nonzero constant:

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

A useful way to apply this theorem is by converting matrices into triangular matrices to obtain the determinant of the original matrix.

Example.

$$A := \begin{bmatrix} 0 & 2 & 5 & 4 \\ 1 & 1 & 3 & 5 \\ 1 & 1 & 3 & 4 \\ -2 & -2 & 1 & 6 \end{bmatrix} \xrightarrow{R_3 - R_2} \underbrace{\begin{bmatrix} 0 & 2 & 5 & 4 \\ 1 & 1 & 3 & 5 \\ 0 & 0 & 0 & -1 \\ -2 & -2 & 1 & 6 \end{bmatrix}}_{\det(A) \to \det(A)} \xrightarrow{R_3 + 2R_2} \underbrace{\begin{bmatrix} 0 & 2 & 5 & 4 \\ 1 & 1 & 3 & 5 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 7 & 16 \end{bmatrix}}_{\det(A) \to \det(A)}$$

$$\xrightarrow{R_3 \leftrightarrow R_4} \underbrace{\begin{bmatrix} 0 & 2 & 5 & 4 \\ 1 & 1 & 3 & 5 \\ 0 & 0 & 7 & 16 \\ 0 & 0 & 0 & -1 \end{bmatrix}}_{\det(A) \to -\det(A)} \xrightarrow{R_1 \leftrightarrow R_2} \underbrace{\begin{bmatrix} 1 & 1 & 3 & 5 \\ 0 & 2 & 5 & 4 \\ 0 & 0 & 7 & 16 \\ 0 & 0 & 0 & -1 \end{bmatrix}}_{-\det(A) \to \det(A)}$$

$$\begin{vmatrix} 1 & 1 & 3 & 5 \\ 0 & 2 & 5 & 4 \\ 0 & 0 & 7 & 16 \\ 0 & 0 & 0 & -1 \end{vmatrix} = (1)(2)(7)(-1) = -14$$

$$\therefore \det(A) = -14$$

Theorem 3.7 (Determinants and Elementary Column Operations)**.** Elementary operations performed on columns rather than rows are known as **elementary column operations**. Similarly, two matrices are column-equivalent when a finite set of elementary column operations transforms one into the other. 3.6 remains valid when elementary column operations are considered, rather than rows.

Theorem 3.8 (Conditions that Yield a Zero Determinant). If A is a square matrix and one of the following is true for A, $\det(A) \equiv 0$.

- An entire row or column consists of 0s.
- Two rows or columns are equivalent.
- One row or column is a multiple of another.

3.3 Properties of Determinants

Lemma 3.9. If E is an elementary matrix, show that $\det(EB) = \det(E)\det(B)$:
proof:
Let E be obtained by interchanging rows of I:

$$\text{By } 3.6,\ \det(EB) = -\det(B)$$

$$\det(E) = -\det(I_n)$$
$$\det(I_n) \equiv 1$$
$$\therefore \det(E) = -1$$
$$\Rightarrow \det(E)\det(B) = -\det(B) = \det(EB)$$

Let E be obtained by multiplying a row/column of I by a contant c:

$$\text{By } 3.6,\ \det EB = c\det B$$

$$\det E = c\det I = c$$
$$\therefore \det E\det B = c\det B = \det EB$$

Let E be olbtained by adding a multilple of a row/column of I to another row/column of I:

$$\text{By } 3.6,\ \det EB = \det B$$

$$\det E = \det I_n$$
$$\therefore \det E\det B = \det B = \det EB$$

Theorem 3.10 (Determinant of a Product). If A and B are square matrices of order n, then $\det(AB) = \det(A)\det(B)$.

Proof. *Proof of 3.10:*

$$\begin{gathered}
\text{Let } A, B \in \mathbb{R}^{n\times n} \text{ and } \det(A) \neq 0 \vee \det(B) \neq 0 \\
\implies \exists (S = \{E_1, E_2, \ldots, E_k\} \subsetneq \mathbb{R}^{n\times n})(E_1 E_2 \cdots E_k = A) \\
\implies \det(A)\det(B) = \det(E_1 \cdots E_k)\det(B) \\
= \det(E_1)\det(E_2 \cdots E_k)\det(B) \text{ 3.9} \\
= \det(E_1)\cdots\det(E_k)\det(B) \text{ 3.9} \\
= \det(E_1 \cdots E_k B) \text{ 3.9} \\
= \det(AB) \\
\text{Suppose } \det(A) = 0 \wedge \det(B) = 0 \\
\implies \text{yea we didnt rlly go over this bit but apparently} \\
\text{the product of 2 singular matrices will be singular idk}
\end{gathered}$$

■

Remark. Because $\det(AB) = \det(A)\det(B)$, we can actually define a homomorphism. The determinant does not preserve addition in the general case, so we cannot consider the vector space or ring of matrices. Rather, we can consider the multiplicative group of $n \times n$ matrices. However, groups necessitate the existance of inverses. Thus, we want to consider the general linear over the reals, that is $\mathrm{GL}(n, \mathbb{R})$. The determinant is not bijective, since it is not injective. For example, consider the following matrices:

$$I_n := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad E := \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

These are both in the general linear of order 4, and the concepts behind them can be generalized to order n. They both have a determinant of 1, but are different matrices. Thus, det is not injective. We can only define a homomorphism, not an isomorphism.

$$\det : \mathrm{GL}(n, \mathbb{R}) \to (\mathbb{R}, \cdot)$$

$$A \mapsto \det(A)$$

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

$$\therefore \det : \mathrm{GL}(n, \mathbb{R}) \simeq (\mathbb{R}, \cdot)$$

We have thus defined a homomorphism between the general linear and the multiplicative group of real numbers.

Theorem 3.11 (Determinant of a Scalar Multiple of a Matrix)**.** If A is a square matrix of order n, and $a; \in \mathbb{F}$, then $\det(\alpha A) = \alpha^n \det(A)$.

The proof of the above theorem follows trivially from the case of an elementary matrix given by a scalar multiple of a row in 3.6.

Theorem 3.12 (Determinant of a Nonsingular Matrix). If A is an invertible matrix, then $\det(A) \neq 0$.

Theorem 3.13 (Determinant of an Inverse Matrix).

$$\text{If } A \text{ is an } n \times n \text{ invertible matrix, then } \det\left(A^{-1}\right) = \frac{1}{\det(A)}$$

Proof. *Proof of 3.13:*

$$\begin{aligned} AA^{-1} &= I_n \\ \implies \det(A)\det\left(A^{-1}\right) &= 1 \\ \implies \det\left(A^{-1}\right) &= \frac{1}{\det(A)} \end{aligned}$$

■

Theorem 3.14 (Determinant of a Transpose). If A is a square matrix, then $\det(A) = \det\left(A^T\right)$.

The proof of the above theorem follows trivially from the fact that a cofactor expansion can be done columnwise and rowwise.

Theorem 3.15 (Conditions for Invertible Matrices). The following conditions are equivalent.

- A is invertible.
- $A\vec{x} = \vec{b}$ has exactly one solution for each $n \times 1$ column vector $\vec{b}$.
- $A\vec{x} = \vec{0}$ has only the trivial solution.
- A is row equivalent to I_n.
- A can be written as the product of elementary matrices.
- $\det(A) \neq 0$.

