MA281: Introduction to Linear Algebra

Dylan C. Beck

Acknowledgements

Primarily, the contents of this document were created in the Fall 2022 semester at Baker University. I would like to express my gratitude toward my students in MA281 (Introduction to Linear Algebra) — especially those who assisted in the enhancement of these notes with comments and suggestions, including Sydney Delfelder, Owen Gonzales, Ben Stubblefield, and April Thomas.

Elsewhere, some of the material from the second chapter on the Smith Normal Form, the Rational Canonical Form, and the Jordan Canonical Form was written independently from August 2018 to January 2022 at the request of students from the Algebra Qualifying Exam Study Group at the University of Kansas. I would like to thank the graduate students there for their motivation and participation in these study sessions — especially Wayne Ng Kwing King, Enrique Salcido, Neethu Suma-Raveendran, and Christopher Wong. I am also grateful to Souvik Dey and Monalisa Dutta for their assistance with these meetings and their peripheral contributions to these notes.

Contents

Chapter 1

Matrices and Vector Spaces

Often, real-world problems require us to deal with large amounts of data and information that can be most efficiently organized by rows and columns in what we will refer to as a matrix. We will soon see that matrices possess an arithmetic that yields a highly sophisticated and useful theory.

1.1 Matrices and Matrix Addition

Unless otherwise specified, we will assume throughout this chapter that m and n are positive integers. We say that a visual representation of any collection of data arranged into m rows and n columns is an $m \times n$ **array**. Each object of an $m \times n$ array A is a **component** or **element** of A. Each component of A can be uniquely identified by specifying its row and column. Explicitly, we use the symbol a_{ij} to indicate the component of A in the ith row and jth column; often, we will refer to a_{ij} as the (i, j)th **entry** of the array A. Collectively, therefore, we may view the array A as **indexed** by its objects a_{ij} for each pair of integers $1 \leq i \leq m$ and $1 \leq j \leq n$. Components of the form a_{ii} are referred to as the **diagonal** entries of A because they lie in the same row and column of A; the collection of all diagonal entries of A is called the **main diagonal** of A. We will adopt the convention that an $m \times n$ array be written using large rectangular brackets, as in the following.

Example 1.1.1. Consider the case that Alice, Bob, Carly, and Daryl play Bridge together. If Alice and Carly belong to one team and Bob and Daryl belong to the opposing team, then we may encode this information (i.e., these teams) as the two columns of the following 2×2 array T.

$$T = \begin{bmatrix} \text{Alice} & \text{Bob} \\ \text{Carly} & \text{Daryl} \end{bmatrix}$$

Observe that $t_{11} = \text{Alice}$, $t_{12} = \text{Bob}$, $t_{21} = \text{Carly}$, and $t_{22} = \text{Daryl}$. One could also just as well swap the rows and columns to display the teams as rows by constructing the following 2×2 array T^t.

$$T^t = \begin{bmatrix} \text{Alice} & \text{Carly} \\ \text{Bob} & \text{Daryl} \end{bmatrix}$$

Our principal concern throughout this course are those $m \times n$ arrays consisting entirely of (real) numbers. Under this restriction, we may refer to an $m \times n$ array as a (real) $m \times n$ **matrix**. Generally, one can define matrices consisting of elements lying in any **ring**, but we will not be so general.

Example 1.1.2. Each real number x may be viewed as a real 1×1 matrix $\begin{bmatrix} x \end{bmatrix}$.

Example 1.1.3. Consider once again the scenario of Example 1.1.1. We may assign to each player a real number called a "skill value" between 0 and 100, e.g., suppose that Alice has skill value 88; Bob has skill value 72; Carly has skill value 95; and Daryl has skill value 90. Under this convention, the matrices of Example 1.1.1 yield new matrices that we could call "skill matrices" as follows.

$$S = \begin{bmatrix} 88 & 72 \\ 95 & 90 \end{bmatrix} \text{ and } S^t = \begin{bmatrix} 88 & 95 \\ 72 & 90 \end{bmatrix}$$

Our previous three examples dealt with **square** matrices, i.e., matrices for which the number of rows and the number of columns were the same (i.e., $m = n$); however, not all matrices are square.

Example 1.1.4. Consider the 1×5 matrix $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix}$ of the first five positive integers.

We refer to matrices with only one row as **row vectors**; likewise, matrices with only one column are called **column vectors**. We will return to the notion of a vector in our study of vector spaces in Section 1.6. Often, we will also use the terminology (horizontal) n-**tuples** when discussing row vectors with n columns and (vertical) m-tuples when discussing column vectors with m rows.

Like we mentioned in the first paragraph of this section, an $m \times n$ matrix A is uniquely determined by the element a_{ij} in its ith row and jth column for each pair of integers $1 \leq i \leq m$ and $1 \leq j \leq n$. For instance, the matrix of Example 1.1.4 is the unique matrix with one row whose jth column consists of the integer j for each integer $1 \leq j \leq 5$. Under this identification, we will adopt the one-line notation $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ for the $m \times n$ matrix A with a_{ij} in its ith row and jth column.

Example 1.1.5. Consider the 2×3 matrix whose ith row and jth column consists of the sum $i + j$. We may write this symbolically (in one-line notation) as $\begin{bmatrix} i + j \end{bmatrix}_{\substack{1 \leq i \leq 2 \\ 1 \leq j \leq 3}}$ or expanded as follows.

$$\begin{array}{cc} & \begin{array}{ccc} j = 1 & j = 2 & j = 3 \end{array} \\ \begin{array}{c} i = 1 \\ i = 2 \end{array} & \begin{bmatrix} 1 + 1 & 1 + 2 & 1 + 3 \\ 2 + 1 & 2 + 2 & 2 + 3 \end{bmatrix} \end{array} \text{ or } \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix}$$

Example 1.1.6. Given any positive integers m and n, there is one and only one matrix consisting entirely of zeros: it is the $m \times n$ **zero matrix**, and it is denoted by $O_{m \times n}$.

Example 1.1.7. We refer to the matrix $I_{m \times n} = \begin{bmatrix} \delta_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ as the $m \times n$ **identity matrix**, where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and} \\ 0 & \text{if } i \neq j \end{cases}$$

is the **Kronecker delta**. Put another way, the $m \times n$ identity matrix is the unique $m \times n$ matrix whose (i, j)th component is one for each pair of integers $1 \leq i \leq m$ and $1 \leq j \leq n$ such that $i = j$ and whose other components are all zero. One can also say that $I_{m \times n}$ is the unique $m \times n$ matrix with ones along the main diagonal and zeros elsewhere. Explicitly, we have the following examples.

$$I_{2 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } I_{2 \times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \text{ and } I_{3 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } I_{3 \times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Observe that the only nonzero components of $I_{n\times n}$ lie on the main diagonal, hence $I_{n\times n}$ is a **diagonal matrix**. Explicitly, a diagonal matrix is an $n \times n$ matrix consisting entirely of zeros off the main diagonal. Even more, $I_{n\times n}$ is the unique diagonal $n \times n$ matrix whose nonzero entries are all one.

Example 1.1.8. Given any $m \times n$ matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, its **matrix transpose** A^t is the $n \times m$ matrix obtained by swapping the rows and columns of A, i.e., we have that $A^t = \begin{bmatrix} a_{ji} \end{bmatrix}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$. Put another way, the (i, j)th entry of A^t is the (j, i)th entry of A, hence the ith row of A^t is precisely the ith column of A. Explicitly, for the matrix A defined in Example 1.1.5, we have the following.

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix} \text{ and } A^t = \begin{bmatrix} 2 & 3 \\ 3 & 4 \\ 4 & 5 \end{bmatrix}$$

Observe that the first row of A becomes the first column of A^t (and likewise for the second row). Consequently, the transpose of any $1 \times n$ row vector is an $n \times 1$ column vector. We will also refer to A^t simply as the transpose of A; the process of computing A^t is called **transposition**. One other thing to notice is that it always holds that $I^t_{m\times n} = I_{n\times m}$, hence we have that $I^t_{n\times n} = I_{n\times n}$.

Definition 1.1.9. We say that an $m \times n$ matrix A is **symmetric** if it holds that $A^t = A$. Observe that a matrix is symmetric only if it is square, i.e., a non-square matrix is never symmetric.

Considering that matrices encode numerical data, it is not surprising to find that they induce their own arithmetic. Using one-line notation, matrix addition can be defined as follows.

Definition 1.1.10. Given any $m \times n$ matrices $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ and $B = \begin{bmatrix} b_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, the **matrix sum** of A and B is the $m \times n$ matrix $A + B = \begin{bmatrix} a_{ij} + b_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$. Put in words, the matrix sum $A + B$ is the $m \times n$ matrix whose (i, j)th entry is the sum of the (i, j)th entries of A and B.

Caution: the matrix sum is not defined for matrices with different numbers of rows or columns.

Example 1.1.11. We compute the matrix sum of the following 2×3 matrices.

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

Example 1.1.12. If A is any $m \times n$ matrix, then we have that $A + O_{m\times n} = A = O_{m\times n} + A$. Consequently, we may view $O_{m\times n}$ as the **additive identity** among all $m \times n$ matrices.

Generally, for any real $m \times n$ matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, then we will typically refer to any (real) number c as a **scalar**, and we define the **scalar multiple** of A by the scalar c as $cA = \begin{bmatrix} ca_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$. Essentially, we may view this as generalizing the sum of the matrix A with itself c times.

Example 1.1.13. Given any $m \times n$ matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, we will write $-A = \begin{bmatrix} -a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$. We have that $A + (-A) = O_{m\times n} = -A + A$, and we say that $-A$ is the **additive inverse** of A.

Our next proposition illustrates that matrix transposition and matrix addition are compatible.

Proposition 1.1.14. *Let A and B be any $m \times n$ matrices. We have that $(A + B)^t = A^t + B^t$. Put another way, the transpose of a sum of matrices is the sum of the matrix transposes.*

Proof. By Definition 1.1.10, the (i, j)th entry of $A+B$ is the sum of the (i, j)th entry of A and the (i, j)th entry of B. By Example 1.1.8, the (i, j)th entry of $(A+B)^t$ is the (j, i)th entry of $A+B$, i.e., the sum of the (j, i)th entry of A and the (j, i)th entry of B. But by the same example, this is the sum of the (i, j)th entry of A^t and the (i, j)th entry of B^t. Ultimately, this shows that the (i, j)th entry of $(A + B)^t$ and the (i, j)th entry of $A^t + B^t$ are the same so that $(A + B)^t = A^t + B^t$. □

1.2 Rotation Matrices and Matrix Multiplication

Let $\mathbb{R}$ denote the set of real numbers. Recall that every point (x, y) in the Cartesian plane $\mathbb{R}\times\mathbb{R}$ can be written as $(r\cos\theta, r\sin\theta)$ for some real number r and some angle θ. Explicitly, this is called the representation of the point (x, y) in **polar coordinates**. Consequently, we may specify any point in the plane by declaring that $x = r\cos\theta$ and $y = r\sin\theta$ for some real numbers r and θ. Rotation of the point (x, y) through an angle ϕ yields a new point defined by $x' = r\cos(\theta+\phi)$ and $y' = r\sin(\theta+\phi)$. Using the addition formulas for sine and cosine, we find that $x' = r(\cos\theta\cos\phi - \sin\theta\sin\phi)$ and $y' = r(\sin\theta\cos\phi + \sin\phi\cos\theta)$. Our objective in this section is to provide a more efficient method of rotating points in the plane through a specified angle ϕ. We achieve this as follows.

We have seen in the previous section that any matrix can be transposed and any two matrices can be added together to obtain new matrices. Even more, if the number of columns (or rows) of a matrix A equals the number of rows (or columns) of a matrix B, then A and B can be multiplied.

Definition 1.2.1. Given any $m \times n$ matrix $A = \big[a_{ij}\big]_{\substack{1\leq i\leq m\\ 1\leq j\leq n}}$ and any $n \times r$ matrix $B = \big[a_{ij}\big]_{\substack{1\leq i\leq n\\ 1\leq j\leq r}}$, the (left) **matrix product** of A and B is the $m \times r$ matrix AB whose (i, j)th entry is given by $\sum_{k=1}^n a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$. Put in words, the matrix product AB is the $m\times r$ matrix whose (i, j)th entry is the sum of the products of the (i, k)th entry of A and the (k, j)th entry of B for all integers $1 \leq k \leq n$. Crucially, the order of the matrices A and B in the matrix product matters; however, if we assume that $r = m$, then the (right) matrix product BA can be defined analogously. Be sure to note also that the number of rows of AB is the same as the number of rows of A, and the number of columns of AB is the same as the number of columns of B.

Caution: the product is not defined for matrices with an incompatible number of rows and columns.

Example 1.2.2. Consider the following matrices.

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

Considering that A is a 2×3 matrix and B is a 3×2 matrix, both of the products AB and BA can be formed: AB is a 2×2 matrix, and BA is a 3×3 matrix. Explicitly, they are as follows.

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix}\begin{bmatrix} -1 & 0 \\ 0 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1(-1)+2(0)+3(-1) & 1(0)+2(1)+3(1) \\ 2(-1)+3(0)+4(-1) & 2(0)+3(1)+4(1) \end{bmatrix} = \begin{bmatrix} -4 & 5 \\ -6 & 7 \end{bmatrix}$$

$$BA = \begin{bmatrix} -1 & 0 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix} = \begin{bmatrix} -1(1)+0(2) & -1(2)+0(3) & -1(3)+0(4) \\ 0(1)+1(2) & 0(2)+1(3) & 0(3)+1(4) \\ -1(1)+1(2) & -1(2)+1(3) & -1(3)+1(4) \end{bmatrix} = \begin{bmatrix} -1 & -2 & -3 \\ 2 & 3 & 4 \\ 1 & 1 & 1 \end{bmatrix}$$

Remark 1.2.3. Example 1.2.2 motivates the following definition of matrix multiplication. Consider a $1 \times n$ row vector $v = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \end{bmatrix}$ and the following $n \times 1$ column vector.

$$w = \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{n1} \end{bmatrix}$$

We define the **dot product** $v \cdot w$ of the vectors v and w as the 1×1 matrix vw^t, i.e.,

$$v \cdot w = vw^t = \begin{bmatrix} v_{11}w_{11} + v_{12}w_{21} + \cdots + v_{1n}w_{n1} \end{bmatrix}.$$

Given any $m \times n$ matrix A and any $n \times r$ matrix B, the ith row of A may be viewed as the $1 \times n$ vector $A_i = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}$ and the jth column of B as the following $n \times 1$ vector.

$$B_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}$$

Ultimately, under this interpretation, the matrix product AB is defined as the $m \times r$ matrix whose (i, j)th component is the dot product $A_i \cdot B_j = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^n a_{ik}b_{kj}$.

We adapt the following example from the example at the bottom of page 50 of [Lan86].

Example 1.2.4. We say that an $n \times n$ matrix A is a **Markov matrix** if each component of A is a non-negative real number and the sum of each column of A is 1. For instance, the 2×2 matrix

$$A = \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix}$$

is a Markov matrix. We may view this Markov matrix as representing a real-life scenario as follows.

Godspeed You! Black Emperor are playing at the Blue Note in Columbia, Missouri, and Alice and Bob are considering attending the concert. Currently, Alice is 90% certain that she will attend, so she is 10% certain that she will not attend. On the other hand, Bob is only 50% sure he will attend. Consequently, the columns of the matrix A represent Alice and Bob, respectively, and the rows represent their certainty or uncertainty that they will attend the concert, respectively.

Even more, suppose that today, Alice has the propensity a to attend the concert and Bob has the propensity b to attend, and tomorrow, Alice has the propensity $0.9a + 0.5b$ to attend the concert and Bob has the propensity $0.1a + 0.5b$ to attend. Under these identifications, tomorrow, the propensity that Alice and Bob will attend the concert is given by the following matrix product.

$$\begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0.9a + 0.5b \\ 0.1a + 0.5b \end{bmatrix}$$

We could continue to iterate this process to predict the propensity that Alice and Bob will attend the concert on any given day in the future; this is called a **Markov process**.

We will demonstrate now that matrix multiplication is associative and distributive.

Proposition 1.2.5. *If A is any $m \times n$ matrix, B is any $n \times r$ matrix, and C is any $r \times s$ matrix, then the matrix products $A(BC)$ and $(AB)C$ are well-defined; in fact, they are equal.*

Proof. By Definition 1.2.1, we have that BC is an $n \times s$ matrix, hence the matrix product $A(BC)$ is well-defined because the number of columns of A is equal to the number of rows of BC; a similar argument shows that $(AB)C$ is well-defined, hence it suffices to prove that $A(BC) = (AB)C$. By the same definition, the (i, j)th entry of $A(BC)$ is the sum of the products of the (i, k)th entry of A and the (k, j)th entry of BC for all integers $1 \leq k \leq n$, and the (k, j)th entry of BC is the sum of the products of the (k, ℓ)th entry of B and the (ℓ, j)th entry of C for all integers $1 \leq \ell \leq r$. Put into symbols, the previous sentence can be expressed as the double summation identity

$$A(BC)_{ij} = \sum_{k=1}^{n} \sum_{\ell=1}^{r} a_{ik} b_{k\ell} c_{\ell j}.$$

Considering that the order of summation of a finite sum does not matter, it follows that

$$A(BC)_{ij} = \sum_{\ell=1}^{r} \sum_{k=1}^{n} a_{ik} b_{k\ell} c_{\ell j}.$$

Observe that $\sum_{k=1}^{n} a_{ik} b_{k\ell}$ is nothing more than the (i, ℓ)th entry of AB, hence we may view the (i, j)th entry of $A(BC)$ as the sum of the products of the (i, ℓ)th entry of AB and the (ℓ, j)th entry of C for all integers $1 \leq i \leq r$, i.e., it is the (i, j)th entry of $(AB)C$. Ultimately, this shows that the (i, j)th entry of $A(BC)$ and the (i, j)th entry of $(AB)C$ are the same so that $A(BC) = (AB)C$. $\square$

Proposition 1.2.6. *If A is any $m \times n$ matrix and B and C are any $n \times r$ matrices, then the product $A(B + C)$ is well-defined; $A(B + C) = AB + AC$; and $A(cB) = c(AB)$ for all scalars c.*

Proof. By Definition 1.1.10, the matrix sum $B + C$ is an $n \times r$ matrix, hence the product $A(B + C)$ is well-defined because the number of columns of A is equal to the number of rows of $B + C$. By Definition 1.2.1, the (i, j)th entry of $A(B + C)$ is the sum of the products of the (i, k)th entry of A and the $(k, j)$the entry of BC for all integers $1 \leq k \leq n$; the latter is by Definition 1.1.10 the sum of the (k, j)th entry of B and the (k, j)th entry of C. Because multiplication is distributive over addition, the (i, j)th entry of $A(B + C)$ is the sum of the products of the (i, k)th entry of A and the (k, j)th entry of B for all integers $1 \leq k \leq n$ plus the sum of the products of the (i, k)th entry of A and the (k, j)th entry of C for all integers $1 \leq k \leq n$, i.e., it is the sum of the (i, j)th entry of AB and the (i, j)th entry of AC, i.e., it is the (i, j)th entry of $AB + AC$. Because the (i, j)th entry of $A(B + C)$ and the (i, j)th entry of $AB + AC$ are the same, we conclude that $A(B + C) = AB + AC$.

We leave it as an exercise for the reader to demonstrate that $A(cB) = c(AB)$ for all scalars c; however, we remark that inspiration can be found in the proof of Proposition 1.2.5. $\square$

Ultimately, Proposition 1.2.6 implies that matrix multiplication is distributive, i.e., if A is any $m \times n$ matrix, B and C are any $n \times r$ matrices, and c is any scalar, then $A(cB + C) = c(AB) + AC$.

