MATH 2210Q – Linear Algebra Lecture Notes

From *Linear Algebra and its Applications 5e*
David C. Lay, Steven R. Lay, Judy J. McDonald

Notes compiled by Bobby McDonald
University of Connecticut, Spring 2019

Contents

1 Linear Equations in Linear Algebra **5**

1.1 Systems of Linear Equations 5
1.2 Row Reduction and Echelon Forms 11
1.3 Vector Equations 17
1.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$ 23
1.5 Solutions Sets of Linear Systems 27
1.6 blah 30
1.7 Linear Independence 31
1.8 Intro to Linear Transformations 35
1.9 Matrix Transformations 40

2 Matrix Algebra **47**

2.1 Matrix Operations 47
2.2 The Inverse of a Matrix 53
2.3 Characterizations of Invertible Matrices 58
2.4 Matrix Factorizations 63

3 Determinants **69**

3.1 Introduction to Determinants 69
3.2 Properties of Determinants 72
3.3 Cramer's Rule 77
3.3.1 Linear Transformations 81

4 Vector Spaces **83**

4.1 Vector Spaces and Subspaces 83
4.2 Null Spaces, Column Spaces, Linear Transformations 89
4.3 Linearly Independent Sets; Bases 95
4.3.1 The spanning set theorem 97
4.3.2 Bases for $\operatorname{Col} A$ and $\operatorname{Nul} A$ 98
4.3.3 Two views of a basis 99
4.4 Coordinate Systems 100

5 Eigenvectors and Eigenvalues **105**

5.1 Eigenvectors and Eigenvalues 105
5.2 The Characteristic Equation 109
5.3 Diagonalization 113
5.4 Eigenvectors and Linear Transformations 118

5.4.1 Linear transformations from V into V 120
5.4.2 Linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$ 121

6 Orthogonality and Least Squares 123

6.1 Inner Product, Length, and Orthogonality 123
6.2 Orthogonal Sets 128
6.2.1 Orthogonal Projection 130
6.2.2 Orthonormal Sets 131
6.3 Orthogonal Projections 133
6.4 The Gram-Schmidt Process 136

Chapter 1

Linear Equations in Linear Algebra

1.1 Systems of Linear Equations

Definition 1.1.1. A **linear equation** in the variables $x_1, \ldots, x_n$ 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** $a_1, \ldots, a_n$ are real or complex numbers.

Example 1.1.2. 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 1.1.3. A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same variables, say $x_1, \ldots, x_n$. A **solution** of the system is a list $(s_1, s_2, \ldots, s_n)$ that give a true statement when the s_i are substituted in for $x_1, \ldots, x_n$ respectively.

Example 1.1.4. Is $(5, 6.5, 3)$ in the solution set of the system

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

Definition 1.1.5. 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.

Proposition 1.1.6. *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.

Remark 1.1.7. 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 1.1.8. What are the solution sets of the following systems? Are they 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 1.1.9. The basic strategy for solving a system is to replace it with an equivalent system that's easier to solve. We can do this by replacing one equation by the sum of itself with the multiple of another equation, interchanging equations, or multiplying an equation by a nonzero constant.

Definition 1.1.10. The essential information in a linear system can be recorded into a rectangular array called a **matrix**. For example, given the system

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

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

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

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

Example 1.1.12. Solve the system in the definition.

Definition 1.1.13 (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 1.1.14. If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

Remark 1.1.15. 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 1.1.16. Determine if the following system is consistent:

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

Example 1.1.17. Determine if the following system is consistent:

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

1.2 Row Reduction and Echelon Forms

Definition 1.2.1. 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 nonzero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row 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 1.2.2. 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 1.2.3. Any nonzero matrix can be **row reduced** into infinitely many matrices in echelon form. However, *reduced* echelon form for a matrix is unique.

Theorem 1.2.4. *Each matrix is row equivalent to one and only one reduced echelon matrix.*

Definition 1.2.5. 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 1.2.6. Label the pivot positions and pivot columns of the matrices above.

Example 1.2.7. 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 1.2.8 (Row Reduction Algorithm)**.** To transform a matrix into echelon form:

1. Begin with the leftmost nonzero column. This is a pivot column. The pivot position is at the top.
2. Select a non-zero entry in the pivot column as a pivot, and interchange rows if necessary to move this entry into the pivot position.
3. Use row replacement to create zeros in all positions below the pivot.
4. Cover (ignore) the row containing the pivot positions and all rows, if any, above it. Apply steps 1-3 to the submatrix that remains. Repeat until there are no more nonzero rows to modify.

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 1.2.9. 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 1.2.10. Steps 1-4 above are called the **forward phase** of the row reduction algorithm. Step 5 is called the **backward phase.**

Example 1.2.11. 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 1.2.12. The variables corresponding to pivot columns of a matrix are called **basic variables**, the other variables are called **free variables**.

Remark 1.2.13. 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 1.2.14. Find the general solution of a system whose augmented matrix is reduced to

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

Remark 1.2.15. The solutions in these examples are called **parametric descriptions**. Finding a general solution means finding a parametric description of the solution set, or determining that it is empty. If a system is inconsistent, then the solution set has no parametric representation.

Example 1.2.16. Solve the following system:

$$\begin{aligned} x_1 - 7x_2 + 2x_3 - 5x_4 + 8x_5 &= 10 \\ x_2 - 3x_3 + 3x_4 + x_5 &= -5 \\ x_4 - x_5 &= 4 \end{aligned}$$

Remark 1.2.17. The method we used above is called **back substitution**. When we reduce to reduced echelon form, this is handled by the backward phase and reduces the likelihood of errors.

Example 1.2.18. 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 1.2.19. *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 1.2.20 (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.

1.3 Vector Equations

Definition 1.3.1 (Vectors in $\mathbb{R}^2$)**.** A matrix 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 1.3.2. 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 1.3.3. 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 1.3.4 (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.

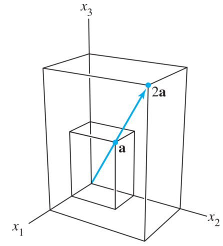
Example 1.3.5. Let $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$. Graph $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u} + \mathbf{v}$ on the plane.

Proposition 1.3.6 (Parallelogram Rule)**.** *If* $\mathbf{u}$ *and* $\mathbf{v}$ *in* $\mathbb{R}^2$ *are represented in the plain, then* $\mathbf{u} + \mathbf{v}$ *corresponds to the last vertex of the parallelogram with vertices are* $\mathbf{u}$*,* $\mathbf{v}$ *and* $\mathbf{0}$.

Example 1.3.7. 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?

Observation 1.3.8 (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,

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



Definition 1.3.9 (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 again, sometimes denote $(a_1, a_2, \ldots, a_n)$. The **zero vector**, denoted $\mathbf{0}$ is the vector whose entries are all zero. We also denote $(-1)\mathbf{u} = -\mathbf{u}$.

Proposition 1.3.10 (Algebraic Properties of $\mathbb{R}^n$)**.** *For* $\mathbf{u}$*,* $\mathbf{v}$*,* $\mathbf{w}$ *in* $\mathbb{R}^n$*, and scalars* c*,* d*:*

(i)	$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$	(v)	$c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$
(ii)	$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$	(vi)	$(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$
(iii)	$\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$	(vii)	$c(d\mathbf{u}) = (cd)\mathbf{u}$
(iv)	$\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$	(viii)	$1\mathbf{u} = \mathbf{u}$

Remark 1.3.11. Sometimes, for ease of notation, we denote $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ as $(a_1, a_2, \ldots, a_n)$.

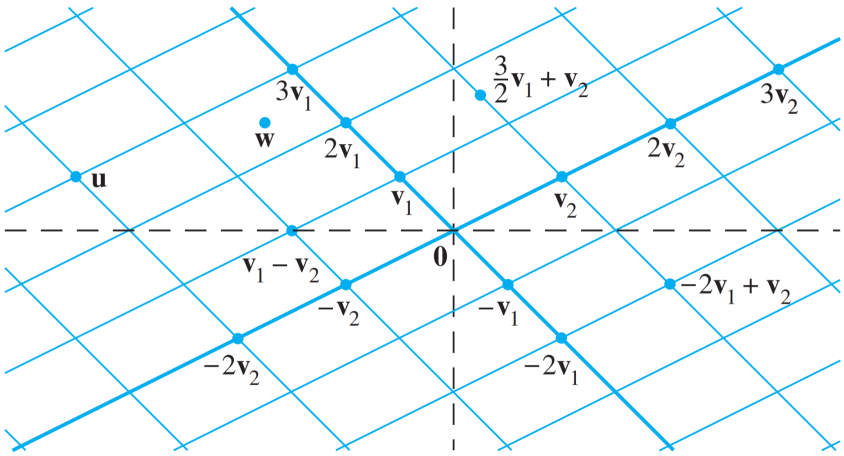
Example 1.3.12. Prove properties (i) and (v) of the Algebraic Properties above.

Definition 1.3.13 (Linear Combinations)**.** Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ in $\mathbb{R}^n$, and scalars $c_1, c_2, \ldots, c_m$. The vector

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

is called a **linear combination** of the $\mathbf{v}_1, \ldots \mathbf{v}_m$ with **weights** $c_1, \ldots, c_m$.

Example 1.3.14. 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 $\mathbf{w}$.



Example 1.3.15. Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$. Is $\mathbf{b}$ a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$?

Remark 1.3.16. In the previous example, the vectors $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{b}$ became the columns of the augmented matrix that we reduced:

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

For brevity, we will write this matrix, using vectors, as $\left[\begin{array}{cc|c} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{b} \end{array}\right]$. This suggests the following.

Procedure 1.3.17. A vector equation $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$, has the same solution set as the linear system whose augmented matrix is

$$\left[\begin{array}{ccc|c} \mathbf{a}_1 & \cdots & \mathbf{a}_n & \mathbf{b} \end{array}\right]$$

In particular, $\mathbf{b}$ can be represented as a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ if and only if there is a solution to the linear system corresponding to this matrix.