3.4 Applications of Determinants

Definition 3.16 (Adjoint of a Matrix)**.** The matrix of cofactors of some matrix A has the form

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}$$

The adjoint of this matrix is given as

$$\operatorname{adj}(A) = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}^T$$

Theorem 3.17 (Matrix Inverse)**.** If A is a square matrix of order n, then

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

Theorem 3.18 (Cramer's Rule)**.** If a system of n linear equations in n variables can be given as $A\vec{x} = \vec{b}$, where A is the coefficient matrix, then it has the solution

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad \cdots, \quad x_n = \frac{\det(A_n)}{\det(A)}$$

where A_i is the matrix A with the ith row replaced with $\vec{b}$.

Theorem 3.19 (Area of a Triangle in $\mathbb{R}^2$)**.** A triangle with vertices $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ has area

$$A = \left| \frac{1}{2} \det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} \right|$$

Theorem 3.20 (Collinear Points in $\mathbb{R}^2$)**.** Three points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ are collinear iff

$$\det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix} = 0$$

Theorem 3.21 (Equation of a Line). A line passing through points $(x_1, y_1), (x_2, y_2)$ is equivalent to

$$\det \begin{bmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{bmatrix} = 0$$

Theorem 3.22 (Volume of a Tetrahedron). A tetrahedron defined by points $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3), (x_4, y_4, z_4)$ has volume

$$V = \left| \frac{1}{6} \det \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{bmatrix} \right|$$

Theorem 3.23 (Coplanar Points in $\mathbb{R}^3$). Points $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3), (x_4, y_4, z_4)$ are coplanar iff

$$\det \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{bmatrix} = 0$$

Theorem 3.24 (Equation of a Plane). A plane through points $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)$ can be given as

$$\det \begin{bmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{bmatrix} = 0$$

Chapter 4

Vector Spaces

In general, this chapter assumes $\mathbb{F} = \mathbb{R}$, unless otherwise noted.

4.1 Vectors in $\mathbb{R}^n$

Definition 4.1 (Properties of Vector Addition and Scalar Multiplication in $\mathbb{R}^n$). Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ and $c, d \in \mathbb{F}$. Then

- $\vec{u} + \vec{v} \in \mathbb{R}^n$
- $\vec{u} + \vec{v} = \vec{v} + \vec{u}$
- $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$
- $\vec{u} + \vec{0} = \vec{u}, \vec{0} \in \mathbb{R}^n$
- $\vec{u} + (-\vec{u}) = \vec{0}, -\vec{u} \in \mathbb{R}^n$
- $c\vec{u} \in \mathbb{R}^n$
- $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$
- $(c + d)\vec{u} = c\vec{u} + d\vec{u}$
- $c(d\vec{u}) = (cd)\vec{u}$
- $1\vec{u} = \vec{u}, \vec{1} \in \mathbb{R}^n$

Theorem 4.2 (Properties of Additive Identity and Inverse). Let $\vec{v} \in \mathbb{R}^n$ and $c \in \mathbb{F}$.

1. The additive identity is unique.
2. There exists a unique additive inverse for every vector.
3. $0\vec{v} = \vec{0}$
4. $c\vec{0} - \vec{0}$
5. $c\vec{v} = \vec{0} \iff c = 0 \vee \vec{v} = \vec{0}$
6. $-(-\vec{v}) = \vec{v}$

Definition 4.3 (Linear Combination). A linear combination is a set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and scalars $c_1, \ldots, c_n$ such that a vector can be written as the sum of these vectors:

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{v}_i$$

Linear combinations can be treated as systems of equations.

Example. Write $\mathbf{x} = \langle 1, 7 \rangle$ as a linear combination of $\mathbf{v}_1 = \langle 2, 4 \rangle$ and $\mathbf{v}_2 = \langle 1, 3 \rangle$.

$$\begin{pmatrix} 1 \\ 7 \end{pmatrix} = c_1 \begin{pmatrix} 2 \\ 4 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$\begin{cases} 1 & = & 2c_1 & + & 1c_2 \\ 7 & = & 4c_1 & + & 3c_2 \end{cases}$$

Remark. An application of linear combinations is quantum computing. We often give

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$$

4.2 Vector Spaces

Definition 4.4 (Vector Space Axioms). A vector space is an abelian group defined over addition $(\mathbb{V}, +)$ with a corresponding scalar field $\mathbb{F}$ for which we define the operations of addition and scalar multiplication:

$$+ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$$

$$\cdot : \mathbb{F} \times \mathbb{V} \to \mathbb{V}$$

wherein the following axioms are satisfied:

- $c\mathbf{u} \in \mathbb{V}$ - closure under scalar multiplication
- $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ - scalar multiplication distributes over vector addition
- $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$ - scalar multiplication distributes over scalar addition
- $c(d\mathbf{u}) = (cd)\mathbf{u}$ - scalar multiplication is associative
- $1\mathbf{u} = \mathbf{u}$ - a scalar multiplicative identity exists

To show that a set is a vector space, it must be shown that the set is abelian under addition and that the rest of the axioms hold.

Remark (Important Vector Spaces).
1. The set of reals - $\mathbb{R}$
2. The set of real ordered pairs - $\mathbb{R}^2$
3. The set of real ordered triples - $\mathbb{R}^3$
4. The set of real ordered n-tuples - $\mathbb{R}^n$
5. The set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ - $C(-\infty, \infty)$
6. The set of all continuous functions $f : [a, b] \subset \mathbb{R} \to \mathbb{R}$ - $C[a, b]$
7. The set of all polynomials - P
8. The set of all polynomials of degree $\leq n$ - P_n
9. The set of all $m \times n$ matrices $M_{m\times n}$
10. The set of all $n \times n$ matrices $M_{n\times n}$

Theorem 4.5 (Properties of Scalar Multiplication). Let $\mathbf{v} \in V$ and $c \in \mathbb{F}$.

$$0\mathbf{v} = \mathbf{0}$$

$$c\mathbf{v} = 0 \iff c = 0 \lor \mathbf{v} = 0$$

$$c\mathbf{0} = \mathbf{0}$$

$$(-1)\mathbf{v} = -\mathbf{v}$$

Proof. Proof of statement 1:

$$\begin{aligned}
0\mathbf{v} &= (0+0)\mathbf{v} \\
0\mathbf{v} &= 0\mathbf{v} + 0\mathbf{v} \\
0\mathbf{v} - 0\mathbf{v} &= 0\mathbf{v} \\
\mathbf{0} &= 0\mathbf{v}
\end{aligned}$$

■

4.3 Subspaces of Vector Spaces

Definition 4.6 (Subspace). If V is a vector space, then a set $W \neq \varnothing$ with $W \subseteq V$ is a **subspace** of V when W is a vector space under the same definition of operations as V.