Example 1.2.7. If A is any $n \times n$ matrix, then the matrix product of A with itself is denoted simply by A^2; it is an $n \times n$ matrix, hence we may form the matrix product of A^2 with A. By Proposition

1.2.5, we have that $(A^2)A = (AA)A = A(AA) = A(A^2)$; we denote this simply by A^3. Continuing in this manner, the k-fold product of A is $A^k = A^{k-1}A = AA^{k-1}$ for all integers $k \geq 2$. Each of these is an $n \times n$ matrix, so we can scale these matrices and add them together to obtain a **matrix polynomial**. By the distributive property for matrices, matrix polynomials behave familiarly, e.g.,

$$(A - I)(A + I) = A^2 + AI - IA - I^2 = A^2 + A - A - I = A^2 - I \text{ and}$$
$$(A + I)^3 = (A^2 + 2A + I)(A + I) = A^3 + A^2 + 2A^2 + 2A + A + I = A^3 + 3A^2 + 3A + I.$$

Even more, like matrix addition, matrix multiplication is compatible with transposition.

Proposition 1.2.8. *If A is any $m \times n$ matrix and B is any $n \times r$ matrix, then $(AB)^t = B^tA^t$. Put another way, the transpose of a matrix product is the reverse matrix product of the transposes.*

Proof. By Example 1.1.8, the (i, j)th entry of $(AB)^t$ is the (j, i)th AB. By Definition 1.2.1, the (j, i)th entry of AB is the sum of the products of the (j, k)th entry of A and the (k, i)th entry of B for all integers $1 \leq k \leq n$. Considering that scalar multiplication is commutative, this is equal to the sum of the products of the (i, k)th entry of B^t and the (k, j)th entry of A^t for all integers $1 \leq k \leq n$, i.e., it is the (i, j)th entry of B^tA^t. We conclude therefore that $(AB)^t = B^tA^t$. □

We return now to the setup of the first paragraph of this section. Once again, we are considering some point (x, y) in the Cartesian plane, and we are identifying this point by its polar coordinates $x = r\cos\theta$ and $y = r\sin\theta$ for some real number r and some angle θ. Our aim is to efficiently write down the rotation of (x, y) through another angle ϕ, resulting in a new point determined by $x' = r\cos(\theta + \phi)$ and $y' = r\sin(\theta + \phi)$. By the addition formulas for sine and cosine, it follows that $x' = r(\cos\theta\cos\phi - \sin\theta\sin\phi)$ and $y' = r(\sin\theta\cos\phi + \sin\phi\cos\theta)$. Consider the following matrices.

$$R(\phi) = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \text{ and } X(r, \theta) = \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix}$$

Observe that $X(r, \theta)$ is the column vector corresponding to the point (x, y) in the Cartesian plane, i.e., it encodes the same data as the point (x, y). By Definition 1.2.1, we have the following.

$$R(\phi)X(r, \theta) = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} r\cos\theta \\ r\sin\theta \end{bmatrix} = \begin{bmatrix} r(\cos\theta\cos\phi - \sin\theta\sin\phi) \\ r(\sin\phi\cos\theta + \sin\theta\cos\phi) \end{bmatrix} = \begin{bmatrix} r\cos(\theta + \phi) \\ r\sin(\theta + \phi) \end{bmatrix}$$

Considering that the last matrix in the above displayed equation is exactly equal to the column vector $X(r, \theta + \phi)$, i.e., the column vector corresponding to the point (x', y'), we conclude that the multiplication by the matrix $R(\phi)$ has the effect of rotating the point (x, y) in the Cartesian plane through the angle ϕ. Consequently, we refer to the matrix $R(\phi)$ as a **rotation matrix**.

Example 1.2.9. Consider the point $(1, 0)$ in the Cartesian plane. Observe that in polar coordinates, this point is determined by $r\cos\theta = 1$ and $r\sin\theta = 0$, hence we obtain the following column vector.

$$X(r, \theta) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

By the previous paragraph, to rotate $X(r, \theta)$ through the angle $\phi = \pi/4$, multiply by the following.

$$R(\pi/4) = \begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

Consequently, we find that rotating the point $(1, 0)$ through the angle $\phi = \pi/4$ results in the point

$$X(r, \theta + \phi) = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{bmatrix}.$$

But if we consider the fact that the point $(1, 0)$ lies on the unit circle and corresponds to the angle $\theta = 0$, then the point obtained by rotating $(1, 0)$ through the angle of $\phi = \pi/4$ must be exactly the point on the unit circle corresponding to the angle $\pi/4$, i.e., it must be $(\sqrt{2}/2, \sqrt{2}/2)$.

1.3 Elementary Row and Column Operations

We will continue to assume that m and n are positive integers. If $x_1, \dots, x_n$ are any variables, then a (real) **linear combination** of $x_1, \dots, x_n$ is an expression of the form $a_1 x_1 + \cdots + a_n x_n$ for some (real) scalars $a_1, \dots, a_n$. Consequently, a (real) $1 \times n$ **linear equation** is any equation of the form $a_1 x_1 + \cdots + a_n x_n = b$ for some (real) scalars $a_1, \dots, a_n$, and b. Even more, a (real) $m \times n$ **system of linear equations** consists of m linear equations in n variables; this is represented as follows.

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

Explicitly, the positive integer m represents the number of equations in the $m \times n$ system of linear equations, and the positive integer n represents the number of variables in each equation.

Example 1.3.1. On 10 June 2022, in Game Four of the 2022 NBA Finals, Steph Curry scored 43 points. Let x_1 be the number of one-pointers made; let x_2 be the number of two-pointers made; and let x_3 be the number of three-pointers made by Curry in this appearance. Observe that Curry's point total is given by the 1×3 (integer) linear equation $x_1 + 2x_2 + 3x_3 = 43$.

We say that the (real) scalars $\xi_1, \dots, \xi_n$ constitute a **solution** to a (real) $m \times n$ system of linear equations if it holds that $a_{i1}\xi_1 + \cdots + a_{in}\xi_n = b_i$ for each integer $1 \leq i \leq m$.

Example 1.3.2. One can find many solutions to the matrix equation of Example 1.3.1. Explicitly, $\xi_1 = 43$ and $\xi_2 = \xi_3 = 0$ or $\xi_1 = 41$, $\xi_2 = 1$, and $\xi_3 = 0$ give rise to two distinct solutions.

Given more information about the game, we can reduce the number of possible solutions. For instance, Curry made seven three-pointers, hence we may substitute $x_3 = 7$ into our equation $x_1 + 2x_2 + 3x_3 = 43$ to find that $x_1 + 2x_2 + 21 = 43$ or $x_1 + 2x_2 = 22$. Even more, Curry made a combined fifteen free throws and two-pointers. Consequently, we have that $x_1 + x_2 = 15$. Observe that these two equations involving x_1 and x_2 induce the following 2×2 system of linear equations.

$$\begin{aligned} x_1 + 2x_2 &= 22 \\ x_1 + x_2 &= 15 \end{aligned}$$

Using this information, we may uniquely determine x_1 and x_2: we have that $x_1 = 15 - x_2$ so that $22 = x_1 + 2x_2 = (15 - x_2) + 2x_2 = 15 + x_2$; cancelling 15 from both sides gives $x_2 = 7$ and $x_1 = 8$.

Using matrices, we can more efficiently rephrase our above observations concerning $m \times n$ systems of linear equations. Explicitly, observe that a (real) $m \times n$ system of linear equations

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

gives rise to a $1 \times n$ matrix $x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$, a $1 \times m$ matrix $b = \begin{bmatrix} b_1 & b_2 & \cdots & b_m \end{bmatrix}$, and an $m \times n$ matrix A whose (i, j)th entry is the coefficient a_{ij} of the jth variable x_j of the ith equation $a_{i1}x_1 + \cdots + a_{in}x_n = b_i$ of the $m \times n$ system of linear equations, i.e., the following $m \times n$ matrix.

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

Conversely, the aforementioned matrices A, x, and b satisfy that $Ax^t = b^t$. We refer to the equation $Ax^t = b^t$ as a (real) $m \times n$ **matrix equation**. Often, the $m \times n$ matrix A and the $1 \times m$ matrix b are known while the $1 \times n$ matrix x consists of n variables. Ultimately, we obtain a one-to-one correspondence between (real) $m \times n$ systems of linear equations and $m \times n$ matrix equations.

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned} \iff Ax^t = b^t, \text{ i.e., } \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Example 1.3.3. We will convert the data of Examples 1.3.1 and 1.3.2 into the language of matrix equations. Consider the matrix $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ whose jth column is the point value of a j-pointer; the matrix $x = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$ whose jth column is the number of j-pointers made by Curry; and the matrix $b = \begin{bmatrix} 43 \end{bmatrix}$ consisting of the total points made by Curry. Observe that the linear equation $x_1 + 2x_2 + 3x_3 = 43$ is in one-to-one correspondence with the matrix equation $Ax^t = b^t$.

We say that a $1 \times n$ (real) matrix ξ forms a **solution** to the matrix equation $Ax^t = b^t$ if it holds that $A\xi^t = b^t$. Observe that this is an analog of a solution of the $m \times n$ system of linear equations.

Example 1.3.4. Rephrasing the results of 1.3.2, the matrices $\xi_1 = \begin{bmatrix} 43 & 0 & 0 \end{bmatrix}$ and $\xi_2 = \begin{bmatrix} 41 & 1 & 0 \end{bmatrix}$ give rise to two distinct solutions of the matrix equation of Example 1.3.3. On the other hand, put into the language of matrix equations, the information that $22 = x_1 + 2x_2$ and $15 = x_1 + x_2$ can be most efficiently synthesized by viewing the coefficients of these linear equations as rows of a matrix. Explicitly, we construct a matrix A whose first row is $\begin{bmatrix} 1 & 2 \end{bmatrix}$, corresponding to the respective coefficients of x_1 and x_2 in the equation $22 = x_1 + 2x_2$; the second row of the matrix A is $\begin{bmatrix} 1 & 1 \end{bmatrix}$, corresponding to the respective coefficients of x_1 and x_2 in the equation $15 = x_1 + x_2$. Once again, the column vector x^t consists of the variables x_1 and x_2 in distinct rows, and the column vector b^t consists of the integers 22 and 15 in distinct rows. Ultimately, yields the matrix equation

$$Ax^t = b^t \text{ or } \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 22 \\ 15 \end{bmatrix}.$$

Once we have extracted an $m \times n$ matrix equation $Ax^t = b^t$ from a (real) $m \times n$ system of linear equations, our next objective is to determine the matrix analog of solving the system. Before we do this, recall the following three valid operations for working with systems of linear equations.

(1.) We may multiply the ith equation by a nonzero (real) scalar c.

(2.) We may add c times the ith equation to the jth equation for all integers $1 \leq i, j \leq m$.

(3.) We may interchange the ith and jth equations for all integers $1 \leq i, j \leq m$.

Consequently, we are looking for matrix analogs of the above three operations. Considering that the coefficients of ith equation are encoded in the ith row of the matrix A and the ith row of the matrix b^t, we must henceforth work with the **augmented matrix** $\left[A \mid b^t\right]$. By definition, this is simply the matrix A with one additional column in the form of b^t. We use the bar $\mid$ notation to emphasize that b^t is appended to the matrix A, i.e., it is not originally a column of A. By definition of matrix multiplication, operation (1.) is analogous to left multiplication by the $m \times m$ matrix with c in row i, column i; 1 in all other entries of the main diagonal; and 0s elsewhere.

(1.) Multiplication of the ith row of an $m \times n$ system of linear equations by a scalar c corresponds to left multiplication of the $m \times (n+1)$ augmented matrix $\left[A \mid b^t\right]$ by the $m \times m$ matrix with c in row i, column i; 1 in all other entries of the main diagonal; and 0s elsewhere.

Example 1.3.5. We obtain the following augmented matrix for the matrices of Example 1.3.4.

$$\left[A \mid b^t\right] = \left[\begin{array}{cc|c} 1 & 2 & 22 \\ 1 & 1 & 15 \end{array}\right]$$

Consequently, to scale the first equation $x_1 + 2x_2 = 22$ by a factor of c, we multiply this augmented matrix by the 2×2 matrix with c in row 1, column 1; 1 in row 2, column 2; and 0s elsewhere.

$$\left[\begin{array}{cc|c} c & 2c & 22c \\ 1 & 1 & 15 \end{array}\right] = \begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix} \left[\begin{array}{cc|c} 1 & 2 & 22 \\ 1 & 1 & 15 \end{array}\right]$$

Likewise, operation (2.) is analogous to left multiplication by the $m \times m$ matrix with c in row j, column i; 1s along the main diagonal; and 0s elsewhere. Explicitly, we obtain the following rule.

(2.) Addition of c times the ith row of an $m \times n$ system of linear equations to the jth row corresponds to left multiplication of the $m \times (n+1)$ matrix $\left[A \mid b^t\right]$ by the $m \times m$ matrix with c in row j, column i; 1s along the main diagonal; and 0s elsewhere.

Example 1.3.6. Consider the augmented matrix $\left[A \mid b^t\right]$ of Example 1.3.5. Observe that if we wish to subtract the first equation $x_1 + 2x_2 = 22$ from the second equation $x_1 + x_2 = 15$, then it suffices to add -1 times the first equation to the second equation. By the previous observation, this can be achieved on the level of matrices by performing the following matrix multiplication.

$$\left[\begin{array}{cc|c} 1 & 2 & 22 \\ 0 & -1 & -7 \end{array}\right] = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \left[\begin{array}{cc|c} 1 & 2 & 22 \\ 1 & 1 & 15 \end{array}\right]$$

Last, operation (3.) is analogous to left multiplication by the $m \times m$ matrix with (i, j)th and (j, i)th entries of 1; 1s along the main diagonal other than in rows i and j; and 0s elsewhere.

(3.) Interchanging rows i and j of an $m \times n$ system of linear equations corresponds to left multiplication of the $m \times (n+1)$ matrix $\left[A \mid b^t\right]$ by the $m \times m$ matrix with 1 in row j, column i; 1 in row i, column j; 1s along the main diagonal other than rows i and j; and 0s elsewhere.

Example 1.3.7. Once again, consider the augmented matrix $\left[A \mid b^t\right]$ of Example 1.3.5. We may interchange the first equation $x_1 + 2x_2 = 22$ and the second equation $x_1 + x_2 = 15$ as follows.

$$\left[\begin{array}{cc|c} 1 & 1 & 15 \\ 1 & 2 & 22 \end{array}\right] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \left[\begin{array}{cc|c} 1 & 2 & 22 \\ 1 & 1 & 15 \end{array}\right]$$

Collectively, we refer to the operations (1.), (2.), and (3.) defined above as the **elementary row operations**; the matrices defined in operations (1.), (2.), and (3.) are then called the $m \times m$ **elementary row matrices**. Explicitly, an elementary row matrix is an $m \times m$ matrix obtain by from the $m \times m$ identity matrix I_m by (1.) multiplying any row of I_m by a nonzero scalar c; (2.) adding c times the ith row of I_m to the jth row of I_m; or (3.) interchanging rows i and j of I_m.

Likewise, the three above operations can be defined for the columns of a matrix to obtain the **elementary column operations** and the **elementary column matrices**: we need only swap all instances of "rows" with "columns" and "left multiplication" with "right multiplication."

1.4 The Method of Gaussian Elimination in Linear Systems

We will soon see that performing elementary row and column operations on a system of linear equations does not affect the solutions to the system, hence it does not alter the solutions of the underlying matrix equation. Even more, if we employ a sequence of elementary row and column operations to reduce a given augmented matrix to a "relatively simple" form and subsequently interpret the resulting augmented matrix "correctly," then we can easily read off all possible solutions to the underlying system of linear equations. We illustrate this in the case of Example 1.3.6.

Example 1.4.1. Consider the augmented matrix $\left[A \mid b^t\right]$ of Example 1.3.6. Converting this back into a system of equations, the second row of the augmented matrix yields that $-x_2 = -7$, hence we conclude that $x_2 = 7$. Consequently, the first row gives that $22 = x_1 + 2x_2 = x_1 + 14$ or $x_1 = 8$. We refer to this as the method of solving a system of linear equations via **back substitution**.

Going forward, we will say that two matrices A and B are **row equivalent** if A can be reduced to B via a sequence of elementary row operations, i.e., there exist elementary row matrices $E_1, \dots, E_k$ such that $B = E_k \cdots E_1 A$. Likewise, we make the analogous definition for **column equivalent** matrices. If A and B are either row or column equivalent, then we will write $A \sim B$.

Example 1.4.2. By Example 1.3.6 of the previous section, we have that

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

are row equivalent because $B = EA$ for the elementary row matrix $E = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$.

By Example 1.4.1, it is clearly advantageous (when possible) to perform a sequence of elementary row operations to reduce a matrix A to a matrix B in which some row has the property that all but one of its entries is nonzero. If this holds, then the row of B consisting of just one nonzero entry can be used to further reduce A to a matrix possessing more zero entries, as we illustrate next.

Example 1.4.3. Consider the row equivalent matrices A and B of Example 1.4.2. Observe that if we add twice the second row of B to the first row of B, then we obtain the matrix

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

Certainly, matrices with more zero entries are easier to interpret as the collection of coefficients corresponding to some system of linear equations because the variables corresponding to the zeros of the ith row of the matrix do not appear in the ith equation of the system. Even more, the zeros of a matrix inform us about other important properties of the matrix that we will soon discuss. Consequently, we turn our attention in this section to an algorithm that we may employ to reduce a given matrix A to a row equivalent matrix consisting of as many zeros as possible.

We say that a row of an $m \times n$ matrix A is **nonzero** if it contains (at least) one nonzero entry. Using this identification, an $m \times n$ matrix A lies in **row echelon form** if and only if

(1.) all rows of A consisting entirely of zeros lie beneath the last nonzero row of A; and

(2.) for any pair of consecutive nonzero rows i and $i+1$, the first nonzero entry of row $i+1$ lies in some column strictly to the right of the column in which the first nonzero entry of row i lies.

Given a matrix A that lies in row echelon form, we distinguish the first nonzero entry of a nonzero row of A as a **pivot**. We have already encountered instances of matrices in row echelon form: the matrices B of Example 1.4.2 and C of Example 1.4.3 lie in row echelon form; however, the matrix A of Example 1.4.2 does not lie in row echelon form because the first nonzero entry of the second row of A lies directly below the first nonzero entry of the first row of A. Even more, the pivots of the aforementioned matrix B (and C) are 1 in the first row and -1 in the second row. Crucially, the following theorem assures us that it is always possible to reduce any matrix to row echelon form.

Theorem 1.4.4. *Every real matrix is row equivalent to a real matrix in row echelon form.*

Proof. Consider a real $m \times n$ matrix A. Begin by relocating all rows of A consisting entirely of zeros to the bottom of the matrix; interchanging rows corresponds to multiplying on the left by an elementary row matrix, hence the resulting matrix is row equivalent to A. We may disregard all columns of A consisting entirely of zeros because the columns of A do not bear on the row echelon form of A, hence we may assume that the first column of A is nonzero; then, find the first nonzero row of A for which the entry in first column of A is nonzero. By interchanging this row with the first row of A, we may ultimately assume that our $m \times n$ matrix A has the form

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

in which the lowermost rows could consist of zeros and a_{11} is nonzero by assumption. Every nonzero real number has a multiplicative inverse, hence we may subtract $a_{i1}a_{11}^{-1}$ times the first row from the ith row; this corresponds to left multiplication by an elementary row matrix and yields that

$$A \sim \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

for some real numbers $b_{22}, \dots, b_{mn}$. Employing this process with the $(m-1) \times (n-1)$ submatrix

$$B = \begin{bmatrix} b_{22} & \cdots & b_{2n} \\ \vdots & & \vdots \\ b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

and subsequently continuing in this manner, we will eventually reduce A to row echelon form. □

We say moreover that a matrix lies in **reduced row echelon form** if and only if

(1.) it lies in row echelon form;

(2.) its pivots are all 1; and

(3.) if the jth column contains a pivot, then all of its non-pivot entries are zero. Put another way, the only nonzero entry of any column containing a pivot is the pivot itself.