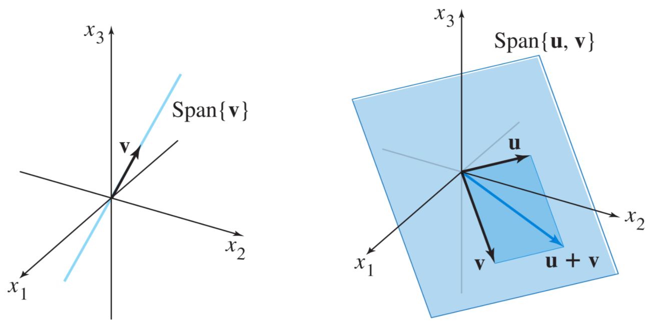
Definition 1.3.18. If $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are in $\mathbb{R}^n$, then the set of all linear combinations of is denoted by $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ and is called the **subset of $\mathbb{R}^n$ spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_m$**. In other words, $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is the collection of all vectors of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m, \text{ with } c_1, \ldots, c_m \text{ scalars.}$$

Example 1.3.19. Let $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Prove that $\mathbf{v}_1$ and $\mathbf{v}_2$ span all of $\mathbb{R}^2$.

Remark 1.3.20. Actually, for *any* $\mathbf{u}$ and $\mathbf{v}$ (which are not multiples) in $\mathbb{R}^3$, Span$\{\mathbf{u}, \mathbf{v}\}$ is a plane!

Observation 1.3.21 (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 $\mathbf{v}$. Then Span$\{\mathbf{v}\}$ is the set of points on the line in $\mathbb{R}^3$ through $\mathbf{0}$ and $\mathbf{v}$, and Span$\{\mathbf{u}, \mathbf{v}\}$ is the plane in $\mathbb{R}^3$ containing $\mathbf{0}$, $\mathbf{u}$ and $\mathbf{v}$, that is, it contains the line in $\mathbb{R}^3$ through $\mathbf{u}$ and the line through $\mathbf{0}$ and $\mathbf{v}$ and $\mathbf{0}$.



Example 1.3.22. If $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 5 \\ -13 \\ -3 \end{bmatrix}$. Is $(-3, 8, 1)$ in the plane spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$?

1.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$

The definition below lets us rephrase some of the concepts from Section 1.3 by viewing linear combinations of vectors as the product of a matrix and a vector

Definition 1.4.1. If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and $\mathbf{x}$ is in $\mathbb{R}^n$, then the **product of** A **and** $\mathbf{x}$**,** denoted by $A\mathbf{x}$, is the linear combination of the columns of A using the corresponding entries in $\mathbf{x}$ as weights:

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

Remark 1.4.2. $A\mathbf{x}$ is only defined the number of columns of A equals the number of entries in $\mathbf{x}$.

Example 1.4.3. Find the following products:

(a) $\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & 1 \\ 3 & 9 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Example 1.4.4. Compute $A\mathbf{x}$, where $A = \begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

Procedure 1.4.5 (Row Vector Rule)**.** If $A\mathbf{x}$ is defined, then the ith entry in $A\mathbf{x}$ is the sum of the products of corresponding entries from row i of A and from the vector $\mathbf{x}$.

Example 1.4.6. Compute

(a) $\begin{bmatrix} 2 & -3 \\ 8 & 0 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

Example 1.4.7. Write the system below as $A\mathbf{x} = \mathbf{b}$ for some A and $\mathbf{b}$.

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

Definition 1.4.8. The equation $A\mathbf{x} = \mathbf{b}$ is called a **matrix equation.**

Theorem 1.4.9. *If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and $\mathbf{b}$ is in $\mathbb{R}^m$, then*

$$A\mathbf{x} = \mathbf{b} \ (\textit{with } \mathbf{x} \textit{ in } \mathbb{R}^n)$$

has the same solutions as the vector equation

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

which has the same solutions as the system of linear equations with augmented matrix

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

Corollary 1.4.10. *The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.*

Example 1.4.11. Let $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ is $A\mathbf{x} = \mathbf{b}$ consistent for all b_1, b_2, b_3?

Theorem 1.4.12. *Let A be an $m \times n$ matrix. Then the following statements are either all true, or all false.*

(a) *For each* $\mathbf{b}$ *in* $\mathbb{R}^m$, *the equation* $A\mathbf{x} = \mathbf{b}$ *has a solution.*

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

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

(d) A *has a pivot position in every row.*

Example 1.4.13. If $A = \begin{bmatrix} 2 & 0 & -2 \\ 2 & 3 & 4 \\ 0 & 1 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 11 \\ 3 \end{bmatrix}$, for what $\mathbf{x}$ is $A\mathbf{x} = \mathbf{b}$ consistent?

We end this section with some important properties of $A\mathbf{x}$, which we will use throughout the course.

Theorem 1.4.14. *If A is an $m \times n$ matrix,* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in* $\mathbb{R}^n$, *and c is a scalar:*

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

(b) $A(c\mathbf{u}) = c(A\mathbf{u})$.

1.5 Solutions Sets of Linear Systems

In this section, we will use vector notation to give explicit and geometric descriptions of solution sets of linear systems. We begin by defining a special type of system.

Definition 1.5.1. 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 1.5.2. 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 1.5.3. 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 1.5.4. *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 1.5.5. Describe all solutions to the homogeneous system

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

Definition 1.5.6. The answers in 1.5.3 and 1.5.5 are **parametric vector equations**. Sometimes, to emphasize that the parameters vary over all real numbers, we write

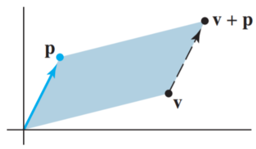
$$\mathbf{x} = s\mathbf{u} + t\mathbf{v} \text{ for } s, t \in \mathbb{R}.$$

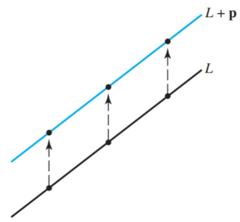
In both examples, we say that the solution is in **parametric vector form.**

Example 1.5.7. Describe all solutions of

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

Definition 1.5.8. 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 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.

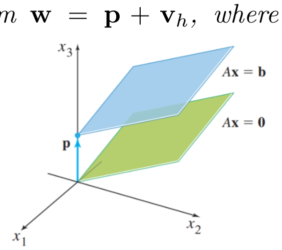




For $t \in \mathbb{R}$, we call $\mathbf{p}+t\mathbf{v}$ **the equation of the line parallel to v through p.**

Example 1.5.9. Use this observation to describe the relationships between the solutions to $A\mathbf{x}=\mathbf{0}$ and $A\mathbf{x}=\mathbf{b}$ using the A and $\mathbf{b}$ from Examples 1.5.3 and 1.5.7.

Theorem 1.5.10. *Suppose the equation* $A\mathbf{x}=\mathbf{b}$ *is consistent for some given* $\mathbf{b}$*, and let* $\mathbf{p}$ *be a solution. Then the solution set of* $A\mathbf{x}=\mathbf{b}$ *is the set of all vecotrs of the form* $\mathbf{w}=\mathbf{p}+\mathbf{v}_h$*, where* $\mathbf{v}_h$ *is any solution of the homogeneous equation* $A\mathbf{x}=\mathbf{0}$*.*



Procedure 1.5.11. 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 1.5.12. Describe and compare the solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$ if

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

1.7 Linear Independence

Definition 1.7.1. 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 1.7.2. 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 1.7.3. 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.

Remark 1.7.4. If $A = \begin{bmatrix} \mathbf{v}_m & \cdots & \mathbf{v}_m \end{bmatrix}$, then the homogeneous equation $A\mathbf{x} = \mathbf{0}$ can be written

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_m = \mathbf{0}.$$

Thus, linear independence is the same as having no non-trivial solutions to this matrix equation.

Definition 1.7.5. The columns of a matrix A are linearly independent if and only if $A\mathbf{x} = \mathbf{0}$ has no non-trivial solutions.

Example 1.7.6. Determine if the columns of $A = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 2 & -1 \\ 5 & 8 & 0 \end{bmatrix}$ are linearly independent.

Example 1.7.7. 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 1.7.8 (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.*

Theorem 1.7.9 (Characterization of Linearly Dependent Sets)**.** *An indexed set* $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ *of two or more vectors is linearly dependent if and only if at least one of the vectors is a linear combination of the others. In fact, if* S *is linearly dependent and* $\mathbf{v}_1 \neq \mathbf{0}$*, then some* $\mathbf{v}_j$ $(j > 1)$ *is a linear combination of the preceding vectors,* $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$*.*

Example 1.7.10. If $\mathbf{u}$ and $\mathbf{v}$ are linearly independent non-zero vectors in $\mathbb{R}^3$. Geometrically describe Span$\{\mathbf{u}, \mathbf{v}\}$. Prove $\mathbf{w}$ is in Span$\{\mathbf{u}, \mathbf{v}\}$ if and only if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a linearly dependent set.

Theorem 1.7.11 (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* $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ *in* $\mathbb{R}^n$ *is linearly dependent if* $p > n$*.*

Theorem 1.7.12. *If a set* S *in* $\mathbb{R}^n$ *contains the zero vector, then* S *is linearly dependent.*

Example 1.7.13. 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) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$

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

(d) $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix}$

1.8 Intro to Linear Transformations

Example 1.8.1. If $A = \begin{bmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 1 \\ 4 \\ -1 \\ 3 \end{bmatrix}$. Find $A\mathbf{0}$, $A\mathbf{u}$, and $A\mathbf{v}$.

We can think of A as *acting* on $\mathbf{0}$, $\mathbf{u}$, and $\mathbf{v}$ like a *function* from one set of vectors to another.

Definition 1.8.2. A **transformation** (also called a **function** or **mapping**) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbb{R}^n$ one (and only one) vector $T(\mathbf{x})$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the **domain** of T and $\mathbb{R}^m$ is called the **codomain** of T, denoted

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

For $\mathbf{x}$ in $\mathbb{R}^n$, the vector $T(\mathbf{x})$ in $\mathbb{R}^m$ is called the **image** of $\mathbf{x}$. The subset of $\mathbb{R}^m$ consisting of all possible images $T(\mathbf{x})$ is called the **range**.

Remark 1.8.3. In this section, we will focus on mappings associated to *matrix multiplication*. For simplicity, we sometimes denote this *matrix transformation* by $\mathbf{x} \mapsto A\mathbf{x}$.

Example 1.8.4. Let $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$,
and define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$.

(a) Write $T(\mathbf{x})$ as a vector.