Theorem 4.7 (Test for a Subspace). Let V be a vector space, and let $W \subseteq V$. Then W is a subspace of V iff

- $W \neq \varnothing$
- If $\mathbf{u}, \mathbf{v} \in W$, then $\mathbf{u} + \mathbf{v} \in W$
- If $\mathbf{u} \in W$ and $c \in \mathbb{F}$, then $c\mathbf{u} \in W$

Theorem 4.8 (Intersections of Subspaces). If V and W are subspaces of U, then $V \cap W$ is a subspace of U.

4.4 Spanning Sets and Linear Combinations

Definition 4.9 (Linear Combination). A vector $\mathbf{x} \in V$ is a linear combination of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$ when

$$x = \sum_{i=1}^{k} c_i \mathbf{v}_i, c_i \in \mathbb{F}$$

To tell if some vector $\mathbf{u} \in V$ can be written as a linear combination of finite set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$, let $c_1, \ldots, c_n \in \mathbb{F}$ where

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

A system of equations can be constructed and solved by giving a matrix representation of the vector.

Remark. By a theorem that will be seen later on, for any vector space V with $\dim V = n$, we have

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

Since $\mathbb{R}^n$ has an equivalent representation as a column vector, all vector spaces of dimension n have the same representation.

Definition 4.10 (Span). Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset V$, where V is a vector space over $\mathbb{F}$. The **span** of S is the set of all linear combinations of S, denoted

$$\operatorname{span}(S) = \left\{ \sum_{i=1}^{n} c_i v_i \mid c_1 \ldots, c_n \in \mathbb{F} \right\}$$

Remark. When $\operatorname{span}(S) = V$, we say V is spanned by S, or the span of S is V.

Example. Let $\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$ be vectors in a vector space.

$$\operatorname{span}\left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\} = \mathbb{R}^2$$

This is just the result of rescaling and rotating the standard basis, so it's obviously going to span the space.

Theorem 4.11. If V is a vector space and $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subseteq V$, then $\operatorname{span}(S)$ is a subspace of V. $\operatorname{span}(S)$ is also the smallest subspace of V that contains S.

Proof. The span of a set $S \subseteq V$ is, by definition, every vector that can be obtained through vector addition with the elements of S, and scalar multiplication with the elements of $\mathbb{F}$. Thus, by definition, $\operatorname{span}(S)$ must be closed under scalar multiplication and vector addition. Furthermore,

$$\mathbf{v} \in S \implies 0\mathbf{v} = \mathbf{0} \in \operatorname{span}(S)$$

Also by definition, $\operatorname{span}(S) \subseteq V$. Therefore, $\operatorname{span}(S)$ is a subspace of V. From this it follows that any other subset of V that contains S must contain $\operatorname{span}(S)$ as well, otherwise it would not be closed. Thus, $\operatorname{span}(S)$ is the smallest subspace of V containing S. ■

Definition 4.12 (Linear Dependence and Independence). Let V be a vector space. A set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subseteq V$ is linearly independent when

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

has only the solution

$$c_1 = \cdots = c_k = 0$$

Any set that has nontrivial solutions is linearly dependent.

Remark. A linearly independent set is any set wherein any vector that is an element of the set cannot be written as a linear combination of the others. A linearly dependent set is a set where this is not true.

Let V be a vector space and S be set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subseteq V$. To determine whether S is linearly independent or dependent, use these steps:
Write $c_1\mathbf{v}_1 + \cdots c_k\mathbf{v}_k = \mathbf{0}$ as a system of linear equations.
Determine whether the system has a unique solution (take determinant of coefficient matrix).

If the set has a unique solution, it's linearly independent.

Example. Let $S = \{(1,3,1),(0,1,2),(1,0,-5)\} \subset \mathbb{R}^3$.

$$\det \begin{bmatrix} 1 & 2 & 1 \\ 3 & 1 & 0 \\ 1 & 0 & -5 \end{bmatrix} = 24 \neq 0$$

Therefore this set is linearly independent.

Theorem 4.13. A set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}, k \geq 2$ is linearly dependent iff at least one vector $\mathbf{v}_i$ can be written as a linear combination of the others.

Corollary 4.14. Two vectors $\mathbf{u}, \mathbf{v} \in V$ are linearly dependent iff there exists some $\alpha \in \mathbb{F}$ where $\mathbf{v} = \alpha\mathbf{u}$

4.5 Basis and Dimension

Definition 4.15 (Basis)**.** A set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k k\}$ in a vector space V is said to be a **basis** of V if

1. S spans V
2. S is linearly independent

Example. The standard basis of $\mathbb{R}^2$ is $\{(1,0),(0,1)\}$.

Example. Let $S = \{(1,2),(1,-1)\}$ be a set in the vector space $\mathbb{R}^2$. Let $(x,y) \in \mathbb{R}^2$ be an arbitrary vector.

$$(x,y) = a(1,2) + b(1,-1)$$
$$\implies (x,y) = (a+b, 2a-b)$$

This has a system representation with the folling matrix representation.

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\det \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} = 1 \neq 0$$

Therefore there exists a unique solution for every $(x,y) \in \mathbb{R}^2$, and thus this set forms a basis of $\mathbb{R}^2$.

Theorem 4.16 (Uniqueness of Basis Represenation)**.** If S is a basis of V, then every vector in V can be written in exactly one way as a linear combination of S.

Proof of this theorem follows from the facts that $\text{span}(S) = V$ and since the set is linearly independent, the coefficient matrix A constructed in the same manner as the previous example will be invertible.

Theorem 4.17 (Bases and Linear Dependence)**.** If S is a basis for V and has n vectors, then every set containing more than n vectors is linearly dependent.

Theorem 4.18 (Number of Vectors in a Basis)**.** If a vector space V has a basis with n vectors, then every basis of V has n vectors.

Definition 4.19 (Dimension of a Vector Space)**.** Let V be a vector space with a basis S with $|S| = n$. The number n is the **dimension** of V, denoted $\dim(V) = n$.

If a vector space V has a basis with a finite number of vectors, V is finite dimensional. Otherwise, V is infinite dimensional. $\{\mathbf{0}\}$ is said to be 0-dimensional.

Remark. The following are true.

- $\dim(\mathbb{R}^n) = n$
- $\dim(P_n) = n + 1$
- $\dim(M_{m,n}) = mn$

Theorem 4.20 (Basis Tests)**.** Let V be a vector space with $\dim(V) = n$.

- If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent, then it is a basis of V.
- If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans V, then S is a basis of V.

4.6 Rank of a Matrix and Systems of Linear Equations

Definition 4.21 (Row Space and Column Space)**.** Let $A \in M_{m,n}$ be an $m \times n$ matrix.

1. The **row space** of A is the subspace of $\mathbb{R}^n$ spanned by the row vectors of A.
2. The **column space** of A is the subspace of $\mathbb{R}^n$ spanned by the column vectors of A.

Example. Let

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

The row space of A is $\operatorname{span}\{(4,3,1),(-2,3,4)\}$ and the column space is $\operatorname{span}\left\{\begin{bmatrix} 4 \\ -2 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \end{bmatrix}\right\}$.

Theorem 4.22. Row equivalent matrices of equivalent dimension have equivalent row space. That is, for two row equivalent matrices $A, B \in M_{m,n}$, the span of their column vectors is equivalent.