Corollary 1.4.5. *Every real matrix is row equivalent to a real matrix in reduced row echelon form.*

Proof. By Theorem 1.4.4, every real matrix A is row equivalent to a real matrix B in row echelon form. By multiplying each nonzero row of B by the multiplicative inverse of its pivot, we obtain a row equivalent matrix C whose pivots are all 1. Last, we must ensure that the only nonzero entry of any column containing a pivot is the pivot itself. Observe that if c_{ij} is nonzero and the jth column of C contains a pivot in row k, then we may add $-c_{ij}$ times the kth row of C to the ith row of C to obtain 0 in the ith row and jth column of C. Continuing in this manner yields the result. □

Essentially, the proofs of Theorem 1.4.4 and Corollary 1.4.5 outline the method of **Gaussian Elimination** in systems of linear equations; for completeness, we summarize the results below.

Algorithm 1.4.6 (Gaussian Elimination)**.** Let A be a nonzero real $m \times n$ matrix. Use the following steps to reduce the matrix A to a row equivalent matrix B that lies in reduced row echelon form.

(1.) Begin by relocating all rows of A consisting entirely of zeros to the bottom of the matrix. We may perform this operation because row interchange yields a row equivalent matrix.

(2.) Find the first nonzero row i of the matrix obtained in the previous step for which the entry a_{i1} in first column is nonzero; if this is not the first row, then interchange the first and ith rows of this matrix so that a_{i1} lies in the first row and column of the resulting matrix.

(3.) Multiply the first row of the resulting matrix by the multiplicative inverse a_{i1}^{-1} of the nonzero real number a_{i1} to obtain an entry of 1 in the first row and first column. We may perform this operation because multiplying a row by a nonzero scalar yields a row equivalent matrix.

(4.) If r_j is the component of the jth row and first column of the matrix obtained in step (3.), then add $-r_j$ times the first row of this matrix to the jth row of this matrix for each integer $1 \leq j \leq m$. We may perform this operation because adding a scalar multiple of a row to another row yields a row equivalent matrix. Observe that the only nonzero entry in the first column of the resulting matrix is the pivot of 1 in the first row and first column.

(5.) Repeat steps (2.), (3.), (4.) for the matrix obtained from the resulting matrix of step (4.) by ignoring the first row and first column; if possible, a pivot of 1 is obtained in the second row of this matrix, and all entries of the matrix below this pivot are zero.

(6.) Repeat step (5.) until the row echelon form of A is obtained and all pivots are 1.

(7.) Eliminate any nonzero entry a_{ij} in row i above the pivot 1 in row k by adding $-a_{ij}$ times the kth row of the matrix of step (6.) to the ith row of the matrix.

(8.) Repeat step (7.) until the matrix lies in reduced row echelon form.

We refer to the matrix obtained from this process as the **reduced row echelon form** $\operatorname{RREF}(A)$.

One of the best ways to understand the method of Gaussian Elimination is to practice using it.

Example 1.4.7. Let us convert the following matrix to reduced row echelon form.

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

Considering that each of the rows of A is nonzero, we may immediately proceed to the second step of the Gaussian Elimination algorithm. Observe that the first nonzero row of A for which the entry in the first column is nonzero is simply the first row of A, so we may proceed to the third step of the algorithm. Explicitly, we multiply the first row of A by $\frac{1}{2}$ (i.e., the multiplicative inverse of 2) to obtain an entry of 1 in the first row and first column of A. We illustrate this as follows.

$$A = \begin{bmatrix} 2 & -3 & 7 \\ -1 & 0 & 3 \\ 2 & 1 & 5 \end{bmatrix} \overset{\frac{1}{2}R_1 \mapsto R_1}{\sim} \begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ -1 & 0 & 3 \\ 2 & 1 & 5 \end{bmatrix}$$

We may subsequently reduce all first column entries beneath the first row of the resulting matrix.

$$\begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ -1 & 0 & 3 \\ 2 & 1 & 5 \end{bmatrix} \overset{R_2 + R_1 \mapsto R_2}{\sim} \begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & -\frac{3}{2} & \frac{13}{2} \\ 2 & 1 & 5 \end{bmatrix} \overset{R_3 - 2R_1 \mapsto R_3}{\sim} \begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & -\frac{3}{2} & \frac{13}{2} \\ 0 & 4 & \frac{3}{2} \end{bmatrix}$$

We have therefore created a pivot of 1 in the first row and first column, so we proceed to do the same for the second row and second column. Explicitly, we multiply the second row of the above matrix by $-\frac{2}{3}$ (i.e., the multiplicative inverse of $-\frac{3}{2}$) to obtain the following row equivalent matrix.

$$\begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & -\frac{3}{2} & \frac{13}{2} \\ 0 & 4 & \frac{3}{2} \end{bmatrix} \overset{-\frac{2}{3}R_2 \mapsto R_2}{\sim} \begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & 1 & -\frac{13}{3} \\ 0 & 4 & \frac{3}{2} \end{bmatrix}$$

We may then create a pivot of 1 in the second row and second column of this matrix by adding -4 times the second row to the third row, reducing the entry in the third row and second column to 0.

$$\begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & 1 & -\frac{13}{3} \\ 0 & 4 & \frac{3}{2} \end{bmatrix} \overset{R_3 - 4R_2 \mapsto R_3}{\sim} \begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & 1 & -\frac{13}{3} \\ 0 & 0 & \frac{95}{6} \end{bmatrix}$$

Last, we obtain a pivot of 1 in the third row and third column by multiplying by the multiplicative inverse $\frac{6}{95}$ of $\frac{95}{6}$. Ultimately, we obtain the row echelon form of A for which all pivots are 1.

$$\begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & 1 & -\frac{13}{3} \\ 0 & 0 & \frac{95}{6} \end{bmatrix} \overset{\frac{6}{95}R_3 \mapsto R_3}{\sim} \begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & 1 & -\frac{13}{3} \\ 0 & 0 & 1 \end{bmatrix}$$

We proceed to the seventh and eighth steps of the Gaussian Elimination algorithm. Because there is a pivot in the second row, we eliminate first the nonzero non-pivot entries in the second column.

$$\begin{bmatrix} 1 & -\frac{3}{2} & \frac{7}{2} \\ 0 & 1 & -\frac{13}{3} \\ 0 & 0 & 1 \end{bmatrix} \overset{R_1 + \frac{3}{2}R_2 \mapsto R_1}{\sim} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -\frac{13}{3} \\ 0 & 0 & 1 \end{bmatrix}$$

Once this is accomplished, we put the matrix in reduced row echelon form as follows.

$$\begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -\frac{13}{3} \\ 0 & 0 & 1 \end{bmatrix} \overset{R_1 + 3R_3 \mapsto R_1}{\sim} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{13}{3} \\ 0 & 0 & 1 \end{bmatrix} \overset{R_2 + \frac{13}{3}R_3 \mapsto R_2}{\sim} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ultimately, the method of Gaussian Elimination illustrates that our original matrix A is in fact row equivalent to the 3×3 identity matrix. We will see in the next section that row equivalence to the $n \times n$ identity matrix is a very important and special property of a square matrix.

1.5 Invertible Matrices

We will assume throughout this section that n is a positive integer. Given any $n \times n$ matrix A, we say that an $n \times n$ matrix L is a **left inverse** of A if it holds that $LA = I_{n \times n}$, where $I_{n \times n}$ is the $n \times n$ identity matrix. Likewise, we say that an $n \times n$ matrix R is a **right inverse** of A if it holds that $AR = I_{n \times n}$. We will establish immediately that every left inverse of A is also a right inverse and vice-versa, hence we may dispense of the distinct notions of left and right inverses of matrices and simply say that an $n \times n$ matrix B is a (two-sided) **inverse** of an $n \times n$ matrix A if it holds that $AB = I_{n \times n} = BA$. Our next proposition shows that a two-sided inverse of a matrix A is unique.

Proposition 1.5.1. *Let A be an $n \times n$ matrix. Every left inverse of A is a right inverse of A and vice-versa (provided they both exist). Even more, if A admits a two-sided inverse, it is unique.*

Proof. Consider any $n \times n$ matrices L and R such that $LA = I_{n\times n} = AR$. By Proposition 1.2.5, we have that $L = LI_{n\times n} = L(AR) = (LA)R = I_{n\times n}R = R$. Consequently, L is a two-sided inverse of A. Even more, if L' is any two-sided inverse of A, then it is a right inverse of A so that $L' = L$. □

Consequently, if an $n \times n$ matrix A admits a (two-sided) inverse, then it is unique, and we may denote it by A^{-1}. We will also say in this case that A is **invertible** (or **non-singular**). Certainly, the zero matrix does not possess an inverse, hence some (and in fact many) matrices are not invertible. We demonstrate next how matrix inverses behave in relation to other matrix operations.

Proposition 1.5.2. *Let A be any $n \times n$ matrix. If A^{-1} exists, then $(A^t)^{-1} = (A^{-1})^t$. Put another way, if A is invertible, then A^t is invertible, and its inverse is the transpose of A^{-1}.*

Proof. By Proposition 1.2.8, it follows that $(A^{-1})^t A^t = (AA^{-1})^t = I^t_{n\times n} = I_{n\times n}$, and we conclude that $(A^t)^{-1} = (A^{-1})^t$ by the uniqueness of the matrix inverse guaranteed by Proposition 1.5.1. □

Proposition 1.5.3. *Let $A_1, \dots, A_k$ be any invertible $n \times n$ matrices. We have that*

$$(A_1 \cdots A_k)^{-1} = A_k^{-1} \cdots A_1^{-1}.$$

Put another way, the inverse of a product of invertible matrices is the reverse product of the inverses.

Proof. By Proposition 1.5.1, it suffices to verify that $(A_k^{-1} \cdots A_1^{-1})(A_1 \cdots A_k) = I_{n\times n}$. Considering that $A_i^{-1}A_i = I_{n\times n}$ for all integers $1 \leq i \leq k$, we may replace every instance of $A_i^{-1}A_i$ with $I_{n\times n}$; then, using the fact that $I_{n\times n}B = B$ for any $n \times r$ matrix B, the result eventually follows. □

Using the method of Gaussian Elimination, we can determine if an $n \times n$ matrix A admits an inverse, and we may subsequently compute A^{-1} in this way, as well. Before we demonstrate this, we remind the reader that two matrices are row equivalent if and only if there exist some elementary row matrices whose product (on the left) of one matrix gives the other. Elementary row matrices are the $n \times n$ matrices obtained from the $n \times n$ identity matrix by performing one of the following.

(1.) We may multiply any row of $I_{n\times n}$ by a nonzero scalar c.

(2.) We may add c times the ith row of $I_{n\times n}$ to the jth row of $I_{n\times n}$.

(3.) We may interchange any pair of rows i and j of $I_{n\times n}$.

We refer to the above operations as the elementary row operations.

Proposition 1.5.4. *Every elementary row matrix is invertible.*

Proof. Let E be an $n \times n$ elementary row matrix. Consider the following three cases.

(1.) If E is obtained from $I_{n\times n}$ by multiplying the ith row of $I_{n\times n}$ by a nonzero scalar c, then E^{-1} is obtained from $I_{n\times n}$ by multiplying the ith row of $I_{n\times n}$ by the nonzero scalar c^{-1}.

(2.) If E is obtained from $I_{n \times n}$ by adding c times the ith row of $I_{n \times n}$ to the jth row of $I_{n \times n}$, then E^{-1} is obtained from $I_{n \times n}$ by adding $-c$ times the ith row of $I_{n \times n}$ to the jth row of $I_{n \times n}$.

(3.) If E is obtained from $I_{n \times n}$ by interchanging rows i and j of $I_{n \times n}$, then E is its own inverse. □

Corollary 1.5.5. *If A and B are row equivalent, then A is invertible if and only if B is invertible.*

Proof. By definition, the $n \times n$ matrix A is row equivalent to the $n \times n$ matrix B if and only if there exist $n \times n$ elementary row matrices $E_1, \dots, E_k$ such that $B = E_k \cdots E_1 A$. Observe that if B is invertible, then A is invertible because $(B^{-1} E_k \cdots E_1)A = I_{n \times n}$. Conversely, if A is invertible, then B is invertible by Propositions 1.5.3 and 1.5.4: $I_{n \times n} = B(E_k \cdots E_1 A)^{-1} = BA^{-1} E_1^{-1} \cdots E_k^{-1}$. □

By Corollary 1.4.5, every $n \times n$ matrix A is row equivalent to its reduced row echelon form $\operatorname{RREF}(A)$. Consequently, by the previous corollary, it follows that A is invertible if and only if $\operatorname{RREF}(A)$ is invertible. Particularly, if $\operatorname{RREF}(A)$ admits any rows consisting entirely of zeros, then it is not invertible (because the last row of $\operatorname{RREF}(A)B$ is zero for all $n \times r$ matrices B), hence A cannot be invertible. Conversely, we will demonstrate that if all rows $\operatorname{RREF}(A)$ are nonzero, then it is invertible, hence A is invertible. Before this, we mention that an **upper-triangular matrix** is an $n \times n$ matrix with the property that if $i < j$, then the (i, j)th component of the matrix is zero. Put another way, all entries below the main diagonal of an upper-triangular matrix are zero.

Theorem 1.5.6. *Every upper-triangular matrix with nonzero diagonal elements is invertible.*

Proof. By definition, every $n \times n$ upper-triangular matrix U can be written as follows.

$$U = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

By hypothesis that a_{ii} is nonzero for each integer $1 \leq i \leq n$, we may multiply the ith row of the above matrix by a_{ii}^{-1} to obtain an upper-triangular matrix whose pivots are all 1. Consequently, we assume from the beginning that this is the case, i.e., we may consider the following case.

$$U = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 0 & 1 & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

By Corollary 1.5.5, it suffices to demonstrate that U is row equivalent to the invertible $n \times n$ identity matrix $I_{n \times n}$. We achieve this by furnishing some elementary row operations that reduces U to $I_{n \times n}$. Observe that if we add $-a_{in}$ times the last row of U to the ith row of U, then we obtain a 0 in the (i, n)th component of the resulting matrix. Continuing in this way, we may reduce the nth column of U to zero except in the bottom right-hand corner. Considering that adding any scalar multiple of a row of U to another row of U is a row equivalence, we conclude that U is row equivalent to this matrix. Continuing in this way for each column of U from right to left, we obtain $I_{n \times n}$. □

Corollary 1.5.7. *An $n \times n$ matrix is invertible if and only if it is row equivalent to the $n \times n$ identity matrix. Even more, we may obtain the unique inverse matrix by performing Gaussian Elimination.*

Proof. By Theorem 1.5.6 and the paragraph that precedes it, an $n \times n$ matrix A is invertible if and only if the upper-triangular matrix $\operatorname{RREF}(A)$ is invertible if and only if $\operatorname{RREF}(A) = I_{n \times n}$. Consequently, there exist some elementary row operations $E_1, \dots, E_k$ such that $E_k \cdots E_1 A = I_{n \times n}$, from which we conclude that the unique inverse of A is given by $A^{-1} = E_k \cdots E_1$. $\square$

Corollary 1.5.8. *Every invertible $n \times n$ matrix is a product of elementary row matrices.*

Proof. By the proof of Corollary 1.5.7, every invertible $n \times n$ matrix A admits some elementary row matrices $E_1, \dots, E_k$ such that $E_k \cdots E_1 A = I_{n \times n}$. By multiplying both sides on the left by $E_1^{-1} \cdots E_k^{-1}$, we obtain that $A = E_1^{-1} \cdots E_k^{-1}$. By the proof of Proposition 1.5.4, each of the matrices $E_1^{-1}, \dots, E_k^{-1}$ is an elementary row matrix, hence A is the product of elementary row matrices. $\square$

Example 1.5.9. Let us illustrate the method of Gaussian Elimination to determine a numerical criterion under which an arbitrary real 2×2 matrix is invertible. Consider any 2×2 matrix

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

such that a, b, c, and d are real numbers. Observe that if $a = 0$ and $c = 0$, then A is not invertible because the first row of the matrix BA will be zero for all real $m \times 2$ matrices B. Consequently, we may assume that a is nonzero. By multiplying the first row of A by a^{-1}, we obtain the following.

$$A \overset{a^{-1}R_1 \mapsto R_1}{\sim} \begin{bmatrix} 1 & a^{-1}b \\ c & d \end{bmatrix}$$

Equivalently, the displayed matrix above is $E_1 A$ for the following elementary row matrix

$$E_1 = \begin{bmatrix} a^{-1} & 0 \\ 0 & 1 \end{bmatrix}$$

We may subsequently create a pivot in the first row and first column of $E_1 A$ by adding $-c$ times the first row of $E_1 A$ to the second row of $E_1 A$. Explicitly, we obtain the following.

$$E_1 A \overset{R_2 - cR_1 \mapsto R_2}{\sim} \begin{bmatrix} 1 & a^{-1}b \\ 0 & d - a^{-1}bc \end{bmatrix}$$

Equivalently, the displayed matrix above is $E_2 E_1 A$ for the following elementary row matrix.

$$E_2 = \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix}$$

Observe that if $d - a^{-1}bc = 0$, then the last row of $E_2 E_1 A$ is zero, hence it is not invertible so that A is not invertible. Consequently, we must have that $d - a^{-1}bc$ is nonzero, i.e., we must have that $ad - bc$ is nonzero. Continuing onward, because $d - a^{-1}bc$ is nonzero, it possesses a multiplicative inverse $(d - a^{-1}bc)^{-1}$. By multiplying the last row of $E_2 E_1 A$ by $(d - a^{-1}bc)^{-1}$, obtain the following.

$$E_2 E_1 A \overset{(d - a^{-1}bc)^{-1} R_2 \mapsto R_2}{\sim} \begin{bmatrix} 1 & a^{-1}b \\ 0 & 1 \end{bmatrix}$$

Equivalently, the displayed matrix above is $E_3E_2E_1A$ for the following elementary row matrix.

$$E_3 = \begin{bmatrix} 1 & 0 \\ 0 & (d - a^{-1}bc)^{-1} \end{bmatrix}$$

Last, by adding $-(d - a^{-1}bc)^{-1}$ times the second row of A to the first row of A, we obtain a pivot in the second row and second column. Explicitly, if we multiply $E_3E_2E_1A$ on the left by

$$E_4 = \begin{bmatrix} 1 & -a^{-1}b \\ 0 & 1 \end{bmatrix},$$

then we obtain $E_4E_3E_2E_1A = I_{2\times 2}$ so that $A^{-1} = E_4E_3E_2E_1$. Explicitly, the following holds.

$$A^{-1} = \begin{bmatrix} 1 & -a^{-1}b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (d - a^{-1}bc)^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -c & 1 \end{bmatrix} \begin{bmatrix} a^{-1} & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Consequently, our original matrix A is invertible if and only if $ad - bc$ is nonzero.

Example 1.5.10. We will compute one more example to demonstrate the method of Gaussian Elimination, but in this case, we will keep track of the elementary row operations in a simpler manner than in Example 1.5.9. Observe that if A is an $n \times n$ matrix, then we may construct the augmented matrix $\left[A \mid I_{n\times n}\right]$ by adjoining the $n \times n$ identity matrix $I_{n\times n}$ on the right-hand side of A. If A is invertible, then by performing elementary row operations to this augmented matrix, we may reduce A to $I_{n\times n}$ and simultaneously convert $I_{n\times n}$ to A^{-1}. Explicitly, we will obtain $\left[I_{n\times n} \mid A^{-1}\right]$.

Consider the following 3×3 matrix A and the resulting augmented matrix $\left[A \mid I_{3\times 3}\right]$.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix} \text{ and } \left[A \mid I_{3\times 3}\right] = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{array}\right]$$

We will carry out the Gaussian Elimination as follows, listing each elementary row operation.