(b) Find $T(\mathbf{u})$, the image of $\mathbf{u}$ under the transformation T.

(c) Find an $\mathbf{x}$ in $\mathbb{R}^2$ such that $T(\mathbf{x}) = \mathbf{b}$. Is there more than one?

(d) Determine if $\mathbf{c}$ is in the range of T.

Example 1.8.5. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and describe $\mathbf{x} \mapsto A\mathbf{x}$.

Remark 1.8.6. The map in Example 1.8.5 is called a **projection map.**

Example 1.8.7. Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, and $T(\mathbf{x}) = A\mathbf{x}$. Find the images of $\mathbf{u} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ under T, and use this to describe T geometrically.

Remark 1.8.8. The map in Example 1.8.7 is called a **shear transformation**.

Definition 1.8.9. A transformation T is called **linear** if

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

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and $\mathbf{u}$ in the domain of T.

Remark 1.8.10. The properties of $A\mathbf{x}$ Section 1.4 show that when T is a *matrix transformation*, T is a linear transformation. Not all transformations are linear, however.

Proposition 1.8.11. *If T is a linear transformation, then*

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

and

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

for all vectors $\mathbf{u}, \mathbf{v}$ in the domain of T, and all scalars c, d.

Remark 1.8.12. If T satisfies the second property above, then it is as linear transformation. Repeated application of this property gives

$$T(c_1\mathbf{u}_1 + \cdots + c_n\mathbf{v}_n) = c_1T(\mathbf{u}_1) + \cdots + c_nT(\mathbf{v}_n)$$

In physics and engineering, this is called a *superposition principle*.

Example 1.8.13. Given a scalar $r \geq 0$, define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = r\mathbf{x}$. Prove T is a linear transformation, and describe it geometrically.

Remark 1.8.14. In Example 1.8.12, the linear transformation T is called a **contraction** when $0 < r < 1$ and a **dilation** when $r > 1$.

Example 1.8.15. Let $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, and $T(\mathbf{x}) = A\mathbf{x}$. Find the images of $\mathbf{u} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, and $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ under T, and use this to describe T geometrically.

1.9 Matrix Transformations

Whenever a function is decribed geometrically or in words, we usually want to find a formula. In linear algebra, the same will be true for linear transformations. It turns out that *every* linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is actually a matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$.

Example 1.9.1. Suppose that T is a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^3$ such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \qquad\qquad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 7 \\ -8 \\ 6 \end{bmatrix}$$

Find a formula for the image of an arbitrary $\mathbf{x}$ in $\mathbb{R}^2$, and a matrix, A, such that $T(\mathbf{x}) = A\mathbf{x}$.

Definition 1.9.2. The **identity matrix**, I_n, is the $n \times n$ matrix with ones on the diagonal $[\diagdown]$, and zeros everywhere else. For example

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad\qquad I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Remark 1.9.3. The key to finding the matrix for a linear transformation is to see what it does I_n.

Theorem 1.9.4. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that*

$$T(\mathbf{x}) = A\mathbf{x} \textit{ for all } \mathbf{x} \textit{ in } \mathbb{R}^n$$

In fact, A is the $m \times n$ matrix whose jth column is the vector $T(\mathbf{e}_j)$, where $\mathbf{e}_j$ is the jth column of the indentity matrix in $\mathbb{R}^n$:

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

Definition 1.9.5. The matrix A in Theorem 1.9.4 is called the **standard matrix** for T.

Example 1.9.6. If $r \geq 0$, find the standard matrix for the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $\mathbf{x} \mapsto r\mathbf{x}$.

Example 1.9.7. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation that rotates each point counter clockwise about the origin through an angle α. Find the standard matrix for T.

The following definitions should sound familiar.

Definition 1.9.8. A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **onto** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at least one* $\mathbf{x}$ in $\mathbb{R}^n$. T is said to be **one-to-one** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at most one* $\mathbf{x}$ in $\mathbb{R}^n$.

Remark 1.9.9. T being *onto* is an *existence* question: for every $\mathbf{b}$ in $\mathbb{R}^m$, does an $\mathbf{x}$ exist such that $T(\mathbf{x}) = \mathbf{b}$? T being *one-to-one* is a *uniqueness* question: for every $\mathbf{b}$ in $\mathbb{R}^m$, if there is a solution to $T(\mathbf{x}) = \mathbf{b}$, is it unique?

Example 1.9.10. Let T be the transformation whose standard matrix is

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

Is T one-to-one? Is T onto?

Theorem 1.9.11. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if the equation $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.*

Theorem 1.9.12. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A. Then:*
(a) T is one-to-one if and only if the columns of A are linearly independent;
(b) T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$.

Example 1.9.13. Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be the transformation that brings $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ to $\begin{bmatrix} 2x_1 + 4x_4 \\ x_1 + x_2 + 3x_4 \\ -2x_1 + x_3 - 4x_4 \end{bmatrix}$.
Find a standard matrix for T and determine if T is one-to-one. Is T onto?

Example 1.9.14. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the transformation that brings $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ to $\begin{bmatrix} x_1 - x_2 \\ -2x_1 + x_2 \\ x_1 \end{bmatrix}$.
Find a standard matrix for T and determine if T is one-to-one. Is T onto?

Example 1.9.15. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. If T is onto, what can you say about m and n? If T is one-to-one, what can you say about m and n?

Example 1.9.16. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that first reflects points through the horizontal x_1-axis, and then reflects them through the line $x_2 = x_1$. Find the standard matrix of T.

Remark 1.9.17. The following tables, taken from Lay's Linear Algebra book, illustrate common geometric linear transformations of the plane. Each shows the transformation of the unit square.

TABLE 1 Reflections

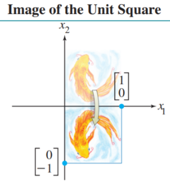
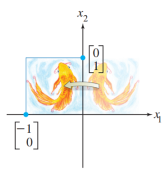
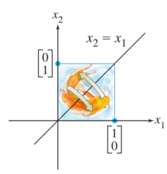
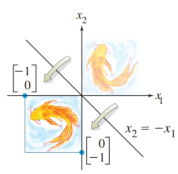
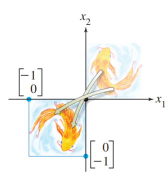
Transformation	Image of the Unit Square	Standard Matrix
Reflection through the x_1-axis		$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Reflection through the x_2-axis		$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
Reflection through the line $x_2 = x_1$		$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
Reflection through the line $x_2 = -x_1$		$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$
Reflection through the origin		$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$

TABLE 2 Contractions and Expansions

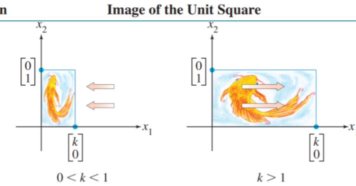
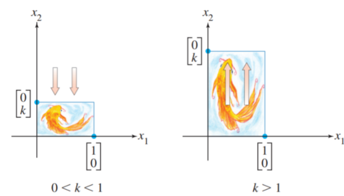
Transformation	Image of the Unit Square	Standard Matrix
Horizontal contraction and expansion		$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$
Vertical contraction and expansion		$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$

TABLE 3 Shears

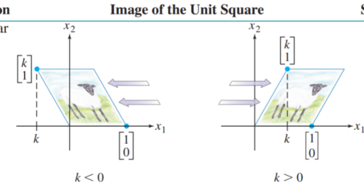
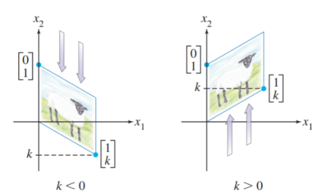
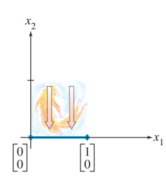
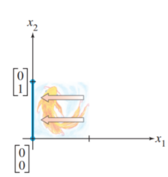
Transformation	Image of the Unit Square	Standard Matrix
Horizontal shear		$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$
Vertical shear		$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

TABLE 4 Projections

Transformation	Image of the Unit Square	Standard Matrix
Projection onto the x_1-axis		$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Projection onto the x_2-axis		$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

Chapter 2

Matrix Algebra

2.1 Matrix Operations

Definition 2.1.1. 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 $\mathbf{a}_1, \ldots \mathbf{a}_n$. We often write:

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

The **diagonal entries** of $A = [a_{ij}]$ are $a_{11}, a_{22}, a_{33}, \ldots$, and they form the **main diagonal** of A. A **diagonal matrix** is an $n \times n$ square matrix whose nondiagonal entries are all zero. A **zero matrix** is an $m \times n$ matrix whose entries are all zero.

Definition 2.1.2. 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 2.1.3. 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 2.1.4. 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 2.1.5. 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 2.1.6. *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 2.1.7. 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 2.1.8. If the number of columns of A doesn't match the number of rows of B, then the product AB is *undefined.*

Example 2.1.9. 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 2.1.10 (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 2.1.11. With A and B from Example 2.1.9, compute AB using the row-column rule.

Theorem 2.1.12. *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. $A(BC) = (AB)C$

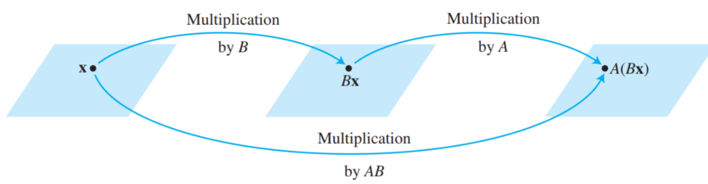
b. $A(B+C) = AB + AC$

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

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

e. $I_m A = A = A I_n$

Remark 2.1.13. When a matrix B multiplies a vector $\mathbf{x}$, it transforms $\mathbf{x}$ into $B\mathbf{x}$. If this vector is multiplied by a second matrix A, the resulting vector is $A(B\mathbf{x})$. We can think about this as a *composition* of mappings. The matrix product is defined in a special way so that $A(B\mathbf{x}) = (AB)\mathbf{x}$.



Example 2.1.14. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that first reflects points through the horizontal x_1-axis, and then reflects them through the line $x_2 = x_1$. Find the standard matrix of T.

Example 2.1.15. 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! 2.1.16. 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 2.1.17. If A is an $n \times n$ square matrix and k is a positive integer, then we denote

$$A^k = AA \cdots A \text{ (k times)}$$

We adopt the convention that $A^0 = I_n$.