Proof. Let $A, B \in M_{m,n}$ such that there exists a finite set of elementary matrices $\{E_1, \ldots, E_n\}$ where

$$E_n E_{n-1} \cdots E_2 E_1 A = B$$

There are three types of elementary matrices. An elementary matrix with a row swap performs a permutation on the row vectors of matrix A, preserving the vectors while modifying their order. A matrix that performs a scalar multiple on one row is equivalent to multiplying that vector by a scalar, which has an equivalent span since the scalar can be divided back out. Adding one row to another simply turns another vector into a linear combination of the original row vectors, which will have equivalent span as seen earlier. ■

Theorem 4.23 (Basis for Row Space)**.** If a matrix A is row equivalent to a matrix B in row-echelon form, then the nonzero row vectors of B form a basis for the row space of A.

Example.

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

Therefore, the row space of A has as a basis the set $\{(1, 2, -2, 0), (0, 0, 0, 1)\}$.

Theorem 4.24. The row and column spaces of any matrix A have equivalent dimension.

Definition 4.25 (Rank)**.** The dimension of the row or column space of some matrix A is the **rank** of A, denoted $\text{rank}(A)$.

Theorem 4.26 (Solutions of a Homogeneous System)**.** Let $A \in M_{m,n}$ be a matrix. The set of all solutions $\mathbf{x}$ to the equation $A\mathbf{x} = \mathbf{0}$ is called the **nullspace** of A, and is denoted

$$N(A) = \{\mathbf{x} \in \mathbb{R}^n | A\mathbf{x} = \mathbf{0}\}$$

and is a subspace of $\mathbb{R}^n$. The dimension of the nullspace of A is the **nullity** of A.

Remark. I think the nullspace is synonymous with $\ker(A)$ where A is a linear transformation.

Definition 4.27 (Dimension of the Nullspace)**.** If $A \in M_{m,n}$ is an $m \times n$ matrix with $\text{rank}(A) = r$, then $\dim(N(A)) = n - r$.

Theorem 4.28 (Solutions of a Nonhomogeneous Linear System)**.** Let $\mathbf{x}_p$ be a solution to $A\mathbf{x} = \mathbf{b}$ with $\mathbf{b} \neq \mathbf{0}$. Every solution of this system can then be written as the form

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$$

where $\mathbf{x}_h$ is a solution to $A\mathbf{x} = \mathbf{0}$.

Theorem 4.29 (Solutions of a System of Linear Equations). Let A be an $m \times n$ matrix. The system $A\mathbf{x} = \mathbf{b}$ is consistent iff $\mathbf{b}$ is in the column space of A.

Remark. Summary of Equivalent Conditions for Square Matrices:
If $A \in M_{n,n}$, then the following conditions are equivalent:

1. $\exists A^{-1} \in M_{n,n}$
2. $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$.
3. $A\mathbf{x} = \mathbf{0} \iff \mathbf{x} = \mathbf{0}$
4. A is row equivalent to I_n ($A \sim I_n$ because row equivalence is an equivalence relation).
5. $\det(A) \neq 0$
6. $\operatorname{rank}(A) = n$
7. The n rows of A are linearly independent.
8. The n columns of A are linearly independent.

4.7 Coordinates and Change of Basis

Definition 4.30 (Coordinate Representation Relative to a Basis). Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an ordered basis for a vector space V over $\mathbb{F}$ and let $\mathbf{x} \in V$ such that

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

The scalars $c_1, \ldots, c_n$ are the coordinates of $\mathbf{x}$ relative to the basis B, and the coordinate matrix of $\mathbf{x}$ relative to B is

$$[\mathbf{x}]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

Definition 4.31 (Transition Matrix). If B and B' are bases for a vector space $\mathbb{R}^n$, then a transition matrix P is a matrix such that

$$P[x]_B = [x]_{B'}$$

Lemma 4.32. Let $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $B' = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ be two bases for a vector space V. If

$$\begin{aligned}
\mathbf{v}_1 &= c_{11}\mathbf{u}_1 + c_{21}\mathbf{u}_2 + \cdots + c_{n1}\mathbf{u}_n \\
\mathbf{v}_2 &= c_{12}\mathbf{u}_1 + c_{22}\mathbf{u}_2 + \cdots + c_{n2}\mathbf{u}_n \\
&\vdots \\
\mathbf{v}_n &= c_{1n}\mathbf{u}_1 + c_{2n}\mathbf{u}_2 + \cdots + c_{nn}\mathbf{u}_n
\end{aligned}$$

Then the transition matrix from B to B' is

$$Q = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

Theorem 4.33 (Transition Matrices are Invertible)**.** If P is the transition matrix from a basis B' to a basis B for $\mathbb{R}^n$, then P is invertible and the transition matrix from B to B' is P^{-1}.

Proof. Let P be a transition matrix for $B' \to B$. It follows that

$$P[x]_{B'} = [x]_B$$

Let Q be a transition matrix for $B \to B'$. It follows that

$$[x]_{B'} = Q[x]_B$$

We now have

$$\begin{aligned}
[x]_B &= P\left(Q[x]_B\right) \\
\iff PQ &= I_n \\
[x]_{B'} &= Q\left(P[x]_B\right) \\
\iff QP &= I \\
\therefore QP = PQ = I &\iff P = Q^{-1}
\end{aligned}$$

The last step holds since inverses are unique. ■

Theorem 4.34 (Transition Matrices via Gaussian Elimination)**.** Let B and B' be bases for $\mathbb{R}^n$. The transition matrix P^{-1} from B to B' can be found with

$$\begin{bmatrix} B' & B \end{bmatrix} \xrightarrow{rref} \begin{bmatrix} I_n & P^{-1} \end{bmatrix}$$

Remark. Transition matrices define automorphisms on $\mathbb{R}^n$. Also all vector spaces of equivalent dimension are isomorphic, so the idea of a transition matrix extends beyond $\mathbb{R}^n$, but this is the space where it can be initially defined most easily, hence the use of $\mathbb{R}^n$ in our definitions.

Chapter 5

Inner Product Spaces

5.1 Length and Dot Product in $\mathbb{R}^n$

Definition 5.1 (Length of a Vector in $\mathbb{R}^n$). Let $\mathbf{v} \in \mathbb{R}^n$. The norm (length) of $\mathbf{v}$ is given as

$$\|\mathbf{v}\| = \sqrt{\sum_{i=1}^{n} v_i^2}$$

where v_i is the ith component of $\mathbf{v}$.

Theorem 5.2 (Length of Scalar Multiple). Let $\mathbf{v} \in \mathbb{R}^n$ and let $c \in \mathbb{F}$. Then $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$, where $|\cdot|$ is the absolute value operator.

Proof. Let $\mathbf{v} \in \mathbb{R}^n$. Thus, $\mathbf{v} = (v_1, v_2, v_3)$. Let $c \in \mathbb{F}$. It follows that

$$\begin{aligned}
\|c\mathbf{v}\| &= \sqrt{c^2v_1^2 + c^2v_2^2 + c^2v_3^2} \\
&= \sqrt{c^2\left(v_1^2 + v_2^2 + v_3^2\right)} \\
&= |c|\sqrt{v_1^2 + v_2^2 + v_3^2} \\
&= |c|\|\mathbf{v}\|
\end{aligned}$$

■

Theorem 5.3 (Unit Vector). Let $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. Then the vector

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

is the unit vector in the direction of $\mathbf{v}$.