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{array}\right] \overset{\substack{R_2 - R_1 \mapsto R_2 \\ R_3 - R_1 \mapsto R_3}}{\sim} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 1 & 1 & -1 & 0 & 1 \end{array}\right] \overset{R_2 \leftrightarrow R_3}{\sim} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right]$$

$$\overset{\substack{R_1 - R_3 \mapsto R_1 \\ R_2 - R_3 \mapsto R_2}}{\sim} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 2 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right]$$

$$\overset{R_1 - R_2 \mapsto R_1}{\sim} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{array}\right]$$

By the first paragraph above, we conclude ultimately that the inverse of A is given as follows.

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

1.6 Vector Spaces

Going forward, we will refer to a collection of like objects (such as real $m \times n$ matrices) as a **set**; the objects of a set are called **elements** or **members**. We will use the symbol $\in$ to denote **set membership**, i.e., we write that $s \in S$ if and only if s is an element of the set S.

Example 1.6.1. Consider the set S consisting of the first five positive integers 1, 2, 3, 4, and 5. We note that the elements of S are precisely the numbers 1, 2, 3, 4, and 5; in particular, we may write that $1 \in S$, $2 \in S$, and so on for each of the remaining three integers. We say in this case that S is a **finite set** because it has only finitely many members. We use curly braces to denote a finite set by its elements, hence we have that $S = \{1, 2, 3, 4, 5\}$. One thing to notice is that the arrangement of the elements of S does not matter because S is only keeping track of what belongs to it. Likewise, the number of times an element of S appears in S does not matter. Explicitly, it is true that $S = \{1, 2, 3, 4, 5\} = \{2, 4, 1, 3, 5\} = \{2, 4, 2, 1, 2, 3, 2, 5\}$; however, it is not true that $S = \{0, 1, 2, 3, 4, 5\}$ because the set $\{0, 1, 2, 3, 4, 5\}$ has the non-negative integer 0 as a member.

Example 1.6.2. Often, we will consider sets consisting of infinitely many elements; we call these **infinite sets**. Clearly, it is not possible to list the infinitely many elements of such a set, hence we turn to the so-called **set-builder notation** to describe the elements of an infinite set. For instance, the set of real numbers $\mathbb{R}$ is an infinite set; its elements are simply real numbers, so in set-builder notation, we write $\mathbb{R} = \{x \mid x \text{ is a real number}\}$, and we read this as, "$\mathbb{R}$ is the set of all elements x such that x is a real number." Explicitly, in set-builder notation, we may describe a set S as

$$S = \{\text{the objects of } S \mid \text{the set membership property for } S\}.$$

Back to our example of the real numbers, the objects in $\mathbb{R}$ are denoted by x, and the set membership property for $\mathbb{R}$ is that x is a real number. Put another way, in set-builder notation for a set S, the objects of the set S come first; then, we put a vertical bar $\mid$ to signify the phrase "such that"; and finally, we put the condition under which an object belongs to the set S in question.

Example 1.6.3. Consider the collection $\mathbb{R}^{m \times n}$ of real $m \times n$ matrices; this is an infinite set whose set membership condition can be expressed as $A \in \mathbb{R}^{m \times n}$ if and only if A is a real $m \times n$ matrix. Consequently, in set-builder notation, we have that $\mathbb{R}^{m \times n} = \{A \mid A \text{ is a real } m \times n \text{ matrix}\}$.

Example 1.6.4. Consider the collection $\mathbb{R}[x]$ of real polynomials in indeterminate x; this is an infinite set whose set membership condition can be expressed as $p(x) \in \mathbb{R}[x]$ if and only if $p(x)$ is a real polynomial in indeterminate x. Consequently, in set-builder notation, we have that

$$\mathbb{R}[x] = \{p(x) \mid p(x) \text{ is a real polynomial in indeterminate } x\}.$$

One other way to realize this set in set-builder notation is to notice that every real polynomial in indeterminate x can be written as $a_n x^n + \cdots + a_1 x + a_0$ for some non-negative integer n and some real numbers $a_n, \dots, a_1, a_0$. Consequently, under this identification, we may also write that

$$\mathbb{R}[x] = \{a_n x^n + \cdots + a_1 x + a_0 \mid n \text{ is a non-negative integer and } a_n, \dots, a_1, a_0 \text{ are real numbers}\}.$$

Back in Example 1.1.4, we referred to any (real) $1 \times n$ matrix as a $1 \times n$ row vector. Our objective throughout this section is to demonstrate that the vector terminology can be applied much more broadly than simply in the scope of matrices. We begin by making the following definition.

Definition 1.6.5. We say that a pair $(V, +)$ is a (real) **vector space** if the following hold.

(1.) We have that $u + v \in V$ for any pair of elements $u, v \in V$.

(2.) We have that $(u + v) + w = u + (v + w)$ for any elements $u, v, w \in V$.

(3.) We have that $u + v = v + u$ for any pair of elements $u, v \in V$.

(4.) We have an element $O_V \in V$ such that $v + O_V = v$ for all elements $v \in V$.

(5.) Given any element $v \in V$, there exists an element $-v \in V$ such that $v + (-v) = O_V$.

(6.) We have that αv is an element of V for all (real) scalars α and elements $v \in V$.

(7.) We have that $1v = v$ for each element $v \in V$.

(8.) We have that $\alpha(\beta v) = (\alpha\beta)v$ for all (real) scalars α and β and elements $v \in V$.

(9.) We have that $\alpha(u + v) = \alpha u + \alpha v$ for all (real) scalars α and elements $u, v \in V$.

(10.) We have that $(\alpha + \beta)u = \alpha v + \beta v$ for all (real) scalars α and β and each element $v \in V$.

We refer to the elements $v \in V$ as (real) **vectors** in this case.

Combined, the first five properties of Definition 1.6.5 demonstrate that any vector space V constitutes an **abelian group** with respect to the addition defined on its elements. Groups form a central object of study in modern algebra, but we will not concern ourselves with their study here.

Our next example illustrates that the collection of real $m \times n$ matrices forms a real vector space.

Example 1.6.6. Consider any positive integers m and n. We denote by $\mathbb{R}^{m \times n}$ the collection of all real $m \times n$ matrices. Observe that the following properties hold, hence $\mathbb{R}^{m \times n}$ is a real vector space.

(1.) By definition, for any pair of $m \times n$ matrices A and B, the matrix sum $A + B$ is the real $m \times n$ matrix whose (i, j)th entry is the sum of the (i, j)th entries of A and B.

(2.) By definition, matrix addition is associative because addition of real numbers is associative.

(3.) Likewise, matrix addition is commutative because addition of real numbers is commutative.

(4.) By Example 1.1.6, the $m \times n$ zero matrix $O_{m \times n}$ is the unique real $m \times n$ matrix with the property that $A + O_{m \times n} = A$ for all real $m \times n$ matrices A.

(5.) By Example 1.1.13, for every real $m \times n$ matrix A, there exists a unique real $m \times n$ matrix $-A$ such that $A + (-A) = O_{m \times n}$ for the $m \times n$ zero matrix $O_{m \times n}$. Explicitly, $-A$ is the $m \times n$ matrix whose (i, j)th entry is the (i, j)th entry of A with the opposite sign.

(6.) By the paragraph preceding Example 1.1.13, if A is a real $m \times n$ matrix, then we have that cA is the real $m \times n$ matrix whose (i, j)th entry is c times the (i, j)th entry of A.

(7.) Likewise, if A is a real $m \times n$ matrix, then we have that $1A = A$.

(8.) Even more, if A is a real $m \times n$ matrix, then $c(dA) = (cd)A$ for all real numbers c and d.

(9.) By definition of matrix addition and the paragraph preceding Example 1.1.13, we have that $c(A + B) = cA + cB$ for all real numbers c and all real $m \times n$ matrices A and B.

(10.) Last, by the paragraph preceding Example 1.1.13, we have that $(c + d)A = cA + dA$ for all real numbers c and d and all real $m \times n$ matrices A.

Example 1.6.7. Consider the collection $F(\mathbb{R}, \mathbb{R})$ of real functions $f : \mathbb{R} \to \mathbb{R}$. We may define function addition so that if $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are any functions, then $f + g$ is the function satisfying $(f + g)(x) = f(x) + g(x)$ for all real numbers x, and we may define scalar multiplication so that $(\alpha f)(x) = \alpha f(x)$. Observe that the following hold, hence $F(\mathbb{R}, \mathbb{R})$ is a real vector space.

(1.) Given any functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, the function $f + g$ sends a real number x to the real number $f(x) + g(x)$. Consequently, we have that $f + g \in F(\mathbb{R}, \mathbb{R})$.

(2.) By definition, function addition is associative because addition of real numbers is associative.

(3.) Likewise, function addition is commutative because addition of real numbers is commutative.

(4.) Consider the function $O : \mathbb{R} \to \mathbb{R}$ defined by $O(x) = 0$ for all real numbers x. Given any function $f : \mathbb{R} \to \mathbb{R}$, we have that $(f + O)(x) = f(x) + O(x) + f(x) + 0 = f(x)$ for all real numbers x. We conclude therefore that $f + O = f$, i.e., $f + O$ and f are the same function.

(5.) Given any function $f : \mathbb{R} \to \mathbb{R}$, we may define the function $-f : \mathbb{R} \to \mathbb{R}$ by $(-f)(x) = -f(x)$. Observe that $(f+(-f))(x) = f(x)-f(x) = 0 = O(x)$ for all real numbers x and $f+(-f) = O$.

(6.) Given any function $f : \mathbb{R} \to \mathbb{R}$ and any real number α, it holds that $(\alpha f)(x) = \alpha f(x)$ is a real number for all real numbers x, from which it follows that $\alpha f \in F(\mathbb{R}, \mathbb{R})$.

(7.) Given any function $f : \mathbb{R} \to \mathbb{R}$, we have that $(1f)(x) = 1f(x) = f(x)$ for all real numbers x.

(8.) Given any function $f : \mathbb{R} \to \mathbb{R}$, we have that $\alpha(\beta f) = (\alpha\beta)f$ for all real numbers α and β: indeed, we have that $(\alpha(\beta f))(x) = \alpha(\beta f)(x) = (\alpha\beta)f(x)$ for all real numbers x.

(9.) Given any functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, we have that $\alpha(f + g) = \alpha f + \alpha g$ for all real numbers α because it holds that $\alpha(f+g)(x) = \alpha[f(x)+g(x)] = \alpha f(x)+\alpha g(x) = (\alpha f+\alpha g)(x)$.

(10.) Given any function $f : \mathbb{R} \to \mathbb{R}$, we have $(\alpha + \beta)f = \alpha f + \beta f$ for all real numbers α and β because it holds that $((\alpha + \beta)f)(x) = (\alpha + \beta)f(x) = \alpha f(x) + \beta f(x) = (\alpha f + \beta f)(x)$.

Given any vector O_V of a vector space V satisfying property (4.) of Definition 1.6.5, we say that O_V is a **zero vector**. We demonstrate that a vector space V has one and only one zero vector.

Proposition 1.6.8. *Let $(V, +)$ be a vector space. Let O_V be a zero vector of V.*

1.) *Given any vector $u \in V$ satisfying that $u + v = v$ for every vector $v \in V$, it must hold that $u = O_V$. Consequently, the zero vector of a vector space is unique.*

2.) *We have that $0v = O_V$ for all vectors $v \in V$.*

Proof. (1.) Observe that if u is any vector of V with the property that $u + v = v$ for every vector v of V, then it holds $u + O_V = u$ by definition of a zero vector O_V. Conversely, we have that $u + O_V = O_V$ by assumption. We conclude therefore that $u = u + O_V = O_V$ so that $u = O_V$.

(2.) Given any vector $v \in V$, we obtain a vector $0v \in V$ satisfying that $0v = (0+0)v = 0v + 0v$. Consequently, by properties (2.) and (5.) of Definition 1.6.5, there exists a vector $-0v$ of V such that $0v = 0v + O_V = 0v + [0v + (-0v)] = (0v + 0v) + (-0v) = 0v + (-0v) = O_V$. □

Generally, throughout all of mathematics, one of the primary means of classifying an object is to study its subobjects. Given any vector space V, we say that a subset W of V is a **vector subspace** of V (or simply a subspace of V) if the ten properties of Definition 1.6.5 hold for W with respect to the addition and scalar multiplication of V. We provide next a short criterion for subspaces.

Proposition 1.6.9 (Three-Step Subspace Test)**.** *Let W be any subset of a vector space $(V, +)$. We have that $(W, +)$ is a vector subspace of V if and only if the following three properties hold.*

(1.) *We have that O_V is an element of W.*

(2.) *We have that $v + w$ is an element of W for any pair of vectors $v, w \in W$.*

(3.) *We have that αw is an element of W for all scalars α and all vectors $w \in W$.*

Proof. Certainly, if W is a vector subspace of V, then by Definition 1.6.5, it satisfies the second and third properties listed above. Even more, we may consider the zero vector O_W of W. Considering that W is a subset of V, we may view O_W as an element of V so that $O_W + O_W = O_W = O_W + O_V$. Cancelling O_W from both sides of this identity yields that $O_W = O_V$, as desired.

Conversely, we will demonstrate that if W is any subset of a vector space V that satisfies the three properties listed above, then it must satisfy all ten properties of Definition 1.6.5. Considering that W is a subset of V, it satisfies properties (2.), (3.), and (7.) through (10.); it satisfies properties (1.), (4.), and (6.) by assumption; hence it suffices to prove that it satisfies property (5.). By the third property above, we have that $-w$ is an element of W for all vectors $w \in W$; then, by the second property above, we have that $w + (-w)$ is an element of W that satisfies $w + (-w) = O_V$. □

Example 1.6.10. Consider the real vector space $\mathbb{R}^{m \times n}$ of real $m \times n$ matrices. Consider the subset $W = \{A \in \mathbb{R}^{m \times n} \mid \text{the first row of } A \text{ is zero}\}$. Observe that the $m \times n$ zero matrix $O_{m \times n}$ lies in W because the first row of $O_{m \times n}$ is zero; the sum of any matrices A and B of W lies in W because the first row of $A + B$ is the sum of the first row of A and the first row of B, and both of these rows are zero; and the scalar multiple cA of any matrix $A \in W$ lies in W for all real numbers c because the first row of cA is c times the first row of A, and this is zero because the first row of A is zero. By the Three-Step Subspace Test, we have that W is a real vector subspace of $\mathbb{R}^{m \times n}$.

Example 1.6.11. Consider the real vector space $\mathbb{R}^{n \times n}$ of real $n \times n$ matrices. Consider the subset $W = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is symmetric}\}$. Observe that the $n \times n$ zero matrix $O_{n \times n}$ lies in W; the sum of any matrices A and B of W lies in W because $(A + B)^t = A^t + B^t$ by Proposition 1.1.14; and the scalar multiple cA lies in W for all real numbers c by [Lan86, Exercise 6] on page 47. Consequently, we conclude that W is a real vector subspace of $\mathbb{R}^{n \times n}$ by the Three-Step Subspace Test.

Example 1.6.12. Consider the real vector space $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ and its subset $\mathcal{C}^1(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ such that f' is continuous. Clearly, the zero function $O : \mathbb{R} \to \mathbb{R}$ is continuous. Likewise, the sum of continuous functions is continuous, hence if f' and g' are continuous, then $(f+g)' = f' + g'$ is continuous. Last, the scalar multiple of a continuous function is continuous, hence if f' is continuous, then $(\alpha f)' = \alpha f'$ is continuous for all real numbers α. We conclude that $\mathcal{C}^1(\mathbb{R})$ is a real vector subspace of $F(\mathbb{R}, \mathbb{R})$ by the Three-Step Subspace Test.

Example 1.6.13. Consider the real vector space $\mathcal{C}^1(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ such that f' is continuous. Consider the set $W = \{f \in \mathcal{C}^1(\mathbb{R}) \mid f(0) = 0\}$. Clearly, the zero function $O : \mathbb{R} \to \mathbb{R}$ lies in W because it satisfies that $O(0) = 0$; the sum of any functions f and g of W lies in W because we have that $(f+g)(0) = f(0) + g(0) = 0 + 0 = 0$; and the scalar multiple αf of a function $f \in W$ satisfies that $(\alpha f)(0) = \alpha f(0) = \alpha \cdot 0 = 0$, so it must lie in W for all real numbers α. We conclude that W is a real vector subspace of $\mathcal{C}^1(\mathbb{R})$ by the Three-Step Subspace Test.

Example 1.6.14. Consider the real vector space $\mathbb{R}^{n \times n}$ of real $n \times n$ matrices. Consider the subset $W = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is invertible}\}$. Observe that the $n \times n$ zero matrix $O_{n \times n}$ is not invertible, hence it does not lie in W. By the Three-Step Subspace Test, we conclude that W is not a vector subspace of $\mathbb{R}^{n \times n}$. Even more, the subset $W' = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is not invertible}\}$ does not constitute a vector subspace of V: all though the $n \times n$ zero matrix $O_{m \times n}$ lies in W', this set does not satisfy the first property of Definition 1.6.5 because the $n \times n$ identity matrix is the sum of non-invertible matrices.

Using the Three-Step Subspace Test, we furnish even shorter characterizations of a subspace.

Proposition 1.6.15 (Two-Step Subspace Test)**.** *Let W be any nonempty subset of a vector space V. We have that W is a vector subspace of V if and only if the following two properties hold.*

(1.) *We have that $v + w$ is an element of W for any pair of vectors $v, w \in W$.*

(2.) *We have that αw is an element of W for all scalars α and all vectors $w \in W$.*

Proof. By the Three-Step Subspace Test, if W is a vector subspace of V, then these conditions hold. Conversely, if the second condition above holds, then it follows that $-w$ is an element of W for all elements w of W. Likewise, if the first condition holds, then by assumption that W is nonempty, we have that $O_V = w + (-w)$ is an element of W; we are done by the Three-Step Subspace Test. □

Proposition 1.6.16 (One-Step Subspace Test)**.** *If W is any nonempty subset of a vector space V, then W is a subspace of V if and only if $\alpha v + \beta w \in W$ for any vectors $v, w \in W$ and scalars α, β.*

Proof. By the Two-Step Subspace Test, if W is a vector subspace of V, then these conditions hold. Conversely, if $\alpha v + \beta w$ lies in W for any vectors $v, w \in W$ and any scalars α and β, then $v + w = 1v + 1v \in W$ and $\alpha w = 0v + \alpha w \in W$; we are done by the Two-Step Subspace Test. □

We will distinguish in our next proposition two very important vector subspaces.

Proposition 1.6.17. *Let V be a vector space with vector subspaces U and W.*

(1.) *Let $U + W$ denote the collection of all vectors $u + w$ such that u is a vector of U and w is a vector of W. We have that $U + W$ is a vector subspace of V that contains both U and W.*

(2.) *Let $U \cap W$ denote the collection of all vectors v such that v is a vector of both U and W. We have that $U \cap W$ is a vector subspace of V contained in both U and W.*

Proof. Use the Three-Step Subspace Test. We leave this as an exercise for the reader. □

1.7 Span and Linear Independence

We will assume throughout this section that V is a (real) vector space. Given any vectors $v_1, \dots, v_n$ of V, a **linear combination** of $v_1, \dots, v_n$ is any vector of the form $\alpha_1 v_1 + \cdots + \alpha_n v_n$ for some (real) scalars $\alpha_1, \dots, \alpha_n$. We refer to the collection of linear combinations of the vectors $v_1, \dots, v_n$ as the **span** of the vectors $v_1, \dots, v_n$, and we write $\operatorname{span}\{v_1, \dots, v_n\}$ to denote this set. Explicitly, an element of $\operatorname{span}\{v_1, \dots, v_n\}$ is of the form $\alpha_1 v_1 + \cdots + \alpha_n v_n$ for some (real) scalars $\alpha_1, \dots, \alpha_n$. We will say that V is **generated** by the vectors $v_1, \dots, v_n$ if it holds that $V = \operatorname{span}\{v_1, \dots, v_n\}$.