Definition 2.1.18. 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 2.1.19. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, $B = \begin{bmatrix} 8 & 4 \\ 5 & 5 \\ 6 & 2 \end{bmatrix}$, and $C = \begin{bmatrix} 5 & -2 & 1 & 3 \\ 3 & 1 & 2 & -6 \end{bmatrix}$.

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

Theorem 2.1.20. *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. $(AB)^T = B^T A^T$

2.2 The Inverse of a Matrix

Definition 2.2.1. 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 2.2.2. 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 2.2.3. 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 2.2.4. *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 2.2.5. *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.*

Remark 2.2.6. The quantity $ad - bc$ is called the **determinant** of A, and we write

$$\det A = ad - bc.$$

The theorem says that a 2×2 matrix A is invertible if and only if $\det A \neq 0$.

Example 2.2.7. Find the inverse of $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Theorem 2.2.8. *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 2.2.9. Solve the system

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

Definition 2.2.10. An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix.

Example 2.2.11. $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$.
Find the products E_1A, E_2A, and E_3A, and describe how these products can be obtained by elementary row operations on A. Find an elementary matrix E such that

$$EA = \begin{bmatrix} a & b & c \\ d-2a & e-2b & f-2c \\ g & h & i \end{bmatrix}.$$

Observation 2.2.12. If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.

Observation 2.2.13. Each elementary matrix E is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

Example 2.2.14. Find the inverses of $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Theorem 2.2.15. *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 2.2.16. 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 2.2.17. Find the inverse of $A = \begin{bmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix}$.

2.3 Characterizations of Invertible Matrices

Definition 2.3.1. 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 2.3.2. 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 2.3.3. 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 2.3.4. *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 2.3.5. *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.*

Remark 2.3.6. The quantity $ad - bc$ is called the **determinant** of A, and we write

$$\det A = ad - bc.$$

The theorem says that a 2×2 matrix A is invertible if and only if $\det A \neq 0$.

Example 2.3.7. Find the inverse of $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Theorem 2.3.8. *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 2.3.9. Solve the system

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

Definition 2.3.10. An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix.

Example 2.3.11. $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$.
Find the products E_1A, E_2A, and E_3A, and describe how these products can be obtained by elementary row operations on A. Find an elementary matrix E such that

$$EA = \begin{bmatrix} a & b & c \\ d-2a & e-2b & f-2c \\ g & h & i \end{bmatrix}.$$

Observation 2.3.12. If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.

Observation 2.3.13. Each elementary matrix E is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

Example 2.3.14. Find the inverses of $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Theorem 2.3.15. *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 2.3.16. 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 2.3.17. Find the inverse of $A = \begin{bmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{bmatrix}$.

2.4 Matrix Factorizations

Definition 2.4.1. A matrix with zeros below the main diagonal is called **upper triangular**. A matrix with zeros above the main diagonal is called **lower triangular**.

Suppose $A = LU$ where L is lower triangular, and U is upper triangular. Then the equation $A\mathbf{x} = \mathbf{b}$ can be written $LU\mathbf{x} = L(U\mathbf{x}) = \mathbf{b}$. Writing $\mathbf{y} = U\mathbf{x}$, we can find $\mathbf{x}$ by solving the *pair* of equations

$$L\mathbf{y} = \mathbf{b} \qquad\qquad U\mathbf{x} = \mathbf{y}$$

First solve $L\mathbf{y} = \mathbf{b}$ for $\mathbf{y}$, and then solve $U\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$.

Example 2.4.2. Suppose $A = \begin{bmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & 0 & -5 \\ -9 & 5 & -5 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 2 & -5 & 1 & 0 \\ -3 & 8 & 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -7 & -2 & 2 \\ 0 & -2 & -1 & 2 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$.

Use this factorization of A to solve $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = (-9, 5, 7, 11)$.

Remark 2.4.3. This factorization is useful for solving equations with the same coefficient matrix:

$$A\mathbf{x} = \mathbf{b}_1,\ A\mathbf{x} = \mathbf{b}_2,\ \ldots,\ A\mathbf{x} = \mathbf{b}_p$$

If we find a factorization when solving $A\mathbf{x} = \mathbf{b}_1$, we can use it to solve the remaining equations.

Definition 2.4.4. Let A be an $m \times n$ matrix that can be reduced to echelon form *without row interchanges*. Then A can be written in the form $A = LU$ where L is an $m \times m$ lower triangular matrix with ones on the diagonal, and U is an $m \times n$ upper triangular matrix. This factorization is called an **LU factorization**. The matrix L is invertible and called a *unit* lower triangular matrix.

Suppose that A can be reduced to echelon form U using only row replacements that add multiples of one row to another row *below* it. In this case, there are unit lower triangular elementary matrices $E_1, \ldots, E_p$ such that $E_p \cdots E_2 E_1 A = U$. Then $A = (E_p \cdots E_1)^{-1} U = LU$, where $L = (E_p \cdots E_1)^{-1}$.

Procedure 2.4.5 (Algorithm for an LU factorization)**.**

1. Reduce A to echelon form U by a sequence of row replacements.
2. Place entries in L such that the *same sequence* of row replacements reduces L to I.

Example 2.4.6. Find an LU factorization of $A = \begin{bmatrix} 3 & -7 & -2 & 2 \\ -3 & 5 & 1 & 0 \\ 6 & -4 & 0 & -5 \\ -9 & 5 & -5 & 12 \end{bmatrix}$.

Example 2.4.7. Find an LU factorization of $A = \begin{bmatrix} 2 & 4 & -1 & 5 & -2 \\ -4 & -5 & 3 & -8 & 1 \\ 2 & -5 & -4 & 1 & 8 \\ -6 & 0 & 7 & -3 & 1 \end{bmatrix}$.

Example 2.4.8. Find an LU factorization of $A = \begin{bmatrix} 2 & -4 & -2 & 3 \\ 6 & -9 & -5 & 8 \\ 2 & -7 & -3 & 9 \\ 4 & -2 & -2 & -1 \\ -6 & 3 & 3 & 4 \end{bmatrix}$.

Example 2.4.9. Let $A = \begin{bmatrix} 4 & 3 & -5 \\ -4 & -5 & 7 \\ 8 & 6 & -8 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} 2 \\ -4 \\ 6 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$. Solve $A\mathbf{x} = \mathbf{b}_1$ and $A\mathbf{x} = \mathbf{b}_2$.

Chapter 3

Determinants

3.1 Introduction to Determinants

Definition 3.1.1. For $n \geq 2$, let $A = [a_{ij}]$ be a $n \times n$ matrix. We define $A_{k\ell}$ to be the $(n-1) \times (n-1)$ matrix obtained by deleting the kth row and ℓth column of A. We also set $\det(a) = a$ for any real number a. The **determinant** of A is the alternating sum

$$|A| = \det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} - a_{14} \det A_{14} + \cdots + (-1)^{n+1} \det A_{1n} a_{1n}.$$

Remark 3.1.2. This is a *recursive* definition. That is, we need to know how to compute the determinants of the $A_{k\ell}$ first, before we can compute the determinant of A.

Example 3.1.3. Compute the determinant of $A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 3 & -2 \\ 0 & -3 & 0 \end{bmatrix}$

Definition 3.1.4. Given $A = [a_{ij}]$, the (i, j)**-cofactor** of A is the number

$$C_{ij} = (-1)^{i+j} \det A_{ij}$$

Theorem 3.1.5. *The determinant of an $n \times n$ matrix A can be computed by a* **cofactor expansion** *across any row or down any column. The expansion of across the ith row is*

$$|A| = \det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}.$$

The cofactor expansion down the jth column is

$$|A| = \det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

Example 3.1.6. Use a cofactor expansion across the third row to compute $\det A$ where

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

Example 3.1.7. Compute the determinant of $A = \begin{bmatrix} 3 & 1 & -2 & 6 & 1 \\ 0 & 2 & 5 & -2 & 3 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 2 & 3 & -2 \\ 0 & 0 & 0 & -3 & 0 \end{bmatrix}$

Theorem 3.1.8. *If A is an $n \times n$ triangular matrix, then*

$$\det A = a_{11}a_{22}a_{33}\cdots a_{nn}.$$

Remark 3.1.9. This suggests a nice strategy. Turn A into a triangular matrix! We could try to reduce A to echelon form, U. How are determinants affected by row operations?

3.2 Properties of Determinants

Theorem 3.2.1 (Row Operations)**.** *Let A be a square matrix.*

(a) *If a multiple of one row of A is added to another to produce B, then* $\det B = \det A$.

(b) *If two rows of A are interchanged to produce B, then* $\det B = -\det A$.

(c) *If one row of A is multiplied by k to produce B, then* $\det B = k \det A$.

Example 3.2.2. Compute $\det A$ where $A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$

Example 3.2.3. Compute $\det A$, where $A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$.

Suppose an $n \times n$ matrix A can be reduced to echelon form U using only row replacements and row interchanges. Since U is in echelon form, it is triangular, so $\det U = u_{11}u_{22}u_{33}\cdots u_{nn}$.

Proposition 3.2.4. *If an $n \times n$ matrix A can be reduced to echelon form U using only row replacements and k row interchanges, then*

$$\det A = (-1)^k u_{11}u_{22}u_{33}\cdots u_{nn}.$$

Theorem 3.2.5. *A square matrix A is invertible if and only if* $\det A \neq 0$.

Example 3.2.6. Compute $\det A$, where $A = \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ -6 & 7 & -7 & 4 \\ -5 & -8 & 0 & 9 \end{bmatrix}$.

Example 3.2.7. Compute $\det A$, where $A = \begin{bmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ -2 & -5 & 4 & -2 \end{bmatrix}$.

Theorem 3.2.8. *If A and B are $n \times n$ matrices, then* $\det AB = (\det A)(\det B)$.

Example 3.2.9. Verify Theorem 3.2.8 for $A = \begin{bmatrix} 1 & 0 \\ 2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$.

Example 3.2.10. Let A and P be square matrices with P invertible, and show that $\det(PAP^{-1}) = \det A$.

Theorem 3.2.11. *If A is an $n \times n$ matrix, then* $\det A^T = \det A$.

Remark 3.2.12. This means we can perform operations on the *columns* of a matrix in the same way that we perform row operations, and expect the same effect on the determinant.