Proof. Let $\mathbf{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. It follows immediately that

$$\frac{1}{\|\mathbf{v}\|} > 0$$

Let

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$$

It follows that

$$\begin{aligned}
\|\mathbf{u}\| &= \left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| \\
&= \left\| \frac{1}{\|\mathbf{v}\|}\mathbf{v} \right\| \\
&= \left| \frac{1}{\|\mathbf{v}\|} \right| \|\mathbf{v}\| \\
&= \frac{1}{\|\mathbf{v}\|}\|\mathbf{v}\| \qquad \because \frac{1}{\|\mathbf{v}\|} > 0 \\
&= \frac{\|\mathbf{v}\|}{\|\mathbf{v}\|} \\
&= 1
\end{aligned}$$

■

Remark. The process of dividing a vector by its magnitude is known as **normalizing the vector**.

Exercise. Normalize $\mathbf{v} = (-1, 3, 2)$.

Solution.

$$\|\mathbf{v}\| = \sqrt{(-1)^2 + 3^2 + 2^2} = \sqrt{1+9+4} = \sqrt{14}$$
$$\implies \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{(-1,3,2)}{\sqrt{14}} = \left(-\frac{1}{\sqrt{14}}, \frac{4}{\sqrt{14}}, \frac{2}{\sqrt{14}}\right) /$$

□

Definition 5.4 (Distance Between Vectors in $\mathbb{R}^n$)**.** Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. The **distance** between $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

Notice that the following rules apply.

1. $d(\mathbf{u}, \mathbf{v}) \geq 0$
2. $d(\mathbf{u}, \mathbf{v}) = 0 \iff \mathbf{u} = \mathbf{v}$
3. $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$

Exercise. Let $\mathbf{u} = (1, 2, 3)$ and $\mathbf{v} = (0, -1, 4)$. Find $d(u, v)$:

Solution.

$$\mathbf{u} - \mathbf{v} = (1 - 0, 2 + 1, 3 - 4) = (1, 3, -1)$$
$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{1^2 + 3^2 + (-1)^2} = \sqrt{1 + 9 + 1} = \sqrt{11}$$

□

Definition 5.5 (Dot Product). Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ with an angle θ between them. We have

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \cos\theta = \sum_{i=1}^{n} u_i v_i$$

where u_i, v_i are the ith entries of the vectors $\mathbf{u}$, $\mathbf{v}$, respectively.

Theorem 5.6 (Properties of the Dot Product). Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $c \in \mathbb{F}$. The following properties hold.

1. $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
2. $\mathbf{u} \cdot (\mathbf{v} + wbv) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
3. $c(\mathbf{u} \cdot \mathbf{v}) = (c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v})$
4. $\mathbf{v} \cdot \mathbf{v} = \|\mathbf{v}\|^2$
5. $\mathbf{v} \cdot \mathbf{v} \geq 0 \wedge \mathbf{v} \cdot \mathbf{v} = 0 \iff \mathbf{v} = \mathbf{0}$

Theorem 5.7 (Cauchy-Schwartz (in $\mathbb{R}^n$)). Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Then

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$$

Proof. Let $\mathbf{v}, \mathbf{u} \in \mathbb{R}^n$ with $\mathbf{u} \neq \mathbf{0}$. Let $t \in \mathbb{F}$.

$$(t\mathbf{u} + \mathbf{v}) \cdot (t\mathbf{u} + \mathbf{v}) \geq 0$$

$$\implies (\mathbf{u} \cdot \mathbf{u})t^2 + 2(\mathbf{u} \cdot \mathbf{v})t + (\mathbf{v} \cdot \mathbf{v}) \geq 0$$
$$\implies \|\mathbf{u}\|^2 t^2 + 2(\mathbf{u} + \mathbf{v})t + \|\mathbf{v}\|^2 \geq 0$$

This is a quadratic equation in the variable t. By the fundamental theorem of algebra, this equation has 2 solutions in $\mathbb{C}$. This equation will have either 2 roots in $\mathbb{R}$, with both roots being equal, or no roots in $\mathbb{R}$. All roots are of the form

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

where $a := \|\mathbf{u}\|^2$, $b := 2(\mathbf{u} \cdot \mathbf{v})$, and $c := \|\mathbf{v}\|^2$.

$$b^2 - 4ac \leq 0$$

$$\implies b^2 \leq 4ac$$
$$\implies (2(\mathbf{u} \cdot \mathbf{v}))^2 \leq 4\|\mathbf{u}\|^2 \|\mathbf{v}\|^2$$
$$\implies |\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

■

Corollary 5.8. From 5.7, we guarentee

$$0 \leq \frac{|\mathbf{u} \cdot \mathbf{v}|}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1$$

Definition 5.9 (Angle Between Two Vectors in $\mathbb{R}^n$)**.** Let θ be the angle between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}, \quad 0 \leq \theta \leq \pi$$

Exercise. Find the angle between vectors $\mathbf{u} = (-4, 0, 2)$ and $\mathbf{v} = (2, 0, 1)$ in $\mathbb{R}^3$.

Solution.

$$\|\mathbf{u}\| = \sqrt{20}$$
$$\|\mathbf{v}\| = \sqrt{5}$$
$$\mathbf{u} \cdot \mathbf{v} = -10$$
$$\implies \cos\theta = \frac{-10}{\sqrt{100}}$$
$$\implies \theta = \arccos(-1) = \pi$$

□

Definition 5.10 (Orthogonal Vectors). Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$.

Exercise. Find all vectors $\mathbf{v}$ orthogonal to $\mathbf{u} = (2, 1)$.

Solution. Let $\mathbf{v} = (v_1, v_2)$ with $\mathbf{v} \cdot \mathbf{u} = 0$. We have

$$2v_1 + v_2 = 0$$

$$\implies v_2 = -2v_1$$

Let $t \in \mathbb{R}$ be some parameter. We define

$$\mathbf{v} = (t, -2t)$$

□

Theorem 5.11 (The Triangle Inequality). Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Proof.

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\
&= \|\mathbf{u}\|^2 + 2(\mathbf{u} \cdot \mathbf{v}) + \|\mathbf{v}\|^2 \\
&\leq \|\mathbf{u}\|^2 + 2\left|\mathbf{u} \cdot \mathbf{v}\right| + \|\mathbf{v}\|^2 \\
&\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \\
\implies \|\mathbf{u} + \mathbf{v}\|^2 &\leq (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \\
\implies \|\mathbf{u} + \mathbf{v}\| &\leq \|\mathbf{u}\| + \|\mathbf{v}\| \because \|\cdot\| : \mathbb{R} \to \mathbb{R}_0^+
\end{aligned}$$

■

Theorem 5.12 (Pythagorean Theorem). Vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are orthogonal iff

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

Proof. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ with $\mathbf{u} \cdot \mathbf{v} = 0$.