Example 1.7.1. Given any positive integer n, consider the real vector space $\mathbb{R}^{1 \times n}$ of real row vectors of length n. By [Lan86, Exercise 11] on page 47, every element of $\mathbb{R}^{1 \times n}$ can be written as $x_1 E_1 + \cdots + x_n E_n$ for some real numbers $x_1, \dots, x_n$, where E_i is the $1 \times n$ row vector consisting of 1 in the ith column and zeros elsewhere. Consequently, it follows that $\mathbb{R}^{1 \times n} = \operatorname{span}\{E_1, \dots, E_n\}$.

Example 1.7.2. Given any positive integer n, consider the real vector space $\mathbb{R}^{2 \times 2}$ of real 2×2 matrices. Let E_{ij} denote the 2×2 matrix whose (i, j)th component is 1 and whose other components are zero. Observe that every real 2×2 matrix can be written as a linear combination

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ c & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix} = aE_{11} + bE_{12} + cE_{21} + dE_{22}$$

for any real numbers a, b, c, and d. Consequently, it follows that $\mathbb{R}^{2 \times 2} = \operatorname{span}\{E_{11}, E_{12}, E_{21}, E_{22}\}$.

Example 1.7.3. Given any positive integer n, consider the collection $P_n(x)$ of real polynomials of degree at most n in indeterminate x. By Example 1.6.7, it follows that $P_n(x)$ is a nonempty subset of the real vector space $\mathcal{C}^1(\mathbb{R})$ of real functions whose first derivative is continuous. By the Two-Step Subspace Test, we conclude that $P_n(x)$ is a real vector space: indeed, the sum of two real polynomials of degree at most n is a real polynomial of degree at most n, and a real scalar multiple of any real polynomial of degree at most n is a real polynomial of degree at most n. Observe that every real polynomial $f(x)$ of degree at most n can be written as $f(x) = a_n x^n + \cdots + a_1 x + a_0$ for some real numbers $a_0, a_1, \dots, a_n$, hence we conclude that $P_n(x) = \operatorname{span}\{1, x, \dots, x^n\}$.

We say that a collection of vectors $v_1, \dots, v_n$ are **linearly independent** whenever it holds that $\alpha_1 v_1 + \cdots + \alpha_n v_n = O_V$ implies that $\alpha_1 = \cdots = \alpha_n = 0$, i.e., the only linear combination of $v_1, \dots, v_n$ that is the zero vector is the linear combination of all zeros. Conversely, if there exist scalars $\alpha_1, \dots, \alpha_n$ not all of which are zero such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$, then we say that $v_1, \dots, v_n$ are **linearly dependent**. Observe that in this case, there exists a nonzero scalar α_i such that $\alpha_i v_i = -\alpha_1 v_1 - \cdots - \alpha_n v_n$ and $v_i = -\alpha_1 \alpha_i^{-1} v_1 - \cdots - \alpha_n \alpha_i^{-1} v_n$, i.e., the vector v_i can be written as a linear combination of the vectors $v_1, \dots, v_n$ excluding v_i. Consequently, any collection of vectors including O_V is linearly dependent, and we restrict our attention to nonzero vectors.

Example 1.7.4. Consider the real $1 \times n$ matrices E_i consisting of 1 in the ith column and zeros elsewhere. By [Lan86, Exercise 11] on page 47, we have that $E_1, \dots, E_n$ are linearly independent.

Example 1.7.5. Consider the real 2×2 matrices E_{ij} whose (i, j)th component is 1 and whose other components are zero. Observe that if a, b, c, and d are real numbers such that

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = aE_{11} + bE_{12} + cE_{21} + dE_{22} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ c & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

then $a = b = c = d = 0$. Consequently, it follows that $E_{11}, E_{12}, E_{21}, E_{22}$ are linearly independent.

Example 1.7.6. Consider the real polynomials $1, x, x^2, x^3$ of degree at most three. Observe that if $a_3x^3 + a_2x^2 + a_1x + a_0 = 0$, then all of the derivatives of this polynomial are zero.

$$3a_3x^2 + 2a_2x + a_1 = \frac{d}{dx}(a_3x^3 + a_2x^2 + a_1x + a_0) = \frac{d}{dx}(0) = 0$$

$$6a_3x + 2a_2 = \frac{d^2}{dx^2}(a_3x^3 + a_2x^2 + a_1x + a_0) = \frac{d^2}{dx^2}(0) = 0$$

$$6a_3 = \frac{d^3}{dx^3}(a_3x^3 + a_2x^2 + a_1x + a_0) = \frac{d^3}{dx^3}(0) = 0$$

Cancelling the factor of six from the last identity $6a_3 = 0$, we find that $a_3 = 0$. Likewise, the second derivative of this polynomial is $2a_2 = 0$ so that $a_2 = 0$. Continuing backwards, we conclude that $a_0 = a_1 = a_2 = a_3 = 0$. Ultimately, it follows that $1, x, x^2, x^3$ are linearly independent.

Example 1.7.7. Consider the real polynomials $1, x, \dots, x^n$ of degree at most n. Observe that if there exist real numbers $a_0, a_1, \dots, a_n$ such that $a_nx^n + \cdots + a_1x + a_0 = 0$, then all of the derivatives of this polynomial are zero. Particularly, the nth derivative of this polynomial is $n(n-1)\cdots 2a_n = 0$. Cancelling $n(n-1)\cdots 2$ from both sides, we find that $a_n = 0$. Likewise, the $(n-1)$th derivative of this polynomial is $(n-1)(n-2)\cdots 2a_{n-1}$ so that $a_{n-1} = 0$. Continuing backwards, we conclude that $a_0 = a_1 = \cdots = a_n = 0$. Ultimately, it follows that $1, x, \dots, x^n$ are linearly independent.

Example 1.7.8. Often, we will deal with (large) collections of vectors for which it is not obvious to detect linear independence. Explicitly, consider the vectors $v = (1, 1)$ and $w = (-3, 2)$ of the real vector space $\mathbb{R}^{1\times 2}$. By definition, the vectors v and w are linearly independent if and only if $\alpha v + \beta w = O$ implies that $\alpha = \beta = 0$. Expanding this equation by adding the corresponding columns of the vectors v and w (i.e., computing the matrix sum), we find that $(\alpha, \alpha) + (-3\beta, 2\beta) = (0, 0)$ or $(\alpha - 3\beta, \alpha + 2\beta) = (0, 0)$. Observe that this equation can be viewed as the following matrix equation.

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

Explicitly, the matrix on the left-hand side is the matrix whose columns are the vectors v and w; the scalars α and β are placed in a column vector and multiplied on the right of the matrix created from the given vectors; and the zero vector O is written as a column vector equal to this matrix product. Consequently, if the matrix whose columns are v and w is row equivalent to the $n \times n$ identity matrix $I_{n\times n}$, then it will follow that $\alpha = \beta = 0$, i.e., v and w are linearly independent. By the method of Gaussian Elimination, we obtain the unique reduced row echelon form as follows.

$$\begin{bmatrix} 1 & -3 \\ 1 & 2 \end{bmatrix} \overset{R_2 - R_1 \mapsto R_2}{\sim} \begin{bmatrix} 1 & -3 \\ 0 & 5 \end{bmatrix} \overset{\frac{1}{5}R_2 \mapsto R_2}{\sim} \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \overset{R_1 + 3R_2 \mapsto R_2}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

We conclude therefore that $v = (1, 1)$ and $w = (-3, 2)$ are linearly independent.

Our previous example gives rise to the following general method for determining all linearly independent vectors among a collection $v_1, \dots, v_n$ of real $1 \times m$ row vectors.

Algorithm 1.7.9. Let m and n be positive integers. Consider any real $1 \times m$ row vectors $v_1, \dots, v_n$. Use the following steps to find a (not necessarily unique) collection of linearly independent vectors of largest size among the vectors $v_1, \dots, v_n$. (Generally, this will depend on the order of $v_1, \dots, v_n$.)

(1.) Construct the real $m \times n$ matrix A whose jth column is the $m \times 1$ column vector v_j^t.

(2.) Use the method of Gaussian Elimination to convert A to its reduced row echelon form.

(3.) Every column of A that contains a pivot corresponds to a $1 \times m$ row vector that is linearly independent from all other vectors. Every column that does not possess a pivot corresponds to a $1 \times m$ row vector that can be written as a nonzero linear combination of some vectors.

Proof. Either there is a pivot in the jth column of the unique reduced row echelon form $\operatorname{RREF}(A)$ of the $m \times n$ matrix A, or there is not. By definition of the reduced row echelon form, if the jth column of $\operatorname{RREF}(A)$ contains a pivot, then this column must be the real $m \times 1$ matrix E_i^t with 1 in row i and zeros elsewhere for some integer $1 \leq i \leq j$; otherwise, for each integer $1 \leq i \leq m$ such that the (i,j)th component of $\operatorname{RREF}(A)$ is nonzero, there exists an integer $1 \leq k \leq j$ such that the (i,k)th component of $\operatorname{RREF}(A)$ is a pivot of 1. Consequently, the jth column of $\operatorname{RREF}(A)$ can be written as a nonzero linear combination of these column vectors, hence v_j is linearly dependent. $\square$

Example 1.7.10. We will use Algorithm 1.7.9 to determine the linearly independent vectors among the real 1×3 row vectors $v_1 = (1,1,1)$, $v_2 = (-1,1,1)$, $v_3 = (-1,-1,1)$, and $v_4 = (0,0,6)$. We must construct the 3×4 matrix whose jth column is v_j^t; then, we must subsequently convert this matrix into its unique reduced row echelon form. We illustrate this process this as follows.

$$\begin{bmatrix} 1 & -1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & 1 & 1 & 6 \end{bmatrix} \overset{(1.)}{\sim} \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 2 & 2 & 6 \end{bmatrix} \overset{(2.)}{\sim} \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 2 & 6 \end{bmatrix} \overset{(3.)}{\sim} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 6 \end{bmatrix} \overset{(4.)}{\sim} \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

(1.) We employed the elementary row operations $R_2 - R_1 \mapsto R_2$ and $R_3 - R_1 \mapsto R_3$.

(2.) We employed the elementary row operation $\frac{1}{2}R_2 \mapsto R_2$.

(3.) We employed the elementary row operations $R_1 + R_2 \mapsto R_1$ and $R_3 - 2R_2 \mapsto R_3$.

(4.) We employed the elementary row operations $\frac{1}{2}R_3 \mapsto R_3$ and $R_1 + R_3 \mapsto R_1$.

Consequently, the vectors v_1, v_2, and v_3 are linearly independent and $v_4 = 3v_1 + 0v_2 + 3v_3$.

We say that the vectors $v_1, \dots, v_n$ constitute a **basis** for the vector space V if and only if

(1.) $V = \operatorname{span}\{v_1, \dots, v_n\}$, i.e., V is spanned by $v_1, \dots, v_n$ and

(2.) $v_1, \dots, v_n$ are linearly independent, i.e., $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$ if and only if $\alpha_1 = \cdots = \alpha_n = 0$.

Example 1.7.11. Examples 1.7.1 and 1.7.4 demonstrate that the real $1 \times n$ matrices E_i consisting of 1 in the ith column and zeros elsewhere form a basis for the real vector space $\mathbb{R}^{1 \times n}$.

Example 1.7.12. Examples 1.7.2 and 1.7.5 demonstrate that the real $m \times n$ matrices E_{ij} consisting of 1 in the (i,j)th component and zeros elsewhere form a basis for the real vector space $\mathbb{R}^{m \times n}$.

Example 1.7.13. Examples 1.7.3 and 1.7.7 demonstrate that the polynomials $1, x, \dots, x^n$ of degree at most n form a basis for the real vector space $P_n(x)$ of real polynomials of degree at most n.

Given any basis $v_1, \dots, v_n$ of a vector space V, by definition, every vector of V can be written as a linear combination of the vectors $v_1, \dots, v_n$. Explicitly, for every vector $v \in V$, there exist scalars $\alpha_1, \dots, \alpha_n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. We refer to the scalars $\alpha_1, \dots, \alpha_n$ as the **coordinates** of v with respect to the **ordered basis** $(v_1, \dots, v_n)$. Often, we will write the coordinates of a vector with respect to an ordered basis as the ordered n-tuple $(\alpha_1, \dots, \alpha_n)$. Conventionally, this is due to the fact that every point $(x_1, \dots, x_n)$ in real n-space can be written as $x_1 E_1 + \cdots + x_n E_n$; we have already seen in Example 1.7.11 that $E_1, \dots, E_n$ form a basis for real n-space, so this terminology is merely a generalization of the familiar language from vector calculus and geometry. We demonstrate next that the coordinates of any vector with respect to an ordered basis are unique.

Proposition 1.7.14. *Let $v_1, \dots, v_n$ be linearly independent vectors that lie in some vector space V. If $\alpha_1 v_1 + \cdots + \alpha_n v_n = \beta_1 v_1 + \cdots + \beta_n v_n$, then we must have that $\alpha_1 = \beta_1, \dots, \alpha_n = \beta_n$. Consequently, the coordinates of every vector in the span of $v_1, \dots, v_n$ are unique (up to arrangement).*

Proof. Observe that if $\alpha_1 v_1 + \cdots + \alpha_n v_n = \beta_1 v_1 + \cdots + \beta_n v_n$, then subtracting $\beta_1 v_1 + \cdots + \beta_n v_n$ from both sides and combining like terms gives $(\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_n - \beta_n)v_n = O_V$. By assumption that $v_1, \dots, v_n$ are linearly independent, we conclude that $\alpha_i - \beta_i = 0$ for each integer $1 \leq i \leq n$. $\square$

Example 1.7.15. Consider the real 1×2 vectors $v = (1, 1)$ and $w = (-3, 2)$ of Example 1.7.8. We have already demonstrated that these vectors are linearly independent, hence in order to conclude that they form a basis for the real vector space $\mathbb{R}^{1 \times 2}$, it suffices to prove that they span $\mathbb{R}^{1 \times 2}$. We will achieve this by finding the coordinates α and β of any vector (a, b) with respect to v and w. By definition, we seek real numbers α and β that form a solution to the following matrix equation.

$$\begin{bmatrix} 1 & -3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

Example 1.7.8 exhibits elementary row operations to convert the matrix on the left to reduced row echelon form; to find α and β, we carry out these operations on the following augmented matrix.

$$\left[\begin{array}{cc|c} 1 & -3 & a \\ 1 & 2 & b \end{array}\right] \overset{R_2 - R_1 \mapsto R_2}{\sim} \left[\begin{array}{cc|c} 1 & -3 & a \\ 0 & 5 & b-a \end{array}\right] \overset{\frac{1}{5}R_2 \mapsto R_2}{\sim} \left[\begin{array}{cc|c} 1 & -3 & a \\ 0 & 1 & \frac{1}{5}(b-a) \end{array}\right] \overset{R_1 + 3R_2 \mapsto R_2}{\sim} \left[\begin{array}{cc|c} 1 & 0 & \frac{1}{5}(2a+3b) \\ 0 & 1 & \frac{1}{5}(b-a) \end{array}\right]$$

Consequently, we find that $(a, b) = \frac{1}{5}(2a + 3b)(1, 1) + \frac{1}{5}(b - a)(-3, 2)$ for all real numbers a and b.

1.8 Vector Space Dimension

Our first objective in this section is to demonstrate that if some vectors $v_1, \dots, v_n$ form a basis for the vector space V, then the non-negative integer n is unique. We refer to this number as the (vector space) **dimension** of V, and we write in this case that $\dim(V) = n$. Essentially, this fact follows as a corollary of the proposition that states that if some nonzero vectors $v_1, \dots, v_n$ span the vector space V, then any collection of linearly independent vectors consists of no more than n vectors.

Proposition 1.8.1. *Let V be a vector space that is spanned by some nonzero vectors $v_1, \dots, v_n$. Given any integer $m > n$, every collection of nonzero vectors $w_1, \dots, w_m \in V$ is linearly dependent.*

Proof. By hypothesis that V is spanned by $v_1, \dots, v_n$, for every collection of nonzero vectors $w_1, \dots, w_m \in V$, there exist scalars $\alpha_{11}, \dots, \alpha_{1n}, \dots, \alpha_{m1}, \dots, \alpha_{mn}$ such that the following hold.

$$\begin{aligned} w_1 &= \alpha_{11} v_1 + \cdots + \alpha_{1n} v_n \\ &\vdots \\ w_m &= \alpha_{m1} v_1 + \cdots + \alpha_{mn} v_n \end{aligned}$$

Consider the $m \times n$ matrix A whose (i, j)th component is α_{ij}. We note that A is a nonzero matrix because at least one of the scalars α_{ij} is nonzero. By hypothesis that $m > n$, the reduced row echelon form for A will have (at least) one zero row at the bottom (because it is impossible for a pivot to exist in row m). Consequently, there exist scalars $\beta_1, \dots, \beta_m$ such that $\beta_1 w_1 + \cdots + \beta_m w_m = O_V$. □

Corollary 1.8.2. *Let V be a vector space. If the vectors $v_1, \dots, v_n$ and the vectors $w_1, \dots, w_m$ form bases for V, then we must have that $m = n$. Consequently, the dimension of V is well-defined.*

Proof. By Proposition 1.8.1, we must have that $m \leq n$ because V is spanned by $v_1, \dots, v_n$ and $w_1, \dots, w_m$ are linearly independent. Conversely, we must have that $n \leq m$ because V is spanned by $w_1, \dots, w_m$ and $v_1, \dots, v_n$ are linearly independent. We conclude that $m = n$, as desired. □

Example 1.8.3. By Example 1.7.11, the real $1 \times n$ matrices E_i consisting of 1 in the ith column and zeros elsewhere form a basis for the real vector space $\mathbb{R}^{1 \times n}$, hence we have that $\dim(\mathbb{R}^{1 \times n}) = n$.

Example 1.8.4. By Example 1.7.12, the real $m \times n$ matrices E_{ij} consisting of 1 in the (i, j)th component and zeros elsewhere form a basis for the real vector space $\mathbb{R}^{m \times n}$ so that $\dim(\mathbb{R}^{m \times n}) = mn$.

Example 1.8.5. By Example 1.7.13, the polynomials $1, x, \dots, x^n$ of degree at most n form a basis for the real vector space $P_n(x)$ of real polynomials of degree at most n, i.e., $\dim(P_n(x)) = n + 1$.

We have therefore demonstrated that for any vector space V that admits a basis $v_1, \dots, v_n$, the non-negative integer n is unique; it is called the vector space dimension of V, and it is denoted by $\dim(V)$. Observe that if V is the **zero vector space** (i.e., the vector space consisting only of the zero vector), then $\dim(V) = 0$ because there are no linearly independent vectors in V; otherwise, we will soon demonstrate that the dimension of a nonzero vector space is always positive. Before this, we must understand the following fundamental properties of vector space dimension.