Example 3.2.13. Compute $\det A$, where $A = \begin{bmatrix} -5 & 2 & 2 & 2 \\ 3 & 0 & 3 & 5 \\ -4 & 0 & 4 & 0 \\ -2 & 0 & 2 & -2 \end{bmatrix}$.

Theorem 3.2.14 ("Column" Operations)**.** *Let A be a square matrix.*

(a) *If a multiple of one column of A is added to another to produce B, then* $\det B = \det A$.

(b) *If two columns of A are interchanged to produce B, then* $\det B = -\det A$.

(c) *If one column of A is multiplied by k to produce B, then* $\det B = k \det A$.

3.3 Cramer's Rule

Theorem 3.3.1 (Cramer's Rule)**.** *Let A be an invertible $n \times n$ matrix. For any $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by*

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \qquad i = 1, 2, \ldots, n$$

where

$$A_i(\mathbf{b}) = [\ \mathbf{a}_1 \quad \cdots \quad \mathbf{a}_{i-1} \quad \mathbf{b} \quad \mathbf{a}_{i+1} \quad \cdots \quad \mathbf{a}_n\ ]$$

Example 3.3.2. Use Cramer's rule to solve the system

$$\begin{cases} 3x_1 - 2x_2 = 6 \\ -5x_1 + 4x_2 = 8 \end{cases}$$

Example 3.3.3. Consider the following system of equations where $a \neq 0$. Prove that if $a \neq 24$ then the system has exactly one solution. What does the solution set look like?

$$\begin{cases} ax - 6y = -1 \\ 4x - y = 3 \end{cases}$$

Definition 3.3.4. The **adjugate** (or **classical adjoint**) of A, is

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

Where $C_{ij} = (-1)^{i+j} \det A_{ij}$.

Theorem 3.3.5 (Inverse Formula)**.** *Let A be an invertible $n \times n$ matrix. Then*

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

Example 3.3.6. Find the inverse of the matrix

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

Remark 3.3.7. This formula for the inverse is really useful for theoretical calculations, but in almost all cases, our algorithm of reducing to the identity is ***much*** more efficient.

Theorem 3.3.8. *If A is a 2×2 matrix, then the area of the parallelogram determined by the columns of A is $|\det A|$. If A is a 3×3 matrix, then the volume of the parallelepiped determined by the columns of A is $|\det A|$.*

Example 3.3.9. Calculate the area of the parallelogram with vertices $(0, 0)$, $(1, 2)$, $(2, 3)$ and $(3, 5)$.

Example 3.3.10. Find the area of the parallelogram with vertices $(-2, -2)$, $(0, 3)$, $(4, -1)$ and $(6, 4)$.

Example 3.3.11. Find the area of the parallelepiped with one vertex at the origin, and adjacent vertices $(1, 0, -1)$, $(4, 5, 6)$, $(7, 3, 9)$.

Example 3.3.12. Find the area of the triangle with vertices $(1, 2)$, $(4, 3)$, and $(3, 5)$.

3.3.1 Linear Transformations

Theorem 3.3.13. *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is a parallelogram in $\mathbb{R}^2$, then*

$$\{\textit{area of } T(S)\} = |\det A| \cdot \{\textit{area of } S\}$$

If $T : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation determined by a 3×3 matrix A, and S is a parallelepiped in $\mathbb{R}^3$, then

$$\{\textit{volume of } T(S)\} = |\det A| \cdot \{\textit{volume of } S\}$$

Example 3.3.14. Let S be the parallelogram determined by the vectors $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, and $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with standard matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & 5 \end{bmatrix}$. Find the area of $T(S)$.

Chapter 4

Vector Spaces

4.1 Vector Spaces and Subspaces

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 4.1.1. 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 4.1.2. Show the set of all $m\times n$ matrices with entries in $\mathbb{R}$, denoted $M_{m\times n}$, with addition and scalar multiplication defined using addition and scalar multiplication of matrices is a vector space.

The following properties of vector spaces are also useful.

Proposition 4.1.3. *For each* $\mathbf{u}$ *in* V *and scalar* c*,*

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

Example 4.1.4. Show that the set $\mathbb{P}_n$ of all polynomials of degree at most n with addition by combining like coefficients, and multiplication by c by scaling each coefficient by c is a vector space.

Example 4.1.5. Show that the set W of all real-valued functions on $\mathbb{R}$ with addition given by $f + g = (f + g)(x)$ for all $f, g \in W$, and scalar multiplication given by $cf = cf(x)$ is a vector space.

Definition 4.1.6. 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 4.1.7. The set $\{\mathbf{0}\}$ is a subspace of any vector space, called the **zero subspace**.

Example 4.1.8. 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 4.1.9. Show that $H = \{f : \mathbb{R} \to \mathbb{R} : f \text{ is differentiable}\}$ is a subspace of W, from 4.1.5.

Example 4.1.10. 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 4.1.11. *If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are in a vector space V, then Span$\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V.*

Definition 4.1.12. We call $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ the **subspace spanned** (or **generated**) by $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$. For any subspace H of V, a **spanning set** (or **generating set**) for H is a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ such that $H = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

Example 4.1.13. Let

$$H = \{(a - 3b, b - a, a, b) : a \text{ and } b \text{ in } \mathbb{R}\}.$$

Show that H is a subspace of $\mathbb{R}^4$.

Example 4.1.14. Show $H = \{at^2 + at + a : a \text{ in } \mathbb{R}\}$ is a subspace of $\mathbb{P}_n$.

Example 4.1.15. Show that $\mathbb{D} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x^2 + y^2 < 1 \right\}$ is not a subspace of $\mathbb{R}^2$.

Example 4.1.16. Let V be the set of all arrows in three dimensional space with two arrows considered as "equal" if they point in the same direction and have the same length. Define addition using the parallelogram rule, and define $c\mathbf{v}$ as the vector whose length is $|c|$ times the length of $\mathbf{v}$, in the same direction as $\mathbf{v}$ if $c \geq 0$, and the opposite direction if $c < 0$. Show that V is a vector space.

4.2 Null Spaces, Column Spaces, Linear Transformations

Definition 4.2.1. 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 4.2.2. 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 $\operatorname{Nul} A$.

Theorem 4.2.3. *The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to a system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.*

Example 4.2.4. Let H be the set of all vectors in $\mathbb{R}^4$ whose coordinates satisfy the equations $x_1 - 2x_2 + 5x_3 = x_4$ and $x_3 - x_1 = x_2$. Show that H is a subspace of $\mathbb{R}^4$.

Example 4.2.5. Find a spanning set for the null space of the matrix

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

Remark 4.2.6. These points will be useful later on:

1. The method in Example 4.2.5 gives a spanning set that's automatically linearly independent.
2. When $\operatorname{Nul} A$ contains nonzero vectors, the number of vectors in the spanning set of $\operatorname{Nul} A$ equals the number of free variables in the equation $A\mathbf{x} = \mathbf{0}$.

Definition 4.2.7. 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 $A = [\ \mathbf{a}_1 \ \cdots \ \mathbf{a}_n\ ]$, then

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

Remark 4.2.8. 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 4.2.9. 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 4.2.10. *The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.*

Theorem 4.2.11. *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 $\mathbb{R}^m$.*

Example 4.2.12. 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* $\mathbb{R}^m$.

Definition 4.2.13. A **linear transformation** T from a vector space V into a vector space W is a rule that assigns to each vector $\mathbf{x}$ in V one and only one vector $T(\mathbf{x})$ in W such that for all $\mathbf{u}, \mathbf{v}$ in V and real number c,

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

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$

The **kernel** (or **null space**) of T is the set of all $\mathbf{u}$ in V such that $T(\mathbf{u}) = \mathbf{0}$. The **range** of T is the set of all vectors in W of the form $T(\mathbf{x})$ for some $\mathbf{x}$ in V.

Example 4.2.14. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $T(\mathbf{x}) = A\mathbf{x}$. What are the kernel and range of T?

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

Example 4.2.15. Let V be the space of all differentiable functions whose derivatives are continuous, and W be the space of all continuous functions. Show that $D : V \to W$ by $f \mapsto f'$ is a linear transformation. What is the kernel of D? What is the range?

4.3 Linearly Independent Sets; Bases

Definition 4.3.1. An indexed set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of two or more vectors in a vector space V is called **linearly independent** if the vector equation

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{0} \qquad (\star)$$

has *only* the trivial solution $c_1 = 0, \ldots, c_p = 0$. The set S is called **linearly dependent** if there are $c_1, \ldots, c_p$ *not all* zero, such that $(\star)$ holds. In this case, $(\star)$ is called a **linear dependence relation**.

Theorem 4.3.2. *An indexed set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of two or more vectors, with $\mathbf{v}_1 \neq \mathbf{0}$, is linearly dependent if and only if some $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors, $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.*

Example 4.3.3. Let $\mathbf{p}_1(t) = 1$, $\mathbf{p}_2(t) = t^2$, $\mathbf{p}_3(t) = 4 - t^2$ in $\mathbb{P}_2$. Is $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ linearly independent?

Example 4.3.4. Let $C[0, 1]$ be the space of real-valued continuous functions on $0 \leq t \leq 1$. Is $\{\sin^2 t, \cos^2 t\}$ linearly independent? Is $\{1, \sin^2 t, \cos^2 t\}$?

Definition 4.3.5. Let H be a subspace of a vector space V. An indexed set of vectors $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ in V is a **basis** for H if

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

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

$$H = \text{Span}(\mathcal{B}) = \text{Span}\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$$

Remark 4.3.6. Since $H = V$ is a subspace of V, we can also talk about a basis for V.

Example 4.3.7. Let A be an invertible $n \times n$ matrix, and $\mathcal{B} = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$. Is $\mathcal{B}$ a basis for $\mathbb{R}^n$?

Example 4.3.8. Let $\mathcal{B} = \{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ be the columns of the $n \times n$ identity matrix I. Show that $\mathcal{B}$ is a basis for $\mathbb{R}^n$. This is called the **standard basis** for $\mathbb{R}^n$.

Example 4.3.9. 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 $\mathbb{R}^3$?

Example 4.3.10. Verify $\mathcal{B} = \{1, t, t^2, \cdots, t^n\}$ is a basis for $\mathbb{P}_n$. This is the **standard basis** for $\mathbb{P}_n$.