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= (\mathbf{u}+\mathbf{v}) \cdot (\mathbf{u}+\mathbf{v}) \\ &= \|\mathbf{u}\|^2 + 2(\mathbf{u} \cdot \mathbf{v}) + \|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\end{aligned}$$

Now let $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ with $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. It follows from the above logic that $2(\mathbf{u} \cdot \mathbf{v}) \iff \mathbf{u} \cdot \mathbf{v} = 0$. ∎

One important note is that the dot product of two vectors

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

can be given as

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

which gives a 1×1 matrix $\begin{bmatrix} \mathbf{u} \cdot \mathbf{v} \end{bmatrix}$ which is isomorphic to the scalar it represents.

5.2 Inner Product Spaces

Definition 5.13 (Inner Product). Let V be a vector space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathbb{F}$. An inner product on V is a function

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$$

that satisfies the following axioms:

- $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
- $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
- $c\langle \mathbf{u}, \mathbf{v} \rangle = \langle c\mathbf{u}, \mathbf{v} \rangle$
- $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0 \iff \mathbf{v} = \mathbf{0}$

A vector space V with an inner product is an *inner product space.*

The dot product is an inner product on Euclidean space known as the Euclidean inner product.

Exercise. Let $\mathbf{v} = (v_1, v_2)$ and $\mathbf{u} = (u_1, u_2)$ be vectors in $\mathbb{R}^2$. Show

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + 9 u_2 v_2$$

is an inner product on $\mathbb{R}^2$.

Solution.

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v} \rangle &= u_1 v_1 + 9 u_2 v_2 \\ &= v_1 u_1 + 9 v_2 u_2 \\ &= \langle \mathbf{v}, \mathbf{u} \rangle\end{aligned}$$

Let $\mathbf{w} \in \mathbb{R}^2$ be defined as $\mathbf{w} = (w_1, w_2)$.

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle &= u_1(v_1 + w_1) + 9u_2\left(v_2 + w_2\right) \\ &= u_1 v_1 + u_1 w_1 + 9 u_2 v_2 + 9 u_2 w_2 \\ &= (u_1 v_1 + 9 u_2 v_2) + (u_1 w_1 + 9 u_2 w_2) \\ &= \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle\end{aligned}$$

Let $c \in \mathbb{R}$.

$$\begin{aligned}c\langle \mathbf{u}, \mathbf{v} \rangle &= c\left(u_1 v_1 + 9 u_2 v_2\right) \\ &= c u_1 v_1 + c 9 u_2 v_2 \\ &= \left(c u_1\right) v_1 + 9\left(c u_2\right) v_2 \\ &= \langle c\mathbf{u}, \mathbf{v} \rangle\end{aligned}$$

$$\begin{aligned}\langle \mathbf{u}, \mathbf{u} \rangle &= u_1^2 + 9 u_2^2 \\ u_1^2 &\geq 0 \\ u_2^2 &\geq 0 \\ u_1^2, u_2^2 = 0 &\iff u_1, u_2 = 0\end{aligned}$$

□

Exercise. Let $\mathbf{v} = (v_1, v_2)$ and $\mathbf{u} = (u_1, u_2)$ be vectors in $\mathbb{R}^2$. Show the function

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 - 7 u_2 v_2$$

does not define an inner product on $\mathbb{R}^2$.

Solution. Let $u_2 = \frac{u_1}{\sqrt{7}}$ with $u_1 \neq 0$. We have

$$\begin{aligned}\langle \mathbf{u}, \mathbf{u} \rangle &= u_1^2 - 7 u_2^2 \\ &= u_1^2 - 7\frac{u_1^2}{\sqrt{7}^2} \\ &= u_1^2 - \frac{7u_1^2}{7} \\ &= u_1^2 - u_1^2 \\ &= 0\end{aligned}$$

Therefore the function does not define an inner product on $\mathbb{R}^2$. □

Exercise. Let $f, , h \in C[a, b]$ be continuous functions defined on the interval $[a, b] \subset \mathbb{R}$ where $a \neq b$.

Show the function

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$$

defines an inner product on $C[a, b]$.

Solution. The functions are real-valued, so multiplication will commute. Thus,

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx = \int_a^b g(x)f(x)\,dx = \langle g, f\rangle$$

Integrals can be split at addition. Thus,

$$\langle f, g+h\rangle = \int_a^b f(x)(g(x)h(x))\,dx = \int_a^b f(x)g(x) + f(x)h(x)\,dx$$

$$= \int_a^b f(x)g(x)\,dx + \int_a^b f(x)h(x)\,dx = \langle f, g\rangle + \langle f, h\rangle$$

Let $c \in \mathbb{R}$. Constants can be factored in and out of an integral. Thus,

$$c\langle f, g\rangle = c\int_a^b \left(f(x)g(x)\right)\,dx = \int_a^b cf(x)g(x)\,dx = \int_a^b (cf(x))g(x)\,dx = \langle cf, g\rangle$$

The square of a real number will always be positive, and the integral of an integral that's always positive will be positive for $a > b$, which it is. Thus,

$$\langle f, f\rangle = \int_a^b f(x)f(x)\,dx \geq 0$$

The integral evaluates to 0 if and only if the integrand is 0, or if $f(a) = -f(b)$, but the square of the function is always $f^2 \geq 0$. Furthermore, the square of the function is 0 $\iff$ $f(x) = 0$. $\square$

Theorem 5.14 (Properties of Inner Products)**.** Let V be an inner product space over $\mathbb{F}$, and let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathbb{F}$. The following properties are satisfied.

- $\langle \mathbf{0}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{0}\rangle = 0$
- $\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$
- $\langle \mathbf{u}, c\mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$

Proof.

$$
\begin{aligned}
\langle \mathbf{0}, \mathbf{v} \rangle &= \langle \mathbf{v}, \mathbf{0} \rangle \\
\langle \mathbf{0}, \mathbf{v} \rangle &= \langle 0\mathbf{0}, \mathbf{v} \rangle \\
&= 0\langle \mathbf{0}, \mathbf{v} \rangle \\
&= 0 \,\forall \mathbf{v}
\end{aligned}
$$

$$\therefore \langle \mathbf{0}, \mathbf{v} \rangle \langle \mathbf{v}, \mathbf{0} \rangle = 0$$

$$
\begin{aligned}
\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= \langle \mathbf{w}, \mathbf{u} + \mathbf{v} \rangle \\
&= \langle \mathbf{w}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle
\end{aligned}
$$

$$\therefore \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$$

$$
\begin{aligned}
\langle \mathbf{u}, c\mathbf{v} \rangle &= \langle c\mathbf{v}, \mathbf{u} \rangle \\
&= c\langle \mathbf{v}, \mathbf{u} \rangle \\
&= c\langle \mathbf{u}, \mathbf{v} \rangle
\end{aligned}
$$

$$\therefore \langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$$

■

Rapid fire definition and theorem time!

Definition 5.15 (Norm). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u} \in V$. The norm of $\mathbf{u}$ is defined as

$$\|\mathbf{u}\| \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$$

Definition 5.16 (Distance). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v} \in V$. The distance between $\mathbf{u}$ and $\mathbf{v}$ is given as

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

Definition 5.17 (Angle). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v} \in V \setminus \{\mathbf{0}\}$. The angle θ between the vectors is

$$\theta = \arccos\left(\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}\right)$$

with $\theta \in [0, \pi]$.