Proposition 1.8.6. *If V is a vector space that is spanned by some vectors $v_1, \dots, v_n$, then the dimension of V is the largest positive integer m not exceeding n for which some vectors $v_{i_1}, \dots, v_{i_m}$ are linearly independent. Put another way, every collection of generators of V induces a basis of V.*

Proof. Consider the largest positive integer m not exceeding n for which some vectors $v_{i_1}, \dots, v_{i_m}$ are linearly independent. We may assume these vectors are simply $v_1, \dots, v_m$; if they are not, then we may rearrange the subscripts. By Corollary 1.8.2, it suffices to demonstrate that $v_1, \dots, v_m$ span V. Observe that for each integer $m + 1 \leq k \leq n$, we have that $v_1, \dots, v_m, v_k$ are linearly dependent by definition of m. Consequently, there exist scalars $\alpha_1, \dots, \alpha_m, \alpha_k$ not all of which are zero such that $\alpha_1 v_1 + \cdots + \alpha_m v_m + \alpha_k v_k = O_V$. Observe that if $\alpha_k = 0$, then $\alpha_1 = \cdots = \alpha_m = 0$ by assumption that $v_1, \dots, v_m$ are linearly independent, so it must be the case that α_k is nonzero. Particularly, we may solve for v_k to find that $v_k = -\alpha_1 \alpha_k^{-1} v_1 - \cdots - \alpha_m \alpha_k^{-1} v_m$. Considering that $m+1 \leq k \leq n$ is an arbitrary integer, it follows that $v_{m+1}, \dots, v_n$ lie in the span of $v_1, \dots, v_m$. By hypothesis that V is spanned by the vectors $v_1, \dots, v_n$, for every vector $v \in V$, there exist scalars $\alpha_1, \dots, \alpha_n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Each of the vectors $v_{m+1}, \dots, v_n$ can be replaced by a linear combination of the vectors $v_1, \dots, v_m$, hence every vector of V can be written as a linear combination of $v_1, \dots, v_m$. □

Proposition 1.8.7. *If V is a vector space that admits linearly independent vectors $v_1, \dots, v_n$ such that $v_1, \dots, v_n, v$ are linearly dependent for all vectors $v \in V$, then $v_1, \dots, v_n$ is a basis for V. Put another way, the dimension of V is the largest number of linearly independent vectors of V.*

Proof. By definition of a basis, it suffices to demonstrate that $v_1, \dots, v_n$ span V. Given any vector $v \in V$, there exist scalars $\alpha_1, \dots, \alpha_n, \alpha$ not all of which are zero and $\alpha_1 v_1 + \cdots + \alpha_n v_n + \alpha v = O_V$ by hypothesis that $v_1, \dots, v_n, v$ are linearly dependent. On the other hand, the linear independence of $v_1, \dots, v_n$ implies that if $\alpha = 0$, then $\alpha_1 = \cdots = \alpha_n = 0$. Consequently, we must have that α is nonzero so that $v = \alpha_1 \alpha^{-1} v_1 + \cdots + \alpha_n \alpha^{-1} v_n$. We conclude that $V = \operatorname{span}\{v_1, \dots, v_n\}$. □

Corollary 1.8.8. *Let V be a vector space of finite dimension n. If $v_1, \dots, v_m$ are linearly independent vectors in V, then there exist nonzero vectors $v_{m+1}, \dots, v_n \in V$ such that $v_1, \dots, v_n$ form a basis for V. Put another way, every linearly independent subset of V can be extended to a basis of V.*

Proof. Begin with a collection of linearly independent vectors $v_1, \dots, v_m$. By Proposition 1.8.7, if $v_1, \dots, v_m, v$ are linearly dependent for all vectors $v \in V$, then $v_1, \dots, v_m$ constitute a basis for V; otherwise, there exists a nonzero vector $v_{m+1} \in V$ such that $v_1, \dots, v_{m+1}$ are linearly independent. Continuing in this manner yields nonzero vectors $v_{m+1}, \dots, v_n \in V$ such that $v_1, \dots, v_n$ are linearly independent and $v_1, \dots, v_n, v$ are linearly dependent for all vectors $v \in V$ by Proposition 1.8.1. Consequently, it follows from Proposition 1.8.7 that $v_1, \dots, v_n$ form a basis for V, as desired. □

Corollary 1.8.9. *Let V be a vector space of finite dimension. Let W be a vector subspace of V. We have that $0 \leq \dim(W) \leq \dim(V)$. Equality holds if and only if $W = \{O_V\}$ or $W = V$, respectively.*

Proof. By Proposition 1.8.7, we have that $\dim(W) = 0$ if and only if W admits no linearly independent vectors if and only if W admits no nonzero vectors if and only if $W = \{O_V\}$. Consequently, it suffices to establish that $\dim(W) \leq \dim(V)$ for every nonzero subspace W of V. Begin with any nonzero vector $w_1 \in W$. By Proposition 1.8.7, if w_1 and w are linearly dependent for every vector $w \in W$, then w_1 forms a basis for W; otherwise, there exists a nonzero vector $w_2 \in W$ such that w_1 and w_2 are linearly independent. Continuing in this manner yields nonzero vectors $w_2, \dots, w_m \in W$ such that $w_1, \dots, w_m$ are linearly independent and $w_1, \dots, w_m, w$ are linearly dependent for all vectors $w \in W$. Explicitly, by viewing the vectors $w_1, \dots, w_m, w$ as elements of V, we may appeal to Proposition 1.8.1 because V has finite dimension. Consequently, we conclude by Proposition 1.8.7 that the linearly independent vectors $w_1, \dots, w_m$ form a basis for W and $\dim(W) = m$. Even more, we must have that $m \leq \dim(V)$ by Proposition 1.8.1. Last, if $\dim(W) = \dim(V) = n$, then a basis for W must be a basis for V. Explicitly, if there were a basis $w_1, \dots, w_n$ of W that were not a basis for V, then there would exist a vector $v \in V$ that is not a linear combination of $w_1, \dots, w_n$, i.e., the vectors $w_1, \dots, w_n, v$ would be linearly independent. But this contradicts Proposition 1.8.7. □

Considering that the preceding four statements are so important, we outline them below. Going forward, we will say that a vector space is **finite-dimensional** if and only if it has finite dimension.

Theorem 1.8.10. *Let V be a finite-dimensional vector space.*

1.) *Every collection of vectors that span V can be refined to a basis for V.*

2.) *Every collection of linearly independent vectors of V can be extended to a basis for V.*

3.) *Every collection of $\dim(V)$ vectors that span V forms a basis for V.*

4.) *Every collection of $\dim(V)$ linearly independent vectors of V forms a basis for V.*

5.) *Every vector subspace W of V admits a basis.*

6.) *Every vector subspace W of V satisfies that $0 \leq \dim(W) \leq \dim(V)$. Even more, we have that $\dim(W) = 0$ if and only if $W = \{O_V\}$ and $\dim(W) = \dim(V)$ if and only if $W = V$.*

Before we conclude this section, we exhibit an example of an infinite-dimensional vector space.

Example 1.8.11. Consider the collection $\mathbb{R}[x]$ of real polynomials in indeterminate x. We claim that $\mathbb{R}[x]$ is an infinite-dimensional real vector space. By Example 1.6.7 and the Two-Step Subspace Test, it follows that $\mathbb{R}[x]$ is a real vector space because addition and scalar multiplication of real polynomials in x yields a real polynomial in x. We claim that the set $\{1, x, x^2, \dots\}$ of all non-negative integer powers of x forms a basis for $\mathbb{R}[x]$. By Example 1.7.7, the polynomials $1, x, \dots, x^n$ are linearly independent for each integer $n \geq 0$, hence $\{1, x, x^2, \dots\}$ is a linearly independent collection of vectors; it spans $\mathbb{R}[x]$ because every real polynomial in indeterminate x can be written as $a_n x^n + \cdots + a_1 x + a_0$ for some integer $n \geq 0$. Consequently, the dimension of $\mathbb{R}[x]$ is infinite.

1.9 Matrix Rank

Consider any $m \times n$ matrix A. Each column of A can be viewed as a $m \times 1$ column vector, hence it is natural to investigate the span of the column vectors that comprise A. Explicitly, suppose that $A_1, \dots, A_n$ are the $m \times 1$ column vectors such that A_j corresponds to the jth column of A. By definition, the span of these column vectors is the collection of all possible linear combinations of the vectors $A_1, \dots, A_n$, i.e., we have that $\operatorname{span}\{A_1, \dots, A_n\} = \{c_1 A_1 + \cdots + c_n A_n \mid c_1, \dots, c_n \text{ are scalars}\}$. We will refer to the vector space $\operatorname{span}\{A_1, \dots, A_n\}$ as the **column space** of A; the dimension of $\operatorname{span}\{A_1, \dots, A_n\}$ is commonly known as the **column rank** of A. Crucially, we note that the column space of A is nothing but the collection of all $m \times 1$ vectors of the form Ac^t, where c is any $1 \times n$ vector of the form $(c_1, \dots, c_n)$. Explicitly, we have that $Ac^t = c_1 A_1 + \cdots + c_n A_n$.

Example 1.9.1. Observe that the columns of the real 3×3 identity matrix $I_{3\times 3}$ are simply the real 3×1 vectors E_1^t, E_2^t, and E_3^t such that $E_1 = (1, 0, 0)$, $E_2 = (0, 1, 0)$, and $E_3 = (0, 0, 1)$. Consequently, the column space of $I_{3\times 3}$ is $\operatorname{span}\{E_1^t, E_2^t, E_3^t\} = \{\alpha_1 E_1^t + \alpha_2 E_2^t + \alpha_3 E_3^t \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}\} = \mathbb{R}^{3\times 1}$ by Example 1.7.1. Considering that $\dim(\mathbb{R}^{3\times 1}) = 3$ by Example 1.8.4, the column rank of $I_{3\times 3}$ is 3.

Example 1.9.2. Consider the real 3×4 matrix of Example 1.7.10 in reduced row echelon form.

$$A = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & 1 & 1 & 6 \end{bmatrix} \text{ and } \operatorname{RREF}(A) = \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Previously, we illustrated that the column vectors $(1,1,1)^t$, $(-1,1,1)^t$, and $(-1,-1,1)^t$ are linearly independent. Considering that $\mathbb{R}^{3\times 1}$ has dimension three by Example 1.8.4, we conclude by Proposition 1.8.7 that these vectors form a basis for $\mathbb{R}^{3\times 1}$, hence they form a basis for the column space of A. Consequently, the column rank of A is three. Likewise, the column rank of $\operatorname{RREF}(A)$ is three by the same rationale because the vectors $(1,0,0)^t$, $(0,1,0)^t$, and $(0,0,1)^t$ are linearly independent.

Example 1.9.3. Consider the following real 2×2 matrix.

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

By definition, the column space of A is $\operatorname{span}\{(1,1)^t, (0,0)^t\} = \{\alpha(1,1)^t + \beta(0,0)^t \mid \alpha, \beta \in \mathbb{R}\}$. Considering that $\beta(0,0)^t = (0,0)^t$ the column space of A is simply $\operatorname{span}\{(1,1)^t\} = \{(\alpha,\alpha)^t \mid \alpha \in \mathbb{R}\}$; it has dimension one, so the column rank of A is one. Observe that the reduced row echelon form

$$\operatorname{RREF}(A) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

for A has column space $\operatorname{span}\{(1,0)^t\} = \{(\alpha,0)^t \mid \alpha \in \mathbb{R}\}$, hence its column rank is also one.

We demonstrate next that this phenomenon is no coincidence: in fact, the column rank of a matrix is always equal to the column rank of its unique reduced row echelon form.

Proposition 1.9.4. *Every matrix has column rank equal to the column rank of its unique reduced row echelon form. Put another way, elementary row operations do not affect column rank.*

Proof. Consider an $m \times n$ matrix A with unique reduced row echelon form R. Let $A_1, \dots, A_n$ and $R_1, \dots, R_n$ denote the columns of A and R, respectively. By definition of the reduced row echelon form of A, there exists an invertible $m \times m$ matrix E such that $R = EA$. Consequently, it follows by matrix multiplication that $R_j = EA_j$ for each integer $1 \leq j \leq n$. Observe that if there exist scalars $c_1, \dots, c_n$ such that $c_1 R_1 + \cdots + c_n R_n = O$, then multiplying both sides of this vector equation on the left by E yields that $c_1 A_1 + \cdots + c_n A_n = O$. Conversely, if there exist scalars $d_1, \dots, d_n$ such that $d_1 A_1 + \cdots + d_n A_n = O$, then multiplying both sides of this vector equation on the left by E^{-1} yields that $d_1 R_1 + \cdots + d_n R_n = O$. We conclude therefore that the columns $A_{i_1}, \dots, A_{i_k}$ of A are linearly independent if and only if the columns $R_{i_1}, \dots, R_{i_k}$ are linearly independent. By Proposition 1.8.7 and the definition of column rank, we conclude that the column ranks of A and R are equal. $\square$

We may also consider the rows $a_1, \dots, a_m$ of an $m \times n$ matrix A, i.e., the $1 \times n$ row vectors a_i corresponding to the ith row of A. We define the **row rank** of A to be the dimension of the **row space** of A, i.e., the vector space $\operatorname{span}\{a_1, \dots, a_m\} = \{c_1 a_1 + \cdots + c_m a_m \mid c_1, \dots, c_m \text{ are scalars}\}$.

Example 1.9.5. Like before, the rows of the real 3×3 identity matrix $I_{3\times 3}$ are the real 3×1 vectors $E_1 = (1,0,0)$, $E_2 = (0,1,0)$, and $E_3 = (0,0,1)$; these vectors are linearly independent, and they span the three-dimensional space $\mathbb{R}^{1\times 3}$, so the row space of $I_{3\times 3}$ is $\mathbb{R}^{1\times 3}$.

Example 1.9.6. Consider the real 3×4 matrix of Example 1.9.1 and its reduced row echelon form.

$$A = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & 1 & 1 & 6 \end{bmatrix} \text{ and } \operatorname{RREF}(A) = \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Consider the row vectors $a_1 = (1, -1, -1, 0)$, $a_2 = (1, 1, -1, 0)$, and $a_3 = (1, 1, 1, 6)$. Certainly, the vector a_3 is linearly independent of the vectors a_1 and a_2 because it has a nonzero entry in its fourth column, and the fourth column of a_1 and a_2 is zero. Likewise, the vectors a_1 and a_2 are linearly independent: indeed, if we take any scalars c_1 and c_2 such that $c_1 a_1 + c_2 a_2 = O$, then it follows that $(c_1, -c_1, -c_1, 0) + (c_2, c_2, -c_2, 0) = (0, 0, 0, 0)$ so that $c_1 + c_2 = 0$ and $-c_1 + c_2 = 0$. By adding the first equation to the second, we find that $2c_2 = 0$ or $c_2 = 0$, from which it follows that $c_1 = 0$. Ultimately, we conclude that the row rank of A is three, and the row space of A is

$$\operatorname{span}\{a_1, a_2, a_3\} = \{(c_1 + c_2 + c_3, -c_1 + c_2 + c_3, -c_1 - c_2 + c_3, 6c_3) \mid c_1, c_2, c_3 \in \mathbb{R}\}.$$

Likewise, the row rank of $\operatorname{RREF}(A)$ is three because the vectors $r_1 = (1, 0, 0, 3)$, $r_2 = (0, 1, 0, 0)$, and $r_3(0, 0, 1, 3)$ are linearly independent: indeed, we have that $c_1 r_1 + c_2 r_2 + c_3 r_3 = O$ if and only if $(c_1, c_2, c_3, 3c_1 + 3c_3) = (0, 0, 0, 0)$ if and only $c_1 = c_2 = c_3 = 0$. Last, the row space of $\operatorname{RREF}(A)$ is

$$\operatorname{span}\{r_1, r_2, r_3\} = \{(c_1, c_2, c_3, 3c_1 + 3c_3) \mid c_1, c_2, c_2 \in \mathbb{R}\}.$$

Example 1.9.7. Consider the following real 2×2 matrix of Example 1.9.3.

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

Observe that the row space of A is $\operatorname{span}\{(1, 0), (1, 0)\} = \operatorname{span}\{(1, 0)\} = \{\alpha(1, 0) \mid \alpha \in \mathbb{R}\}$; this is also the row space for the unique reduced row echelon form of A below.

$$\operatorname{RREF}(A) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Consequently, A and $\operatorname{RREF}(A)$ have the same row space, and their row ranks are equal to one.

Like before, the previous examples are illustrative of a more general observation that the row space of any matrix is equal to the row space of its unique reduced row echelon form.

Proposition 1.9.8. *Every matrix has row space equal to the row space of its unique reduced row echelon form. Consequently, the row rank of a matrix is equal to the row rank of its reduced row echelon form. Put another way, elementary row operations do not affect row space or row rank.*

Proof. Consider an $m \times n$ matrix A with unique reduced row echelon form R. Let $a_1, \dots, a_m$ and $r_1, \dots, r_m$ denote the rows of A and R, respectively. Certainly, it does not affect the row space of A to interchange two rows of A because this amounts to relabelling the indices of some row vectors a_i and a_j, and the indices of the vectors in the span by definition do not matter. Likewise, taking a nonzero scalar multiple c of any row a_i of A does not affect the span of $a_1, \dots, a_m$ because any vector $c_1 a_1 + \cdots + c_m a_m$ in the span of $a_1, \dots, a_m$ is now given by $c_1 a_1 + \cdots + (c_1 c^{-1}) c a_i + \cdots + c_m a_m$.

Last, replacing any row a_j of A by the linear combination $ca_i + a_j$ for any scalar c and any integer $1 \leq i \leq m$ does not affect the span of $a_1, \dots, a_m$ because any vector $c_1 a_1 + \cdots + c_m a_m$ in the span of $a_1, \dots, a_m$ can be achieved as $c_1 a_1 + \cdots + (c_i - c_j c) a_i + \cdots + c_j (ca_i + a_j) + \cdots + c_m a_m$. Consequently, every vector in the span of $a_1, \dots, a_m$ lies in the span of $r_1, \dots, r_m$. Conversely, every row of R is a linear combination of some rows of A, hence every vector in the span of $r_1, \dots, r_m$ lies in the span of $a_1, \dots, a_m$. We conclude therefore that $\operatorname{span}\{a_1, \dots, a_m\} = \operatorname{span}\{r_1, \dots, r_m\}$, i.e., the row spaces of A and R are equal. Clearly, now, the row rank of A and the row rank of R are equal. □

Corollary 1.9.9. *Elementary column operations do not affect column rank.*

Proposition 1.9.10. *Elementary column operations do not affect row rank.*

Proof. By definition of the matrix transpose, elementary column operations on a matrix are equivalent to elementary row operations on the matrix transpose. By Proposition 1.9.4, elementary row operations on the matrix transpose do not affect the column rank of the matrix transpose, so elementary column operations on the matrix do not affect the row rank of the matrix. □

Proposition 1.9.11. *Every matrix can be reduced via a sequence of elementary row and column operations to a matrix containing the $r \times r$ identity matrix in the top left-hand corner and whose other rows and columns are all zero, where the non-negative integer r is equal to the row rank of the matrix. Even more, the row rank and the column rank of any matrix are equal.*

Proof. Consider an $m \times n$ matrix A with unique reduced row echelon form R. Observe that if A is the zero matrix, then its row rank and column rank are both zero, and the proposition is vacuously true. Consequently, we may assume that R is nonzero. By definition of the reduced row echelon form of a matrix, the nonzero rows of R are linearly independent; they span the row space of R, hence the number of nonzero rows of R is the row rank of R. By Proposition 1.9.8, the row rank of R is equal to the row rank of A, hence there are precisely r nonzero rows of R, where r is the row rank of A. Each of the r nonzero rows of R possesses a pivot of 1 in some column, and all other entries of any column containing a pivot are zero. By successively interchanging the columns of R, we obtain a matrix with the $r \times r$ identity matrix in the top left-hand corner and zeros in all subsequent rows. By construction of R, there exists a sequence of elementary row operations that reduce A to R, so in conjunction with the aforementioned column interchanges, we obtain a sequence of elementary row and column operations that reduces A to a matrix containing the $r \times r$ identity matrix in the top-left hand corner and whose subsequent rows are all zero. Considering that adding a scalar multiple of one column to another column is an elementary column operation, we can reduce any nonzero columns strictly to the right of column r to zero. Explicitly, if a is the (i, j)th component of the matrix and $1 \leq i \leq r$ and $r + 1 \leq j \leq n$, then $C_j - cC_i \mapsto C_j$ yields a 0 in the (i, j)th component of the resulting matrix. Each of these is an elementary column operation, so after a sequence of elementary column operations, we obtain the desired matrix of the proposition statement. Last, neither elementary row operations nor elementary column operations affect column rank by Propositions 1.9.4 and 1.9.9, hence the column rank of A is equal to the column rank of this matrix, which equals the row rank of the matrix, i.e., the row rank of A. □

Consequently, by Proposition 1.9.11, the row rank and column rank of any matrix coincide; their common value is referred to simply as the **rank** of A. Even more, the previous proposition is constructive in the sense that it gives a simple recipe to find the rank of a matrix.

Corollary 1.9.12. *The rank of a matrix is equal to the number of pivots of its row echelon form.*

Example 1.9.13. Consider the following real 3×3 matrix.