4.3.1 The spanning set theorem

Example 4.3.11. Let $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 6 \\ 16 \\ -5 \end{bmatrix}$, and $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. Verify that $\mathbf{v}_3 = 5\mathbf{v}_1 + 3\mathbf{v}_2$, and $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. What is a basis for H?

Definition 4.3.12. Let $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ be a set in V, and let $H = \text{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$.

(a) If one of the vectors in S, say $\mathbf{v}_k$, is a linear combination of the remaining vectors in S, then the set formed by removing $\mathbf{v}_k$ from S still spans H.

(b) If $H \neq \{\mathbf{0}\}$, some subset of S is a basis for H.

4.3.2 Bases for $\operatorname{Col} A$ and $\operatorname{Nul} A$

Example 4.3.13. 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 4.3.14. Below, A is row equivalent to U from the last example. Find a basis for $\operatorname{Col} A$.

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

Theorem 4.3.15. *The pivot columns of a matrix A form basis for* $\operatorname{Col} A$.

Watchout! 4.3.16. We need to reduce A to echelon form U to find pivot columns. However, the pivot columns of U ***do not*** form a basis for $\operatorname{Col} A$. You have to use the pivot columns of A.

Example 4.3.17. Find a basis for $\operatorname{Nul} A$, where A is the same as the previous example:

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

4.3.3 Two views of a basis

Example 4.3.18. 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 4.3.19. 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.

4.4 Coordinate Systems

An important reason for specifying a basis $\mathcal{B}$ for a vector space V is to give V a "coordinate system." We will show that if $\mathcal{B}$ contains n vectors, then the coordinate system makes V look like $\mathbb{R}^n$.

Theorem 4.4.1. *Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. Then for each $\mathbf{x}$ in V, there exists a unique set of scalars $c_1, \ldots, c_n$ such that*

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

Definition 4.4.2. Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis for V, and $\mathbf{x}$ is in V. The **coordinates of x relative to** $\mathcal{B}$ (or the $\mathcal{B}$**-coordinates of x**) are the weights $c_1, \ldots, c_n$ such that $\mathbf{x} = c_1\mathbf{b}_1 + \cdots + c_p\mathbf{b}_n$. The vector in $\mathbb{R}^n$

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

is called the **coordinate vector of x** (relative to $\mathcal{B}$), and the mapping from V to $\mathbb{R}^n$ by $\mathbf{x} \mapsto \big[\,\mathbf{x}\,\big]_{\mathcal{B}}$ is called the **coordinate mapping** (determined by $\mathcal{B}$).

Example 4.4.3. Consider the basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ of $\mathbb{R}^2$. Suppose $\mathbf{x}$ has the coordinate vector $\big[\,\mathbf{x}\,\big]_{\mathcal{B}} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$. Find $\mathbf{x}$.

Example 4.4.4. Consider the **standard basis** for $\mathbb{R}^2$, $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$. Let $\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$. Find $\big[\,\mathbf{x}\,\big]_{\mathcal{B}}$.

Example 4.4.5. In the previous two examples, we considered the coordinates of $\mathbf{x} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}$ in $\mathbb{R}^2$ relative to the bases $\mathcal{B}$ and $\mathcal{E}$. Interpret these examples graphically.

Once we fix a basis $\mathcal{B}$ for $\mathbb{R}^n$, the $\mathcal{B}$-coordinates of a specified $\mathbf{x}$ are easy to find:

Example 4.4.6. Let $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$ be a basis for $\mathbb{R}^2$. Find the coordinate vector of $\begin{bmatrix} 4 \\ 5 \end{bmatrix}$.

The matrix we used in the previous example changed the $\mathcal{B}$-coordinates of a vector $\mathbf{x}$ into the standard coordinates for $\mathbf{x}$. We can generalize this to $\mathbb{R}^n$:

Definition 4.4.7. Suppose $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is a basis for $\mathbb{R}^n$, and define the matrix $P_{\mathcal{B}} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix}$. Then the vector equation $\mathbf{x} = c_1\mathbf{b}_1 + \cdots c_n\mathbf{b}_n$ is equivalent to

$$\mathbf{x} = P_{\mathcal{B}}\big[\,\mathbf{x}\,\big]_{\mathcal{B}}.$$

$P_{\mathcal{B}}$ is called the **change-of-coordinates matrix** from $\mathcal{B}$ to the standard basis for $\mathbb{R}^n$.

Theorem 4.4.8. *Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. Then the mapping $\mathbf{x} \mapsto \left[\,\mathbf{x}\,\right]_{\mathcal{B}}$ is a one-to-one and onto linear transformation from V to $\mathbb{R}^n$.*

Remark 4.4.9. A one-to-one linear transformation between a vector space V onto another vector space W is called an ***isomorphism***, from the Greek words *iso* meaning "the same," and *morph* meaning "structure." The map $\mathbf{x} \mapsto \left[\,\mathbf{x}\,\right]_{\mathcal{B}}$ gives us a way to view V as indistinguishable from $\mathbb{R}^n$.

Example 4.4.10. Let $\mathcal{B} = \{1, t, t^2\}$ be the standard basis of $\mathbb{P}_2$. Let $\mathbf{p}_0 = a_0 + a_1 t + a_2 t^2$, $\mathbf{p}_1 = t^2$, $\mathbf{p}_2 = 4 + t + 5t^2$ and $\mathbf{p}_3 = 3 + 2t$.

(a) Find $\left[\,\mathbf{p}\,\right]_{\mathcal{B}}$ for $\mathbf{p} = \mathbf{p}_0, \ldots, \mathbf{p}_3$.

(b) Use coordinate vectors to show that $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$ are linearly independent.

Remark 4.4.11. In this example, $\mathbb{P}_2$ is *isomorphic* to $\mathbb{R}^3$. In general, $\mathbb{P}_n$ is isomorphic to $\mathbb{R}^{n+1}$.

Example 4.4.12. Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 3 \\ 12 \\ 7 \end{bmatrix}$. Suppose $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

(a) Find a basis $\mathcal{B}$ for H.

(b) Show that $\big[\,\mathbf{x}\,\big]_{\mathcal{B}}$ is a map from H to $\mathbb{R}^2$, hence, an *isomorphism* between H and $\mathbb{R}^2$.

(c) Show $\mathbf{x}$ is in H, and find the coordinate vector of $\mathbf{x}$ relative to $\mathcal{B}$.

Remark 4.4.13. This example shows that $\mathbf{v}_1, \mathbf{v}_2$ span a plane in $\mathbb{R}^3$ that is *isomorphic* to $\mathbb{R}^2$. In fact, if $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a linearly independent set of vectors in $\mathbb{R}^m$, then $H = \text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is isomorphic to $\mathbb{R}^n$ under the map $\mathbf{x} \mapsto \big[\,\mathbf{x}\,\big]_{\mathcal{B}}$ where $\mathcal{B} = S$.

Definition 4.4.14. If V is spanned by a finite set, then the **dimension** of V is the number of vectors in a basis for V. The dimension of the zero space, $\{\mathbf{0}\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be **infinite-dimensional**.

Chapter 5

Eigenvectors and Eigenvalues

5.1 Eigenvectors and Eigenvalues

Example 5.1.1. Let $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$.

Remark 5.1.2. In this example, it turns out $A\mathbf{v}$ is just $2\mathbf{v}$, so A only stretches $\mathbf{v}$.

Definition 5.1.3. An **eigenvector** of an $n \times n$ matrix is a nonzero vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x} = \mathbf{v}$ of the equation $A\mathbf{x} = \lambda\mathbf{x}$; such a $\mathbf{v}$ is called an *eigenvector corresponding to* λ.

Example 5.1.4. Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$.

(a) Are $\mathbf{u}$ and $\mathbf{v}$ eigenvectors of A?

(b) Show that 7 is an eigenvalue of $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$.

Procedure 5.1.5 (Determining if λ is an eigenvalue)**.** The scalar λ is an eigenvalue for a matrix A if and only if the equation

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

has a nontrivial solution. Just reduce the associated augmented matrix!

Definition 5.1.6. The set of all solutions to $A\mathbf{x} = \lambda\mathbf{x}$ is the nullspace of the matrix $A - \lambda I$, and therefore is a subspace of $\mathbb{R}^n$. We call this the **eigenspace** of A corresponding to λ.

Remark 5.1.7. Even though we used row reduction to find eigen*vectors*, we cannot use it to find eigen*values*. An echelon for a matrix A doesn't usually have the same eigenvalues as A.

Example 5.1.8. Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. Find a basis for the eigenspace corresponding to $\lambda = 2$.

Theorem 5.1.9. *The eigenvalues of a triangular matrix are the entries on its main diagonal.*

Example 5.1.10. Let $A = \begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & 3 & 4 \end{bmatrix}$. What are the eigenvalues of A and B? What does it mean for A to have an eigenvalue of 0?

Theorem 5.1.11. *If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent.*

Example 5.1.12. Let $C = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Find the eigenspaces corresponding to $\lambda = 0, 1$.

Remark 5.1.13. Note, the matrix C is RREF form for A, but the eigenvalues are different.

In the next section, we'll be using determinants to find eigenvalues of a matrix. We'll close this section by reviewing some of the properties we know for determinants.

Proposition 5.1.14. *Suppose A is an $n \times n$ matrix that can be reduced to echelon form U using only row replacements and r row interchanges. Then the determinant of A is*

$$\det A = (-1)^r \cdot u_{11}u_{22} \cdots u_{nn}.$$

Proposition 5.1.15. *Let A and B be $n \times n$ matrices.*

(a) *A is invertible if and only if* $\det A \neq 0$.

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

(c) $\det A^T = \det A$.

(d) *If A is triangular,* $\det A = a_{11}a_{22} \cdots a_{nn}$.

(e) *A row replacement does not change the determinant. A row interchange changes the sign of the determinant. Scaling a row scales the determinant by the same factor.*

We also recall the invertible matrix theorem.