Definition 5.18 (Orthogonality). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v} \in V$. These vectors are orthogonal if

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0$$

If $\|\mathbf{u}\| = 1$, then we say $\mathbf{u}$ is a unit vector. For some $\mathbf{v} \in V$, we say

$$\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

is the unit vector in the direction of $\mathbf{v}$.

Theorem 5.19 (Cauchy-Schwartz Inequality). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v} \in V$. We have

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$$

Theorem 5.20 (Triangle inequality). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v} \in V$. We have

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Theorem 5.21 (Pythagorean Theorem). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v} \in V$. $\mathbf{u}$ and $\mathbf{v}$ are orthogonal iff

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

Definition 5.22 (Orthogonal Projection). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v} \in V$ with $\mathbf{v} \neq \mathbf{0}$. The orthogonal projection of $\mathbf{u}$ onto $\mathbf{v}$ is defined as

$$\operatorname{proj}_{\mathbf{v}} \mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v}$$

Theorem 5.23 (Orthogonal Projection and Distance). Let V be an inner product space over $\mathbb{F}$ with $\mathbf{u}, \mathbf{v} \in V$ with $\mathbf{v} \neq \mathbf{0}$. Then

$$d(\mathbf{u}, \operatorname{proj}_{\mathbf{v}} \mathbf{u}) < d(\mathbf{u}, c\mathbf{v})$$

where

$$c \neq \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}$$

note - do 5.3 at home

5.3 Orthonormal Bases: Gram-Schmidt Process

Remark. When the inner product space being considered is $\mathbb{R}^n$ or some subspace of $\mathbb{R}^n$, assume the inner product being used is the Euclidean inner product.

We say that the basis $B = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is the standard basis for $\mathbb{R}^3$, since it's so convenient to use. Its convenience comes from the following properties:

1. All vectors in B are **mutually orthogonal**, meaning

$$(1, 0, 0) \cdot (0, 1, 0) = 0$$

$$(1,0,0)\cdot(0,0,1)=0$$
$$(0,1,0)\cdot(0,0,1)=0$$

2. All vectors in B are **unit vectors**.

Definition 5.24 (Orthogonal and Orthonormal Sets)**.** A set S of vectors in an inner product space V is **orthogonal** when every pair of vectors in S is orthogonal. If every vector is also a unit vector, then S is orthonormal.

If S is a basis for V, we say it is an **orthogonal basis** or **orthonormal basis**, respectively.

Remark. There are multiple orthonormal bases for many vector spaces. Consider the following basis for $\mathbb{R}^3$:

$$B=\{(\cos\theta,\sin\theta,0),(-\sin\theta,\cos\theta,0),(0,0,1)\}$$

Theorem 5.25 (Orthogonal Sets are Linearly Independent)**.** Let V be an inner product space with $S=\{\mathbf{v}_1,\dots,\mathbf{v}_n\}\subset V$, where $\mathbf{0}\notin S$. Then S is linearly independent.

Proof. Since any linearly dependent set has some solution $(c_1,\dots,c_n)\neq(0,\dots,0)$ for the equation

$$c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$$

we must show that this equation necessarily implies $c_1=\cdots=c_n=0$. Let $\mathbf{v}_i\in S$ be an arbitrary member of the set. Without loss of generality, we can give

$$\langle c_1\mathbf{v}_1+\cdots+c_i\mathbf{v}_i+\cdots+c_n\mathbf{v}_n\rangle=\langle\mathbf{0},\mathbf{v}_i\rangle$$

By 5.14, we have

$$\Longrightarrow c_1\langle\mathbf{v}_1,\mathbf{v}_i\rangle+\cdots+c_i\langle\mathbf{v}_i,\mathbf{v}_i\rangle+\cdots+c_n\langle\mathbf{v}_n,\mathbf{v}_i\rangle+\cdots=0$$

By 5.18, we have

$$c_i\langle\mathbf{v}_i,\mathbf{v}_i\rangle=0$$

By the definition of an inner product, we have

$$\langle\mathbf{v}_i,\mathbf{v}_i\rangle=0\iff\mathbf{v}_i=\mathbf{0}$$

However, by definition of S, we have $\mathbf{v}_i\neq 0$. Therefore,

$$c_i\langle\mathbf{v}_i,\mathbf{v}_i\rangle=0\Longrightarrow c_i=0$$

Thus, all $c_i=0$, and we have $(c_1,\dots,c_n)=(0,\dots,0)$. ■

Corollary 5.26. Let V be an inner product space with $\dim(V)=n$. Any orthogonal set of n nonzero vectors is a basis for V.

The above corrollary follows from the definition of a basis, and the proof is considered trivial.

Exercise. Using 5.26, show that S is a basis for $\mathbb{R}^4$.

$$S = \{(2,3,2,1), (1,0,0,1), (-1,0,2,1), (-1,2,-1,1)\}$$

Solution.

$$\begin{aligned}
(2,3,2,1)\cdot(1,0,0,1) &= 2+0+0-2 = 0\\
(2,3,2,1)\cdot(-1,0,2,1) &= -2+0+4-2 = 0\\
(2,3,2,1)\cdot(-1,2,-1,1) &= -2+6-2-2 = 0\\
(1,0,0,1)\cdot(-1,0,2,1) &= -1+0+0+1 = 0\\
(1,0,0,1)\cdot(-1,2,-1,1) &= -1+0+0+1 = 0\\
(-1,0,2,1)\cdot(-1,2,-1,1) &= 1+0-2+1 = 0
\end{aligned}$$

By 5.26, S is a basis for $\mathbb{R}^4$. □

Theorem 5.27 (Coordinates Relative to an Orthonormal Basis)**.** Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for an inner product space V. The coordinate representation of a vector $\mathbf{w}$ relative to B is given as

$$\mathbf{w} = \langle \mathbf{w}, \mathbf{v}_1\rangle \mathbf{v}_1 + \cdots + \langle \mathbf{w}, \mathbf{v}_n\rangle \mathbf{v}_n$$

Proof. Let B be an orthonormal basis for an inner product space V with $\mathbf{w} \in V$. It follows from the definition of a basis that there exists scalars $c_1, \ldots, c_n \in \mathbb{F}$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

It follows that

$$\begin{aligned}
\langle \mathbf{w}, \mathbf{v}_i\rangle &= \langle (c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n), \mathbf{v}_i\rangle\\
&= c_1\langle \mathbf{v}_1, \mathbf{v}_i\rangle + \cdots c_i\langle \mathbf{v}_i, \mathbf{v}_i\rangle + \cdots c_n\langle \mathbf{v}_n, \mathbf{v}_i\rangle\\
&= c_i\langle \mathbf{v}_i, \mathbf{v}_i\rangle
\end{aligned}$$

Since B is orthonormal with $\mathbf{v}_i \in B$, we have

$$\langle \mathbf{w} = \mathbf{v}_i\rangle = c_i\langle \mathbf{v}_i, \mathbf{v}_i\rangle = c_i\|\mathbf{v}_i\| = c_i$$

$$\therefore \langle \mathbf{w}, \mathbf{v}_i\rangle = c_i$$

■

The **Gram-Schmidt orthonormalization process** is a systematic method for constructing orthonormal bases from some arbitrary basis.