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

By Corollary 1.9.12, in order to find the rank of A, it suffices to find the row echelon form for A. We accomplish this by performing elementary row operations on A as follows.

$$\begin{bmatrix} 1 & -1 & 2 \\ 2 & 0 & 1 \\ 1 & -1 & 2 \end{bmatrix} \overset{\substack{R_3 - R_1 \mapsto R_3 \\ R_2 - 2R_1 \mapsto R_2}}{\sim} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

We have obtained pivots in rows one and two. Consequently, it follows that the rank of A is two.

1.10 Linear Transformations

We turn our attention next to the structure-preserving functions between vector spaces. Explicitly, if V and W are vector spaces, then a **linear transformation** is a function $T : V \to W$ such that

1.) $T(u + v) = T(u) + T(v)$ for all vectors $u, v \in V$ and

2.) $T(\alpha v) = \alpha T(v)$ for all vectors $v \in V$ and all scalars α.

Conveniently, it is possible to summarize these two **linearity** conditions as follows.

Proposition 1.10.1. *If V and W are vector spaces, then a function $T : V \to W$ is a linear transformation if and only if $T(\alpha u + v) = \alpha T(u) + T(v)$ for all vectors $u, v \in V$ and all scalars α.*

Proof. Certainly, if $T : V \to W$ is a linear transformation, then by the definition above, it holds that $T(\alpha u + v) = T(\alpha u) + T(v) = \alpha T(u) + T(v)$ for all vectors $u, v \in V$ and scalars α. Conversely, if $T(\alpha u + v) = \alpha T(u) + T(v)$ for all vectors $u, v \in V$ and all scalars α, then in particular, we have that $T(O_V) = T(0v + O_V) = 0T(v) + T(O_V) = O_W + T(O_V)$ for every vector $v \in V$. Cancelling $T(O_V)$ from both sides, we find that $T(O_V) = O_W$. Consequently, it follows that

1.) $T(u + v) = T(1u + v) = 1T(u) + T(v) = T(u) + T(v)$ and

2.) $T(\alpha u) = T(\alpha u + O_V) = \alpha T(u) + T(O_V) = \alpha T(u) + O_W = \alpha T(u)$

for all vectors $u, v \in V$ and scalars α. Consequently, the claim holds. □

Example 1.10.2. Consider the real vector spaces $\mathbb{R}^{m \times n}$ of real $m \times n$ matrices and $\mathbb{R}^{n \times m}$ of real $n \times m$ matrices for any positive integers m and n. We claim that matrix transposition is a linear transformation, i.e., we will demonstrate that the function $T : \mathbb{R}^{m \times n} \to \mathbb{R}^{n \times m}$ defined by $T(A) = A^t$ is a linear transformation. By Proposition 1.1.14 and [Lan86, Exercise 6] on page 47, we have that

$$T(cA + B) = (cA + B)^t = (cA)^t + B^t = cA^t + B^t = cT(A) + T(B)$$

for all real $m \times n$ matrices A and B and all scalars c. Consequently, by Corollary 1.10.1, we conclude that T is a linear transformation, hence matrix transposition is a linear transformation.

Example 1.10.3. Consider the real vector spaces $\mathbb{R}^{n\times r}$ of real $n \times r$ matrices and $\mathbb{R}^{m\times r}$ of real $m \times r$ matrices for any positive integers m, n, and r. We claim that matrix multiplication is a linear transformation, i.e., if A is any real $m \times n$ matrix, then the function $T_A : \mathbb{R}^{n\times r} \to \mathbb{R}^{m\times r}$ defined by $T_A(B) = AB$ is a linear transformation. By Proposition 1.2.6, we have that

$$T_A(cB + C) = A(cB + C) = A(cB) + AC = c(AB) + AC = cT_A(B) + T_A(C).$$

We conclude by Corollary 1.10.1 that $T_A : \mathbb{R}^{n\times r} \to \mathbb{R}^{m\times r}$ is a linear transformation.

Example 1.10.4. Consider the real vector spaces $\mathbb{R}^{1\times 3}$ of real 1×3 matrices and $\mathbb{R}^{1\times 2}$ of real 1×2 matrices. We claim that the function $T : \mathbb{R}^{1\times 3} \to \mathbb{R}^{1\times 2}$ defined by $T(x, y, z) = (x, y)$ is a linear transformation called the **projection** of (x, y, z) into the xy-plane. Observe that

$$\alpha(x_1, y_1, z_1) + (x_2, y_2, z_2) = (\alpha x_1, \alpha y_1, \alpha z_1) + (x_2, y_2, z_2) = (\alpha x_1 + x_2, \alpha y_1 + y_2, \alpha z_1 + z_2),$$

hence the image of $\alpha(x_1, y_1, z_1) + (x_2, y_2, z_2)$ under T is $(\alpha x_1 + x_2, \alpha y_1 + y_2) = \alpha(x_1, y_1) + (x_2, y_2)$. Considering that this is $\alpha T(x_1, y_1, z_1) + T(x_2, y_2, z_2)$, we conclude that T is a linear transformation.

Example 1.10.5. Consider the real vector space $P_1(x)$ of real linear polynomials. Explicitly, we have that $P_1(x) = \{mx + b \mid m \text{ and } b \text{ are real numbers}\}$, hence every element of $P_1(x)$ is graphically represented by a line in the Cartesian plane. Consider the function $D : P_1(x) \to P_1(x)$ defined by $D(mx + b) = m$. Explicitly, the function D maps a polynomial to its first derivative. Observe that

$$\alpha(m_1 x + b_1) + (m_2 x + b_2) = \alpha m_1 x + \alpha b_1 + m_2 x + b_2 = (\alpha_1 m_1 + m_2)x + (b_1 + b_2),$$

and the derivative of this function is $\alpha m_1 + m_2 = \alpha D(m_1 x + b_1) + D(m_2 x + b_2)$. Consequently, the derivative is a linear transformation from the real vector space $P_1(x)$ to itself.

Example 1.10.6. Consider the real vector spaces $\mathbb{R}$ of the real numbers and $\mathbb{R}^{2\times 2}$ of real 2×2 matrices. Consider the **determinant** function $\det : \mathbb{R}^{2\times 2} \to \mathbb{R}$ defined by

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

We claim that the determinant is not a linear transformation. Explicitly, for any real 2×2 matrix A and any real number c, we have that $\det(cA) = c^2 \det(A)$. Consider any two real 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

By definition of scalar multiplication, for any scalar c, we have that

$$cA = \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix}.$$

By definition of the determinant $\det(A)$ and $\det(cA)$, it follows that

$$\det(cA) = \det \begin{bmatrix} ca_{11} & ca_{12} \\ ca_{21} & ca_{22} \end{bmatrix} = c^2 a_{11}a_{22} - c^2 a_{12}a_{21} = c^2(a_{11}a_{22} - a_{12}a_{21}) = c^2 \det(A).$$

Consequently, the determinant is not a linear transformation.

Example 1.10.7. Consider the real vector space $\mathcal{C}^0(\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$. By the Fundamental Theorem of Calculus, for every function $f \in \mathcal{C}^0(\mathbb{R})$, there exists a function $F \in \mathcal{C}^1(\mathbb{R})$ such that $F'(x) = f(x)$; we refer to $F'(x)$ as an **antiderivative** of $f(x)$. Observe that for any antiderivative $F(x)$ of $f(x)$, we have that $G(x) = F(x) + C$ is an antiderivative of $f(x)$ for all real numbers C. Consequently, the function $A : \mathcal{C}^0(\mathbb{R}) \to \mathcal{C}^1(\mathbb{R})$ defined by $A(f) = F$ is not a linear transformation. Explicitly, every real number is an antiderivative of the zero function.

Conversely, if a is any real number, then we may define a function $R_a : \mathcal{C}^0(\mathbb{R}) \to \mathcal{C}^1(\mathbb{R})$ by declaring that $R_a(f) = \int_a^x f(t)\,dt$. We note that R_a is a linear transformation: indeed, we have that

$$R_a(\alpha f + g) = \int_a^x [\alpha f(t) + g(t)]\,dt = \alpha \int_a^b f(t)\,dt + \int_a^b g(t)\,dt = \alpha R_a(f) + R_a(g).$$

We collect in the next proposition two useful properties of linear transformations.

Proposition 1.10.8. *Let $T : V \to W$ be a linear transformation of the vector spaces V and W.*

1.) *We have that $T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$ for all vectors $v_1, \dots, v_n \in V$ and scalars $\alpha_1, \dots, \alpha_n$. Put another way, the image of a linear combination of vectors under a linear transformation is the linear combination of the images of the vectors.*

2.) *We have that $T(O_V) = O_W$, where O_V and O_W are the respective zero vectors of V and W.*

Proof. We prove the first property by the Principle of Mathematical Induction applied to the number of vectors n. By definition of a linear transformation, the claim holds for $n = 1$. We will assume inductively that $T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$ for all vectors $v_1, \dots, v_n \in V$ and scalars $\alpha_1, \dots, \alpha_n$. By definition of a linear transformation, we have that

$$T(\alpha_1 v_1 + \cdots + \alpha_n v_n + \alpha_{n+1} v_{n+1}) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) + T(\alpha_{n+1} v_{n+1}).$$

By hypothesis, the first summand is equal to $\alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$, from which it follows that $T(\alpha_1 v_1 + \cdots + \alpha_n v_n + \alpha_{n+1} v_{n+1}) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) + \alpha_{n+1} T(v_{n+1})$, as desired.

On the matter of the second property, we use the linearity of the function T to first illustrate that $T(O_V + O_V) = T(O_V) + T(O_V)$. On the other hand, it holds that $O_V + O_V = O_V$, hence we have that $T(O_V) + T(O_V) = T(O_V + O_V) = T(O_V)$. Cancelling $T(O_V)$ yields that $T(O_V) = O_W$. □

By Example 1.6.7, the collection of real functions $f : \mathbb{R} \to \mathbb{R}$ that have a continuous first derivative constitutes a real vector space; however, with a view toward linear algebra, there is nothing particularly special about real functions whose first derivative is continuous. Even more, one can prove that the collection of real functions $f : \mathbb{R} \to \mathbb{R}$ forms a real vector space by the same rationale as provided in the aforementioned example. Generalizing this idea, our next proposition states that the collection of all linear transformations between vector spaces is itself a vector space. Eventually, vector spaces of linear transformations will come to occupy much of our attention.

Proposition 1.10.9. *Let V and W be vector spaces. Let $\mathcal{L}(V, W)$ denote the collection of all linear transformations from V to W, i.e., $\mathcal{L}(V, W) = \{T : V \to W \mid T \text{ is a linear transformation}\}$. We have that $\mathcal{L}(V, W)$ is a vector space with respect to function addition and scalar multiplication.*

Proof. We must verify each of the ten axioms of a vector space from Definition 1.6.5.

(1.) Observe that if $S : V \to W$ and $T : V \to W$ are linear transformations, then $S + T : V \to W$ is the function defined by $(S + T)(v) = S(v) + T(v)$ for all vectors $v \in V$. By hypothesis that S and T are linear transformations, for all vectors $u, v \in V$ and all scalars α, it follows that

$$\begin{aligned}(S + T)(\alpha u + v) &= S(\alpha u + v) + T(\alpha u + v)\\ &= \alpha S(u) + S(v) + \alpha T(u) + T(v)\\ &= \alpha(S(u) + T(u)) + (S(v) + T(v))\\ &= \alpha(S + T)(u) + (S + T)(v).\end{aligned}$$

We conclude by Corollary 1.10.1 that $S + T : V \to W$ is a linear transformation.

(4.) Consider the function $O : V \to W$ defined by $O(v) = O_W$ for all vectors $v \in V$. Observe that for every vector $v \in V$, we have that $(T + O)(v) = T(v) + O(v) = T(v) + O_W = T(v)$, hence we conclude that $T + O = T$. Even more, O is a linear transformation.

(5.) Given any linear transformation $T : V \to W$, consider the function $-T : V \to W$. We have that $(T + (-T))(v) = T(v) - T(v) = O_W$ for all vectors $v \in V$, from which it follows that $T + (-T) = O$. Even more, $-T$ is a linear transformation by assumption that T is linear.

(6.) Last, if $T : V \to W$ is any linear transformation, then the function $\alpha T : V \to W$ defined in the obvious way is a linear transformation because T is a linear transformation.

Each of the remaining six vector space axioms is self-evident: by definition, for every vector $v \in V$, we have that $T(v)$ is a vector of W, hence function addition is associative and commutative because it is essentially vector addition. Likewise, scalar multiplication is associative and distributive. □

Example 1.10.10. Consider the real vector space of real numbers $\mathbb{R}$. By definition, we have that $\mathcal{L}(\mathbb{R}, \mathbb{R})$ is the real vector space of linear transformations $T : \mathbb{R} \to \mathbb{R}$. Consequently, the elements of $\mathcal{L}(\mathbb{R}, \mathbb{R})$ are functions $T : \mathbb{R} \to \mathbb{R}$ that satisfy that $T(x + y) = T(x) + T(y)$ and $T(\alpha x) = \alpha T(x)$ for all real numbers x, y, and α. Observe that if $T(\alpha x) = \alpha T(x)$, then in particular, we must have that $T(x) = T(x \cdot 1) = xT(1)$ for all real numbers x. Consequently, the elements of $\mathcal{L}(\mathbb{R}, \mathbb{R})$ are precisely the lines through the origin in $\mathbb{R}^2$, i.e., we have that $\mathcal{L}(\mathbb{R}, \mathbb{R}) = \{mx \mid m \in \mathbb{R}\}$.

1.11 Kernels and Images of Linear Transformations

Considering that a linear transformation $T : V \to W$ between two vector spaces V and W is nothing more than a linear function, it is natural to ask about the vectors of V that are mapped to the zero vector of W under T. Explicitly, we will consider the **kernel** of the linear transformation

$$\ker(T) = \{v \in V \mid T(v) = O_W\}.$$

Once again, the kernel of the linear transformation $T : V \to W$ is nothing more than the set of all vectors of V that result in the zero vector of W when we apply the linear transformation T to them. Our first order of business is to demonstrate that $\ker(T)$ is a vector subspace of V.

Proposition 1.11.1. *If $T : V \to W$ is a linear transformation of vector spaces V and W, then the kernel $\ker(T) = \{v \in V \mid T(v) = O_W\}$ of T is a subspace of V.*

Proof. By the Three-Step Subspace Test, it suffices to prove the following three properties.

(1.) By the second part of Proposition 1.10.8, we have that $T(O_V) = O_W$ so that $O_V \in \ker(T)$.

(2.) Consider two vectors $u, v \in \ker(T)$. By definition of the kernel, we have that $T(u) = O_W$ and $T(v) = O_W$, hence the linearity of T yields that $T(u + v) = T(u) + T(v) = O_W + O_W = O_W$. We conclude that $u + v$ lies in the kernel of T.

(3.) Last, if $v \in \ker(T)$ and α is any scalar, then $T(\alpha v) = \alpha T(v) = \alpha O_W = O_W$ because T is a linear transformation and $T(v) = O_W$; this demonstrates that $\alpha v \in \ker(T)$. $\square$

Example 1.11.2. Consider the linear transformation $T : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times m}$ of Example 1.10.2 defined by $T(A) = A^t$. By definition, we have that $\ker(T) = \{A \in \mathbb{R}^{m\times n} \mid A^t = T(A) = O_{n\times m}\}$. But if it holds that $A^t = O_{n\times m}$, then we must have that $A = O_{m\times n}$ so that $\ker(T) = \{O_{m\times n}\}$.

Example 1.11.3. Observe that if A is any real $m \times n$ matrix, then the function $T_A : \mathbb{R}^{n\times r} \to \mathbb{R}^{m\times r}$ of Example 1.10.3 defined by $T_A(B) = AB$ is a linear transformation; its kernel is given by

$$\ker(T_A) = \{B \in \mathbb{R}^{n\times r} \mid AB = T_A(B) = O_{m\times r}\}.$$

Consequently, if A is invertible, then $AB = O_{n\times r}$ if and only if $B = A^{-1}(AB) = A^{-1}O_{n\times r} = O_{n\times r}$. Put another way, the kernel of T_A for an invertible real $n \times n$ matrix A is $\ker(T_A) = \{O_{n\times r}\}$.

Concretely, let us find the kernel of T_A for the following real 2×2 matrix.

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

By definition, a real 2×2 matrix B is in the kernel of T_A if and only if $T_A(B)$ is the zero matrix if and only if the matrix product AB is the zero matrix, i.e., we have that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \ker(T_A) \text{ if and only if } \begin{bmatrix} a-c & b-d \\ -a+c & -b+d \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

if and only if $a - c = 0$ and $b - d = 0$ and $-a + c = 0$ and $-b + d = 0$ if and only if $a = c$ and $b = d$. Consequently, the kernel of T_A consists precisely of those 2×2 matrices of the form

$$\begin{bmatrix} a & b \\ a & b \end{bmatrix} = \begin{bmatrix} a & 0 \\ a & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \\ 0 & b \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} = a(E_{11} + E_{21}) + b(E_{12} + E_{22}).$$

We conclude therefore that $\ker(T_A) = \operatorname{span}\{E_{11} + E_{21}, E_{12} + E_{22}\}$.

Example 1.11.4. Consider the linear transformation $T : \mathbb{R}^{1\times 3} \to \mathbb{R}^{1\times 2}$ of Example 1.10.4 defined by $T(x, y, z) = (x, y)$, i.e., the projection of (x, y, z) into the xy-plane. We have that

$$\ker(T) = \{(x, y, z) \in \mathbb{R}^{1\times 3} \mid (x, y) = T(x, y, z) = (0, 0)\} = \{(0, 0, z) \mid z \in \mathbb{R}\} = \operatorname{span}\{(0, 0, 1)\}.$$

Example 1.11.5. Consider the differentiation transformation $D : P_1(x) \to P_1(x)$ of Example 1.10.5 defined by $D(mx + b) = m$. Observe that a polynomial $mx + b$ lies in the kernel of D if and only if $D(mx + b) = 0$ if and only if $m = 0$, i.e., $\ker(D) = \{mx + b \mid m = 0\} = \{b \mid b \in \mathbb{R}\}$ consists of all constant functions on $\mathbb{R}$. We note that this agrees with our intuition: by the Fundamental Theorem of Calculus, the derivative of any function is zero if and only if the function is constant.

We are especially interested in those linear transformations $T : V \to W$ with $\ker(T) = \{O_V\}$. We will say that $T : V \to W$ is **injective** if and only if $T(v_1) = T(v_2)$ implies that $v_1 = v_2$.

Proposition 1.11.6. *If $T : V \to W$ is a linear transformation of vector spaces V and W, then T is injective if and only if $\ker(T) = \{O_V\}$. Explicitly, $\ker(T)$ measures the failure of T to be injective.*

Proof. We will assume first that $T : V \to W$ is injective. Consider any vector $v \in \ker(T)$. By the definition of $\ker(T)$, we have that $T(v) = O_W = T(O_V)$. By assumption that T is injective, we conclude that $v = O_V$ and $\ker(T) = \{O_V\}$. Conversely, suppose that $\ker(T) = \{O_V\}$. Given any vectors $v_1, v_2 \in V$ such that $T(v_1) = T(v_2)$, we must have that $O_W = T(v_1) - T(v_2) = T(v_1 - v_2)$ by the linearity of T. Consequently, we have that $v_1 - v_2 \in \ker(T)$, from which it follows that $v_1 - v_2 = O_V$. By adding v_2 to both sides of this identity, we conclude that $v_1 = O_V + v_2 = v_2$. $\square$

Example 1.11.7. By Example 1.11.2, matrix transposition is an injective linear transformation.

Example 1.11.8. By Example 1.11.3, left-multiplication by an invertible (real) $n \times n$ matrix is an injective linear transformation from the vector space of (real) $n \times r$ matrices to itself.

Even more, we demonstrate in the next proposition that the linear transformations that preserve linear independence are precisely the injective linear transformations.