Theorem 5.1.16 (The Invertible Matrix Theorem)**.** *Let A be a square $n \times n$ matrix. Then the following statements are equivalent (i.e. they're either all true or all false).*

(a) *A is an invertible matrix.*

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

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

(d) *A is row equivalent to I_n.*

(e) *A^T is an invertible matrix.*

(f) *A has n pivot positions.*

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

(h) *$A\mathbf{x} = \mathbf{b}$ has a solution for all $\mathbf{b}$ in $\mathbb{R}^n$.*

(i) *The columns of A span $\mathbb{R}^n$*

(j) *The columns of A are linearly independent.*

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

(l) *The transformation $\mathbf{x} \mapsto A\mathbf{x}$ is onto.*

We can also add the following to the list:

(m) *The determinant of A is not zero.*

(n) *The number 0 is not an eigenvalue of A*

5.2 The Characteristic Equation

Example 5.2.1. Find the eigenvalues of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

Definition 5.2.2. The equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A.

Proposition 5.2.3. *A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the characteristic equation*

$$\det(A - \lambda I) = 0.$$

Example 5.2.4. Find the characteristic equation and eigenvalues of $A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Definition 5.2.5. If A is an $n \times n$ matrix, then $\det(A - \lambda I)$ is a polynomial of degree n called the **characteristic polynomial** of A. The **multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic polynomial.

Example 5.2.6. The characteristic polynomial of a 6×6 matrix A is $\lambda^6 - 4\lambda^5 - 12\lambda^4$. Find the eigenvalues of A and their multiplicities.

Example 5.2.7. Find the eigenvalues and bases for the corresponding eigenspaces of

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

1. Let H be a plane through the origin in $\mathbb{R}^3$. Let T be a transformation with standard matrix A that reflects points through the plane H. Without finding a matrix, find the eigenvalues of A and their multiplicities, and explain your reasoning.

5.3 Diagonalization

In this section, we will use eigenvalues and eigenvectors to factor a matrix A as $A = PDP^{-1}$, where D is a diagonal matrix. This factorization helps us compute A^k very quickly.

Example 5.3.1. Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$, $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$ and $P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}$. Then, $A = PDP^{-1}$. Find a formula for A^k.

Definition 5.3.2. If A and B are $n \times n$ matrices, then A is **similar** to B if there is an invertible matrix P such that $A = PBP^{-1}$. We call A **diagonalizable** if A is similar to a diagonal matrix D.

Theorem 5.3.3. If A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues.

Theorem 5.3.4 (The Diagonalization Theorem)**.** An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors. In fact, $A = PDP^{-1}$, with D diagonal, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are the eigenvalues of A that correspond, respectively, to the eigenvectors in P.

Remark 5.3.5. In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis for $\mathbb{R}^n$, called an **eigenvector basis** for $\mathbb{R}^n$.

Example 5.3.6. Let $A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$. Diagonalize A, if possible.

Strategy 5.3.7 (Diagonalizing a matrix)**.** If A is an $n \times n$ matrix, then to try to diagonalize A we

1. Find the eigenvalues of A
2. Find n linearly independent eigenvectors of A
3. Construct P from the vectors in step 2
4. Construct D from the corresponding eigenvalues
5. Profit

If A is *not* diagonalizable, we will find out in step 2, when we can't find enough eigenvectors.

Example 5.3.8. Let $A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$. Diagonalize A, if possible.

Theorem 5.3.9. An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

Example 5.3.10. Let $A = \begin{bmatrix} 5 & -8 & 1 \\ 0 & 0 & 7 \\ 0 & 0 & -2 \end{bmatrix}$.

(a) Determine if A is diagonalizable.

(b) Diagonalize A, if possible.

Theorem 5.3.11. Let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_p$.

(a) For $1 \leq k \leq p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.

(b) The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, which happens if and only if

  (i) the characteristic polynomial factors completely into linear factors, and

  (ii) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.

(c) If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k$ forms an eigenvector basis for $\mathbb{R}^n$.

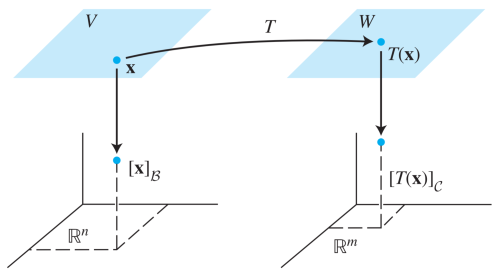
Example 5.3.12. Diagonalize the following matrix, if possible,

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

5.4 Eigenvectors and Linear Transformations

Recall from Section 1.9 that any linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be represented by left multiplication by the *standard matrix* of T. The goal of this section will be to find the same sort of representation for any linear transformation between two finite-dimensional vector spaces.

Example 5.4.1. Let V be an n-dimensional vector space with basis $\mathcal{B}$, let W be an m-dimensional vector space with basis $\mathcal{C}$, and let T be any linear transformation from V to W. Show that the mapping $\big[\,\mathbf{x}\,\big]_{\mathcal{B}} \mapsto \big[\,T(\mathbf{x})\,\big]_{\mathcal{C}}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and find its standard matrix, that is, find a matrix M such that $\big[\,T(\mathbf{x})\,\big]_{\mathcal{C}} = M\big[\,\mathbf{x}\,\big]_{\mathcal{B}}$.



Definition 5.4.2. In the example above, the matrix

$$M = \begin{bmatrix} \big[\,T(\mathbf{b}_1)\,\big]_{\mathcal{C}} & \big[\,T(\mathbf{b}_2)\,\big]_{\mathcal{C}} & \cdots & \big[\,T(\mathbf{b}_n)\,\big]_{\mathcal{C}} \end{bmatrix}$$

is called the **matrix for T relative to the bases $\mathcal{B}$ and $\mathcal{C}$.**

Example 5.4.3. Suppose $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$ is a basis for V and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3\}$ is a basis for W. Let $T : V \to W$ be a linear transformation such that

$$T(\mathbf{b}_1) = 3\mathbf{c}_1 - 2\mathbf{c}_2 + 5\mathbf{c}_3 \text{ and } T(\mathbf{b}_2) = 4\mathbf{c}_1 - 7\mathbf{c}_2 - 1\mathbf{c}_3.$$

Find the matrix M for T relative to $\mathcal{B}$ and $\mathcal{C}$ and determine if T is one-to-one and/or onto.

Example 5.4.4. Let $\mathbb{P}_2$ have basis $\mathcal{B} = \{1, t, t^2\}$ and let $T : \mathbb{P}_2 \to \mathbb{P}_2$. Suppose that M is the matrix for T relative to $\mathcal{B}$ (and $\mathcal{B}$) where

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

Describe T, and find its kernel and range using M.

5.4.1 Linear transformations from V into V

Definition 5.4.5. In the common case (like the previous example) where $W = V$ and $\mathcal{C} = \mathcal{B}$, the matrix M is called the **matrix for T relative to $\mathcal{B}$**, or just the **$\mathcal{B}$-matrix for T**, denoted $\big[\,T\,\big]_{\mathcal{B}}$. The $\mathcal{B}$ matrix for $T : V \to V$ satisfies

$$\big[\,T(\mathbf{x})\,\big]_{\mathcal{B}} = \big[\,T\,\big]_{\mathcal{B}}\big[\,\mathbf{x}\,\big]_{\mathcal{B}}.$$

$$\begin{tikzcd} \mathbf{x} \arrow[r, "T"] \arrow[d] & T(\mathbf{x}) \arrow[d] \\ {[\mathbf{x}]_{\mathcal{B}}} \arrow[r, "\text{Multiplication}", "\text{by } [T]_{\mathcal{B}}"'] & {[T(\mathbf{x})]_{\mathcal{B}}} \end{tikzcd}$$

Example 5.4.6. Let $D : \mathbb{P}_3 \to \mathbb{P}_3$ be the transformation that sends $\mathbf{p}(t)$ to $\mathbf{p}'(t)$. That is,

$$D(a_0 + a_1 t + a_2 t^2 + a_3 t^3) = a_1 + 2a_2 t + 3a_3 t^2.$$

(a) Find the $\mathcal{B}$-matrix for D, when $\mathcal{B} = \{1, t, t^2, t^3\}$.

(b) Verify that $\big[\,D(\mathbf{x})\,\big]_{\mathcal{B}} = \big[\,D\,\big]_{\mathcal{B}}\big[\,\mathbf{x}\,\big]_{\mathcal{B}}$

5.4.2 Linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$

When $T : \mathbb{R}^n \to \mathbb{R}^n$ is linear with standard matrix A, then if A is diagonalizable, we can find a basis $\mathcal{B}$ of $\mathbb{R}^n$ consisting of eigenvectors of A. In this case, the $\mathcal{B}$-matrix for T will be diagonal!

Theorem 5.4.7. Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If $\mathcal{B}$ is the basis for $\mathbb{R}^n$ formed by the columns of P (that is, an eigenvector basis), then D is the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$.

Example 5.4.8. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. The eigenvalues of A are $\lambda = 3, 5$.

(a) Find a basis for $\mathbb{R}^2$ with the property that the $\mathcal{B}$-matrix for T is a diagonal matrix.

(b) Define $T^n : \mathbb{R}^2 \to \mathbb{R}^2$ by applying T to $\mathbf{x}$ n-times, that is,

$$T^n(\mathbf{x}) = T(T(T(\cdots T(\mathbf{x}))) = A(A(A \cdots A\mathbf{x})) = A^n\mathbf{x}$$

Find a formula for $T^n(\mathbf{x})$.

Example 5.4.9. Define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$ where $A = \begin{bmatrix} 1/3 & 1 \\ 0 & 2/3 \end{bmatrix}$.

(a) Find a basis for $\mathbb{R}^2$ with the property that the $\mathcal{B}$-matrix for T is a diagonal matrix.

(b) Define $T^n : \mathbb{R}^2 \to \mathbb{R}^2$ by applying T to $\mathbf{x}$ n-times, that is,

$$T^n(\mathbf{x}) = T(T(T(\cdots T(\mathbf{x}))) = A(A(A \cdots A\mathbf{x})) = A^n\mathbf{x}$$

Find $\lim_{n\to\infty} T^n(\mathbf{x})$.

Chapter 6

Orthogonality and Least Squares

6.1 Inner Product, Length, and Orthogonality

Concepts of length, distance, and perpendicularity are well known for $\mathbb{R}^2$ and $\mathbb{R}^3$. The goal of this section is to extend these concepts to $\mathbb{R}^n$ for $n \geq 4$.