Theorem 5.28 (Gram-Schmidt Orthonormalization Process). Let V be an inner product space with a basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Define the set $B' = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$, where

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \mathbf{v}_2 - \tfrac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 \\
\mathbf{w}_3 &= \mathbf{v}_3 - \tfrac{\langle \mathbf{v}_3, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \tfrac{\langle \mathbf{v}_3, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 \\
&\vdots \\
\mathbf{w}_n &= \mathbf{v}_n - \tfrac{\langle \mathbf{v}_n, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \tfrac{\langle \mathbf{v}_n, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \cdots - \tfrac{\langle \mathbf{v}_n, \mathbf{w}_{n-1} \rangle}{\langle \mathbf{w}_{n-1}, \mathbf{w}_{n-1} \rangle}\mathbf{w}_{n-1}
\end{aligned}$$

Then B' is an orthogonal basis for V. To normalize B', let B'' be the set $B'' = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$, where

$$\mathbf{u}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$$

Then B'' is an orthonormal basis for V.

"Rather than give a proof of this theorem, it is more instructive to discuss a special case for which you can use a geometric model." -the textbook lmfao

Proof. I honestly don't know how to derive the formula, but we can use proof by induction to verify that B' is an orthogonal basis.
Let $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for an inner product space V. Let $\mathbf{w}_1 \coloneqq \mathbf{v}_1$. Let

$$\mathbf{w} \coloneqq \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1$$

It follows that

$$\begin{aligned}
\langle \mathbf{w}_1, \mathbf{w}_2 \rangle &= \left\langle \mathbf{w}_1, \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 \right\rangle \\
&= \langle \mathbf{w}_1, \mathbf{v}_2 \rangle - \left\langle \mathbf{w}_1, \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 \right\rangle \\
&= \langle \mathbf{w}_1, \mathbf{v}_2 \rangle - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle} \langle \mathbf{w}_1, \mathbf{w}_1 \rangle \\
&= \langle \mathbf{w}_1, \mathbf{v}_2 \rangle - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|} \|\mathbf{w}_1\| \\
&= \langle \mathbf{w}_1, \mathbf{v}_2 \rangle - \langle \mathbf{w}_1, \mathbf{v}_2 \rangle \\
&= 0
\end{aligned}$$

Therefore, $\mathbf{w}_1$ is orthogonal to $\mathbf{w}_2$. Now suppose we have defined $k < n$ orthogonal vectors. We must show that this implies the $k+1$th vector will be orthogonal to all other vectors. Using the above theorem, let

$$\mathbf{w}_{k+1} \coloneqq \mathbf{v}_{k+1} - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_k \rangle}{\langle \mathbf{w}_k, \mathbf{w}_k \rangle}\mathbf{w}_k$$

Now consider $\mathbf{w}_i$, where $1 \leq i \leq k$.

$$\begin{aligned}
\langle \mathbf{w}_i, \mathbf{w}_{k+1} \rangle &= \left\langle \mathbf{w}_i, \mathbf{v}_{k+1} - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_k \rangle}{\langle \mathbf{w}_k, \mathbf{w}_k \rangle}\mathbf{w}_k \right\rangle \\
&= \left\langle \mathbf{w}_i, \mathbf{v}_{k+1} - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_i \rangle}{\langle \mathbf{w}_i, \mathbf{w}_i \rangle}\mathbf{w}_i \right\rangle \because \mathbf{w}_i \text{ is orthogonal to all } \mathbf{w}_j, i \neq j \\
&= \langle \mathbf{w}_i, \mathbf{v}_{k+1} \rangle - \left\langle \mathbf{w}_i, \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_i \rangle}{\langle \mathbf{w}_i, \mathbf{w}_i \rangle}\mathbf{w}_i \right\rangle \\
&= \langle \mathbf{w}_i, \mathbf{v}_{k+1} \rangle - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_i \rangle}{\langle \mathbf{w}_i, \mathbf{w}_i \rangle} \langle \mathbf{w}_i, \mathbf{w}_i \rangle \\
&= \langle \mathbf{w}_i, \mathbf{v}_{k+1} \rangle - \frac{\langle \mathbf{v}_{k+1}, \mathbf{w}_i \rangle}{\|\mathbf{w}_i\|} \|\mathbf{w}_i\| \\
&= \langle \mathbf{w}_i, \mathbf{v}_{k+1} \rangle - \langle \mathbf{v}_{k+1}, \mathbf{w}_i \rangle \\
&= \langle \mathbf{w}_i, \mathbf{v}_{k+1} \rangle - \langle \mathbf{w}_i, \mathbf{v}_{k+1} \rangle \\
&= 0
\end{aligned}$$

$$\therefore \mathbf{w}_i \perp \mathbf{w}_{k+1} \, \forall 1 \leq i \leq k$$

When constructing a basis, we simply stop the process for $k = n$. Thus, by induction, this step of 5.28 produces an orthogonal basis. Simply normalizing each vector will preserve orthogonality, but will produce an orthonormal basis. Thus, the proof is complete. ■

Remark. An orthonormal set derived using 5.28 will be different depending on the ordering of the vectors in the basis.

Exercise. Use Gram-Schmidt orthonormalization on the basis $B = \{1, x, x^2\}$ on the vector space P_2, with the inner product

$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

Solution. Let $B' = \langle \mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3 \rangle$ be the set obtained after applying the first step of Gram-Schmidt. Let

$$\mathbf{v}_1 = 1 \quad \mathbf{v}_2 = x \quad \mathbf{v}_3 = x^2$$

We have

$$\mathbf{w}_1 = \mathbf{v}_1 = 1$$

Thus, we also have

$$\begin{aligned}
\mathbf{w}_2 &= \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 \\
&= x - \frac{\int_{-1}^{1} x\,dx}{\int_{-1}^{1}\,dx} 1 \\
&= x
\end{aligned}$$

Finally, we have

$$\begin{aligned}
\mathbf{w}_3 &= \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_3, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 \\
&= x^2 - \frac{\int_{-1}^{1} x^2 \, dx}{\int_{-1}^{1} \, dx} 1 - \frac{\int_{-1}^{1} x^3}{\int_{-1}^{1} x^2} \, dx \\
&= x^2 - \frac{1}{3}
\end{aligned}$$

Let $B'' = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, where

$$\mathbf{u}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} \quad \mathbf{u}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} \quad \mathbf{u}_3 = \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|}$$

We have

$$\begin{aligned}
\mathbf{u}_1 &= \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = \frac{1}{\sqrt{2}} \\
\mathbf{u}_2 &= \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \frac{\sqrt{3}}{\sqrt{2}} x \\
\mathbf{u}_3 &= \frac{\mathbf{w}_3}{\|\mathbf{w}_3\|} = \frac{\sqrt{5}}{2\sqrt{2}} \left(3x^2 - 1\right)
\end{aligned}$$

□

Remark. The polynomials $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ in the previous example are the first three **normalized Legendre polynomials**.