Definition 6.1.1. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, then $\mathbf{u}$ and $\mathbf{v}$ are $n \times 1$ matrices, the transpose $\mathbf{u}^T$ is a $1 \times n$ matrix, and the product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix. The number $\mathbf{u}^T\mathbf{v}$ is called the **dot product** (or **inner product**) of $\mathbf{u}$ and $\mathbf{v}$. It's usually denoted

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T\mathbf{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

Example 6.1.2. Compute $\mathbf{u} \cdot \mathbf{v}$ and $\mathbf{v} \cdot \mathbf{u}$ for $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}$

Theorem 6.1.3. Let $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then,

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$

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

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

(d) $\mathbf{u} \cdot \mathbf{u} \geq 0$, and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

Definition 6.1.4. The **length** (or **norm**) of $\mathbf{v}$ is the nonnegative scalar $\|\mathbf{v}\|$ defined by

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

Example 6.1.5. Let $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ Show that $\|\mathbf{v}\|$ is the length of the line from the origin to (a, b).

Definition 6.1.6. A vector whose length is 1 is called a **unit vector**. If we divide a vector by its length, we obtain a unit vector *in the same direction as* $\mathbf{v}$. This is called **normalizing v**.

Example 6.1.7. Let $\mathbf{v} = (1, -2, 2, 0)$. Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$.

Example 6.1.8. Let $\mathbf{x} = \begin{bmatrix} \frac{3}{4} \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$ and $W = \mathrm{Span}(\mathbf{x})$. Find a unit vector $\mathbf{z}$ that is a basis for W.

Recall that if a and b are real numbers, then the distance on a number line between a and b is the number $|a - b|$. This definition of distance has a direct analogue in $\mathbb{R}^n$.

Definition 6.1.9. For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the **distance between u and v**, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is,

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

Example 6.1.10. Compute the distance between the vectors $\mathbf{u} = (7, 1)$ and $\mathbf{v} = (3, 2)$.

Example 6.1.11. Compute the distance between the vectors $\mathbf{u} = (1, 2, 3)$ and $\mathbf{v} = (4, 6, 5)$.

The rest of this chapter will explore an analogue of perpendicularity in $\mathbb{R}^n$.

Definition 6.1.12. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (or perpendicular) if $\mathbf{u} \cdot \mathbf{v} = 0$.

Example 6.1.13. Show that $\mathbf{u} = (2, 6)$ and $\mathbf{v} = (-3, 1)$ are orthogonal.

Example 6.1.14. Find a nonzero vector in $\mathbb{R}^3$ that is orthogonal to $\mathbf{u} = (1, 2, 3)$.

Theorem 6.1.15 (Pythagorean Theorem)**.** Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if

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

The rest of the concepts we introduce in this section will be discussed in more detail in Section 6.3.

Definition 6.1.16. If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\mathbf{z}$ is said to be **orthogonal to** W. The set of all vectors that are orthogonal to W is called the **orthogonal complement** of W, and is denoted $W^\perp$.

Theorem 6.1.17.

1. A vector $\mathbf{x}$ is in $W^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in a spanning set of W.
2. $W^\perp$ is a subspace of $\mathbf{R}^n$.
3. $(W^\perp)^\perp = W$

Example 6.1.18. Find the orthogonal complement to $W = \text{Span}(\mathbf{u})$ where $\mathbf{u} = (1, 2, 3)$.

We'll close with an interesting connection between the column and null space of a matrix.

Theorem 6.1.19. Let A be an $m \times n$ matrix. Then,

$$(\text{Nul}\, A)^\perp = \text{Col}\, A^T \qquad (\text{Col}\, A)^\perp = \text{Nul}\, A^T$$

6.2 Orthogonal Sets

Definition 6.2.1. A set of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal. That is, if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

Example 6.2.2. Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \qquad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

Theorem 6.2.3. If $S = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent, hence a basis for the subspace spanned by S.

Definition 6.2.4. An **orthogonal basis** for the subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

The next theorem demonstrates why an orthogonal basis is nicer than other bases.

Theorem 6.2.5. Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights of the linear combination

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

are given by

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j}, \qquad (j = 1, \ldots, p)$$

Example 6.2.6. The set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^3$, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \qquad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

Express the vector $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ as a linear combination of the vectors in S.

Remark 6.2.7. If the basis in the previous example was not orthogonal, then we would have to solve a system of linear equations to find the weights, like in Chapter 1.

6.2.1 Orthogonal Projection

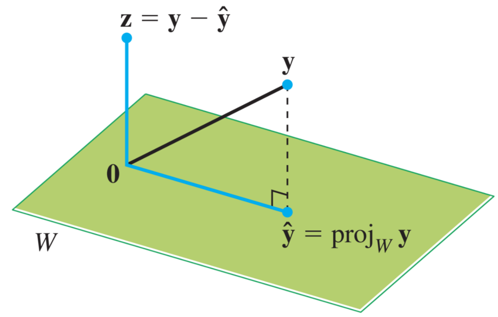
Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, and suppose we wanted to decompose a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and one orthogonal to $\mathbf{u}$. That is, we want to write

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}$$

where $\widehat{\mathbf{y}} = \alpha\mathbf{u}$ for some α.

Definition 6.2.8. In the set up above, the vector $\widehat{\mathbf{y}}$ is called the **orthogonal projection of y onto u**, and the vector $\mathbf{z}$ is called the **component of y orthogonal to u**. If L is the subspace spanned by $\mathbf{u}$, then the orthogonal projection of $\mathbf{y}$ onto any $c\mathbf{u}$ in L is the same as the projection of $\mathbf{y}$ onto $\mathbf{u}$. Thus, sometimes $\widehat{\mathbf{y}}$ is denoted $\text{proj}_L\,\mathbf{y}$ and is called the **orthogonal projection of y onto** L. It can be shown that

$$\widehat{\mathbf{y}} = \text{proj}_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$$



Example 6.2.9. Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$, then write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span$\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

6.2.2 Orthonormal Sets

Definition 6.2.10. A set $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is called an **orthonormal set** if it is an orthogonal set of unit vectors. If W is spanned by S, then S is called an **orthonormal basis** for W.

Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis for $\mathbb{R}^3$, where

$$\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \qquad \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

Theorem 6.2.11. An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

Definition 6.2.12. An invertible matrix with orthonormal columns is called an **orthogonal matrix**. Orthogonal matrices have the nice property that $U^{-1} = U^T$.

Example 6.2.13. Show $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$.
Show U has orthonormal columns and compute $\|\mathbf{x}\|$ and $\|U\mathbf{x}\|$

Theorem 6.2.14. Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then,

(a) $\|U\mathbf{x}\| = \|\mathbf{x}\|$
(b) $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
(c) $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$

Example 6.2.15. Show that $U = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix}$ is an orthogonal matrix.

6.3 Orthogonal Projections

Orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin has an important analogue in $\mathbb{R}^n$. Given a vector $\mathbf{y}$ and a subspace W in $\mathbb{R}^n$, there is a vector $\widehat{\mathbf{y}}$ in W such that (1) $\widehat{\mathbf{y}}$ is the unique vector $\mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}$ is orthogonal to W, and (2) $\widehat{\mathbf{y}}$ is the unique vector in W closest to $\mathbf{y}$.

Example 6.3.1. Let $\mathbf{u}_1, \cdots, \mathbf{u}_5$ be an orthogonal basis for $\mathbb{R}^5$ and let

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_5\mathbf{u}_5$$

Let $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, and write $\mathbf{y}$ as the sum of a vector $\mathbf{z}_1$ and a vector $\mathbf{z}_2$ in $W^\perp$.

Remark 6.3.2. Here, we were able to find the decomposition $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$ because we had a basis for both W and $\mathbb{R}^5$. Actually, we only need a basis for W!

Theorem 6.3.3. Let W be a subspace of $\mathbb{R}^n$. Each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely as

$$\mathbf{y} = \widehat{\mathbf{y}} + \mathbf{z}$$

where $\widehat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$. In fact, if $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

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

and $\mathbf{z} = \mathbf{y} - \widehat{\mathbf{y}}$. The vector $\widehat{\mathbf{y}}$ is called the **orthogonal projection y onto** W, and is usually denoted as $\operatorname{proj}_W \mathbf{y}$.

Example 6.3.4. Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Show $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, and write $\mathbf{y}$ as the sum of a vector in W and a vector in $W^{\perp}$.

Theorem 6.3.5 (The Best Approximation Theorem)**.** Let W be a subspace of $\mathbb{R}^n$, $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and $\widehat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\widehat{\mathbf{y}}$ is the closest point in W closest to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \widehat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|$$

for all $\mathbf{v}$ in W distinct from $\widehat{\mathbf{y}}$. The vector $\widehat{\mathbf{y}}$ is called **the best approximation to y by the elements of** W.

Example 6.3.6. Let $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ and find the closest point in W to $\mathbf{y}$ where

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

Example 6.3.7. The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W. Find the distance from $\mathbf{y}$ to $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ where

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

Theorem 6.3.8. If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

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

If $U = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{bmatrix}$ then

$$\widehat{\mathbf{y}} = \text{proj}_W \mathbf{y} = UU^T\mathbf{y} \text{ for all } \mathbf{y} \in \mathbb{R}^n.$$

6.4 The Gram-Schmidt Process

The Gram-Schmidt process is a simple algorithm for producing an orthogonal or orthonormal basis for any nonzero subspace of $\mathbb{R}^n$. These first two examples will try to demonstrate the process.

Example 6.4.1. Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ and construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W, where

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Next, we give a very important example that illustrates the Gram-Schmidt process.

Example 6.4.2. Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ and construct an orthogonal basis for W, where

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \text{and } \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Theorem 6.4.3. Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

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

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

$$\mathrm{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \mathrm{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \text{ for all } 1 \le k \le p$$

Remark 6.4.4. Once we have an orthogonal basis, we can easily find an orthonormal basis just by scaling all of the basis elements by their norms. Don't normalize until you've found the whole orthogonal basis, though, unless you want to work with lots of ugly square roots!

Let $W = \mathrm{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ and construct an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ for W, where

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Theorem 6.4.5 (QR Factorization)**.** A is an $m \times n$ matrix with linearly independent columns, then A can be factored as QR, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\operatorname{Col} A$ and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

Example 6.4.6. Find a QR factorization of $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.