Proposition 1.11.9. *If $T : V \to W$ is a linear transformation of vector spaces V and W, then the following statements are equivalent.*

1.) *If $v_1, \dots, v_n$ are linearly independent, then $T(v_1), \dots, T(v_n)$ are linearly independent.*

2.) *We have that $\ker(T) = \{O_V\}$, i.e., T is injective.*

Put another way, a linear transformation is injective if and only if it preserves linear independence.

Proof. We will assume first that if $v_1, \dots, v_n$ are linearly independent, then $T(v_1), \dots, T(v_n)$ are linearly independent. Consider any vector $v \in \ker(T)$. By definition of the kernel, we have that $T(v) = O_W$, hence $T(v)$ is not linearly independent; this implies that v is not linearly independent, hence we must have that $v = O_V$ and $\ker(T) = \{O_V\}$. Conversely, suppose that $\ker(T) = \{O_V\}$. Given any linearly independent vectors $v_1, \dots, v_n$ of V, consider any scalars $\alpha_1, \dots, \alpha_n$ such that $\alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) = O_W$. By the first part of Proposition 1.10.8, we have that

$$O_W = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n)$$

so that $\alpha_1 v_1 + \cdots + \alpha_n v_n$ lies in $\ker(T)$. By hypothesis that $\ker(T) = \{O_V\}$, we must have that $\alpha_1 v_1 + \cdots + \alpha_n v_n$; then, the linear independence of $v_1, \dots, v_n$ yields that $\alpha_1 = \cdots = \alpha_n = 0$. $\square$

Conversely, we may consider the collection of all possible **images** $T(v)$ of the vectors v of V under a linear transformation $T : V \to W$. Explicitly, we refer to this as the **range**

$$\operatorname{range}(T) = \{w \in W \mid w = T(v) \text{ for some vector } v \in V\} = \{T(v) \mid v \in V\}$$

of the linear transformation. Occasionally, we will write $T(V) = \{T(v) \mid v \in V\}$ to emphasize that the linear transformation is acting on vectors of the vector space V. Under this identification, we may also define $T^{-1}(U) = \{v \in V \mid T(v) \in U\}$ for any vector subspace U of W; we refer to $T^{-1}(U)$ as the **pre-image** (or **inverse image**) of U under T. Like with the kernel of a linear transformation, it is true that the range of a linear transformation is a subspace of the target space W.

Proposition 1.11.10. *If $T : V \to W$ is a linear transformation of vector spaces V and W, then the range* $\operatorname{range}(T) = \{T(v) \mid v \in V\}$ *of T is a subspace of W.*

Proof. Once again, we must verify the following conditions of the Three-Step Subspace Test.

(1.) By the second part of Proposition 1.10.8, we have that $T(O_V) = O_W$ so that $O_W \in \operatorname{range}(T)$.

(2.) Consider any vectors $w, x \in \operatorname{range}(T)$. By definition of $\operatorname{range}(T)$, there exist vectors $u, v \in V$ such that $w = T(u)$ and $x = T(v)$. By assumption that T is a linear transformation, we find that $w + x = T(u) + T(v) = T(u + v)$. Considering that u and v are vectors of the vector space V, their sum $u + v$ is also a vector of V, from which it follows that $w + x \in \operatorname{range}(T)$.

(3.) Last, for any vector $w \in \operatorname{range}(T)$ and any scalar α, then there exists a vector $v \in V$ such that $\alpha w = \alpha T(v) = T(\alpha v)$. Like before, we find that αv lies in V so that αw lies in $\operatorname{range}(T)$. □

Example 1.11.11. Consider the linear transformation $T : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times m}$ of Example 1.10.2 defined by $T(A) = A^t$. By definition, we have that $\operatorname{range}(T) = \{A^t \mid A \in \mathbb{R}^{m\times n}\}$. Considering that any $n\times m$ matrix B can be written as $(B^t)^t$ and B^t is an $m\times n$ matrix, it follows that $\operatorname{range}(T) = \mathbb{R}^{n\times m}$.

Example 1.11.12. Given any real $m \times n$ matrix A, as in Example 1.10.3, we may define a linear transformation $T_A : \mathbb{R}^{n\times r} \to \mathbb{R}^{m\times r}$. We have that $\operatorname{range}(T_A) = \{AB \mid B \in \mathbb{R}^{n\times r}\}$. Observe that if $m = n$ and A is an invertible real $n \times n$ matrix, then for every real $n \times r$ matrix C, we have that

$$C = I_{n\times n}C = (AA^{-1})C = A(A^{-1}C) = T_A(A^{-1}C).$$

Consequently, in this case, every real $n \times r$ matrix C is the image of the real $n \times r$ matrix $A^{-1}C$ under the linear transformation T_A, from which it follows that $\operatorname{range}(T_A) = \mathbb{R}^{n\times r}$.

Example 1.11.13. Consider the linear transformation $T : \mathbb{R}^{1\times 3} \to \mathbb{R}^{1\times 2}$ of Example 1.10.4 defined by $T(x, y, z) = (x, y)$. We have that $\operatorname{range}(T) = \{(x, y) \mid x, y \in \mathbb{R}\} = \mathbb{R}^{1\times 2}$.

Example 1.11.14. Consider the differentiation transformation $D : P_1(x) \to P_1(x)$ of Example 1.10.5 defined by $D(mx + b) = m$. Observe that $\operatorname{range}(D) = \{m \mid m \in \mathbb{R}\}$ consists of all constant functions on $\mathbb{R}$. Coincidentally, it holds that $\operatorname{range}(T) = \ker(T)$; this is not typically true.

We will say that a linear transformation $T : V \to W$ is **surjective** if it holds that $\operatorname{range}(T) = W$. Consequently, the range of T measures the degree to which T is surjective. We illustrate next that surjective linear transformations are exactly those that preserve the span of a collection of vectors.

Proposition 1.11.15. *If $T : V \to W$ is a linear transformation of finite-dimensional vector spaces V and W, then the following statements are equivalent.*

1.) *If $V = \operatorname{span}\{v_1, \dots, v_n\}$, then $W = \operatorname{span}\{T(v_1), \dots, T(v_n)\}$.*

2.) *We have that $\operatorname{range}(T) = W$, i.e., T is surjective.*

Put another way, a linear transformation is surjective if and only if it preserves spanning sets.

Proof. By assumption that V is a finite-dimensional vector space, there exist vectors $v_1, \dots, v_n \in V$ such that $V = \operatorname{span}\{v_1, \dots, v_n\}$. Consequently, if the first statement of the proposition holds, then $W = \operatorname{span}\{T(v_1), \dots, T(v_n)\}$. By definition, for every vector $w \in W$, there exist scalars $\alpha_1, \dots, \alpha_n$ such that $w = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n)$. Considering that $\alpha_1 v_1 + \cdots + \alpha_n v_n$ lies in V, we conclude that $\operatorname{range}(T) = W$, hence T is surjective. Conversely, if T is surjective, then for every vector $w \in W$, there exists a vector $v \in V$ such that $w = T(v)$. By hypothesis that $V = \operatorname{span}\{v_1, \dots, v_n\}$, there exist scalars $\alpha_1, \dots, \alpha_n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $w = T(v) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$, i.e., $W = \operatorname{span}\{T(v_1), \dots, T(v_n)\}$. □

We demonstrate next that linear transformations preserve vector subspaces.

Proposition 1.11.16. *Let $T : V \to W$ be a linear transformation of vector spaces V and W.*

1.) *If U is a subspace of V, then $T(U)$ is a subspace of W.*

2.) *If U is a subspace of W, then $T^{-1}(U)$ is a subspace of V.*

Proof. We proceed by the Three-Step Subspace Test. Observe that if U is a subspace of V, then it holds that $O_V \in U$ so that $T(O_V) = O_W$ lies in $T(U)$. Even more, if $T(u)$ and $T(v)$ are any vectors in $T(U)$, then their sum $T(u) + T(v) = T(u + v)$ lies in $T(U)$ because $u + v$ lies in U. Last, if $T(u)$ is any vector in $T(U)$ and α is any scalar, then $\alpha T(u) = T(\alpha u)$ lies in $T(U)$ because αu lies in U.

Likewise, if U is a subspace of W, then we have that $T(O_V) = O_W \in U$ so that O_V lies in $T^{-1}(U)$. Given any vectors $u, v \in T^{-1}(U)$, by definition, there exist vectors $w, x \in U$ such that $T(u) = w$ and $T(v) = x$. By assumption that U is a subspace of W, it follows that $w + x = T(u) + T(v) = T(u + v)$ lies in U, hence we conclude that $u + v$ lies in $T^{-1}(U)$. Like before, if u is an element of $T^{-1}(U)$, then $T(\alpha u) = \alpha T(u)$ lies in U because $T(u)$ lies in the subspace U of W, hence αu lies in $T^{-1}(U)$. □

Before we conclude this section, we prove a result whose importance in practice cannot be understated. Briefly stated, the following proposition ensures that we may define a linear transformation $T : V \to W$ uniquely by declaring the images $T(v_i)$ for all basis vectors v_i of V under T; the image of any ordinary vector $v \in V$ is then determined by **extending linearly** according to the unique expression $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ of v in terms of some of these basis vectors.

Proposition 1.11.17. *Every linear transformation of vector spaces is uniquely determined by the images of any basis for the domain space. Explicitly, if $S : V \to W$ and $T : V \to W$ are linear transformations of vector spaces V and W such that $S(v_i) = T(v_i)$ for all vectors v_i of a basis for V, then it must hold that $S(v) = T(v)$ for all vectors $v \in V$, i.e., S and T must be the same function.*

Proof. Every vector $v \in V$ can be written uniquely as $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ for some basis vectors $v_1, \dots, v_n$ and scalars $\alpha_1, \dots, \alpha_n$. Consequently, if $S(v_i) = T(v_i)$ for all basis vectors v_i, then $S(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha S(v_1) + \cdots + \alpha S(v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n)$. □

1.12 The Rank-Nullity Theorem

Given any linear transformation $T : V \to W$ of vector spaces V and W, we obtain two vector spaces $\ker(T) = \{v \in V \mid T(v) = O_W\} \subseteq V$ and $\operatorname{range}(T) = \{T(v) \mid v \in V\} \subseteq W$ called the kernel and the range of T, respectively. Previously, we showed that $\ker(T)$ measures the failure of T to be injective and that $\operatorname{range}(T)$ measures the degree to which T is surjective. Even more, we noticed that T is injective if and only if it preserves linear independence, and likewise, T is surjective if and only if it preserves spanning sets. Consequently, if T is **bijective** (i.e., it is injective and surjective), then T preserves linear independence and spanning sets, hence it preserves bases.

Proposition 1.12.1. *If $T : V \to W$ is a linear transformation of finite-dimensional vector spaces V and W, then the following statements are equivalent.*

1.) *If $v_1, \dots, v_n$ form a basis for V, then $T(v_1), \dots, T(v_n)$ form a basis for W.*

2.) *We have that T is bijective, i.e., it is injective and surjective.*

Ultimately, we will come to find that a bijective linear transformation $T : V \to W$ encodes many desirable properties of the vector spaces V and W: in some sense, the existence of a bijective linear transformation between vector spaces V and W implies that V and W are "indistinguishable" other than by the "labels" of the vectors. We will elaborate on this property in due time.

One other way to measure certain properties of a linear transformation $T : V \to W$ is to find the dimensions of its kernel and range, i.e., the **nullity** $\operatorname{nullity}(T) = \dim(\ker(T))$ and the **rank** $\operatorname{rank}(T) = \dim(\operatorname{range}(T))$ of T. Often, this data provides a sufficient measure of the properties of T and will be preferable to the detailed information of the entire kernel or range of T.

Example 1.12.2. Consider the transposition transformation $T : \mathbb{R}^{m \times n} \to \mathbb{R}^{n \times m}$ of Examples 1.11.2 and 1.11.11 defined by $T(A) = A^t$. Previously, we demonstrated that $\operatorname{range}(T) = \mathbb{R}^{n \times m}$, hence we have that $\operatorname{rank}(T) = \dim(\mathbb{R}^{n \times m}) = mn = \dim(\mathbb{R}^{m \times n})$. On the other hand, we have that $\ker(T) = \{O_{m \times n}\}$ so that $\operatorname{nullity}(T) = 0$ and $\dim(\mathbb{R}^{m \times n}) = \operatorname{rank}(T) + \operatorname{nullity}(T)$.

Example 1.12.3. Given any real $m \times n$ matrix A, as in Examples 1.11.3 and 1.11.12, we may define a linear transformation $T_A : \mathbb{R}^{n \times r} \to \mathbb{R}^{m \times r}$. Like before, if we assume that $m = n$ and A is an invertible real $n \times n$ matrix, then $\operatorname{range}(T_A) = \mathbb{R}^{n \times r}$ so that $\operatorname{rank}(T_A) = nr$ and $\operatorname{nullity}(T_A) = 0$. Once again, in this case, we have that $\dim(\mathbb{R}^{n \times r}) = nr = \operatorname{rank}(T_A) + \operatorname{nullity}(T_A)$.

Example 1.12.4. Consider the linear transformation $T : \mathbb{R}^{1 \times 3} \to \mathbb{R}^{1 \times 2}$ of Examples 1.11.4 and 1.11.13 defined by $T(x, y, z) = (x, y)$. We have that $\operatorname{range}(T) = \mathbb{R}^{1 \times 2}$ and $\ker(T) = \operatorname{span}\{(0, 0, 1)\}$ so that $\operatorname{rank}(T) = 2$ and $\operatorname{nullity}(T) = 1$ and $\dim(\mathbb{R}^{1 \times 3}) = 3 = \operatorname{rank}(T) + \operatorname{nullity}(T)$.

Example 1.12.5. Consider the differentiation transformation $D : P_1(x) \to P_1(x)$ of Examples 1.11.5 and 1.11.14 defined by $D(mx + b) = m$. We showed before that $\operatorname{range}(T) = \ker(T) = \mathbb{R}$, hence we have that $\operatorname{rank}(T) = \ker(T) = \dim(\mathbb{R}) = 1$. Considering that the real polynomials 1 and x form a basis for $P_1(x)$, we find that $\dim(P_1(x)) = 2 = \operatorname{rank}(T) + \operatorname{nullity}(T)$.

Our main results of this section establish that the previous examples are illustrative of a more general relationship between the rank and nullity of a linear transformation.

Proposition 1.12.6. *If $T : V \to W$ is a linear transformation of vector spaces V and W, the linearly independent vectors of $\operatorname{range}(T)$ induce linearly independent vectors of V, i.e., $\operatorname{rank}(T) \leq \dim(V)$.*

Proof. Given any vectors $v_1, \dots, v_n \in V$, if $\alpha_1, \dots, \alpha_n$ are scalars such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = O_V$, then $O_W = T(O_V) = T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots \alpha_n T(v_n)$. Consequently, if $T(v_1), \dots, T(v_n)$ are linearly independent in W, then we must have that $\alpha_1 = \cdots = \alpha_n = 0$, hence $v_1, \dots, v_n$ are linearly independent. Ultimately, this shows that if $T(v_1), \dots, T(v_n)$ form a basis for range(T), then $v_1, \dots, v_n$ are linearly independent in V, hence we conclude that range$(T) = n \leq \dim(V)$. $\square$

Theorem 1.12.7 (Rank-Nullity Theorem)**.** *If $T : V \to W$ is a linear transformation of finite-dimensional vector spaces V and W, then it holds that* $\dim(V) = \operatorname{rank}(T) + \operatorname{nullity}(T)$.

Proof. Observe that if $\ker(T) = \{O_V\}$, i.e., if T is injective, then by Proposition 1.11.9, we have that $\dim(V) \leq \operatorname{rank}(T)$. Conversely, by Proposition 1.12.6, it always holds that $\dim(V) \geq \operatorname{rank}(T)$, hence in this case, we conclude that $\dim(V) = \operatorname{rank}(T) = \operatorname{rank}(T) + 0 = \operatorname{rank}(T) + \operatorname{nullity}(T)$. Consequently, we may assume that there exists a nonzero vector $v_1 \in \ker(T)$. By Theorem 1.8.10, there exist vectors $v_2, \dots, v_r \in \ker(T)$ such that $v_1, \dots, v_r$ form a basis for $\ker(T)$; likewise, there exist vectors $v_{r+1}, \dots, v_n \in V$ such that $v_1, \dots, v_r, v_{r+1}, \dots, v_n$ form a basis for V. We claim that $T(v_{r+1}), \dots, T(v_n)$ form a basis for range(T). Every vector v of V can be written as

$$v = \alpha_1 v_1 + \cdots + \alpha_r v_r + \alpha_{r+1} v_{r+1} + \cdots + \alpha_n v_n$$

for some scalars $\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_n$, hence every vector of range(T) can be written as

$$T(v) = T(\alpha_1 v_1 + \cdots + \alpha_r v_r + \alpha_{r+1} v_{r+1} + \cdots + \alpha_n v_n).$$

By the linearity of T, this above expression can be expanded to the following.

$$T(v) = \alpha_1 T(v_1) + \cdots + \alpha_r T(v_r) + \alpha_{r+1} T(v_{r+1}) + \cdots + \alpha_n T(v_n)$$

By assumption that $v_1, \dots, v_r$ lie in $\ker(T)$, it follows that every vector of W can be written as $\alpha_{r+1} T(v_{r+1}) + \cdots + \alpha_n T(v_n)$; this in turn implies that range$(T) = \operatorname{span}\{T(v_{r+1}), \dots, T(v_n)\}$. We must demonstrate next that $T(v_{r+1}), \dots, T(v_n)$ are linearly independent in W. Given any scalars $\alpha_{r+1}, \dots, \alpha_n$ such that $O_W = \alpha_{r+1} T(v_{r+1}) + \cdots + \alpha_n T(v_n) = T(\alpha_{r+1} v_{r+1} + \cdots + \alpha_n v_n)$, we have that $\alpha_{r+1} v_{r+1} + \cdots + \alpha_n v_n$ lies in $\ker(T)$. Consequently, there exist scalars $\alpha_1, \dots, \alpha_r$ such that

$$\alpha_{r+1} v_{r+1} + \cdots + \alpha_n v_n = \alpha_1 v_1 + \cdots + \alpha_r v_r.$$

By subtracting the right-hand side from the left-hand side, we obtain a relation of linear dependence $-\alpha_1 v_1 - \cdots - \alpha_r v_r + \alpha_{r+1} v_{r+1} + \cdots + \alpha_n v_n = O_V$. Considering that $v_1, \dots, v_r, v_{r+1}, \dots, v_n$ form a basis for V, they are linearly independent so that $\alpha_1 = \cdots = \alpha_r = \alpha_{r+1} = \cdots = \alpha_n = 0$. $\square$

Corollary 1.12.8. *If $T : V \to W$ is a linear transformation of finite-dimensional vector spaces V and W such that* $\dim(V) = \dim(W)$, *then the following statements are equivalent.*

1.) *We have that T is injective.*

2.) *We have that* $\operatorname{nullity}(T) = 0$.

3.) *We have that* $\operatorname{rank}(T) = \dim(W)$.

4.) *We have that T is surjective.*

Proof. We will assume first that T is injective. By Proposition 1.11.6 and the definition of nullity, we have that $\operatorname{nullity}(T) = 0$. By the Rank-Nullity Theorem, if $\operatorname{nullity}(T) = 0$, then we conclude that $\dim(W) = \dim(V) = \operatorname{rank}(T)$. Even more, if it holds that $\operatorname{rank}(T) = \dim(W)$, then $\operatorname{range}(T)$ is a subspace of W of the same dimension as W, hence we must have that $W = \operatorname{range}(T)$ by Propositions 1.11.10 and 1.8.9. Last, if T is surjective, then $\operatorname{range}(T) = W$ by definition, from which it follows that $\operatorname{rank}(T) = \dim(W) = \dim(V)$. By the Rank-Nullity Theorem, once again, we conclude that $\operatorname{nullity}(T) = 0$; this condition is equivalent to $\ker(T) = \{O_T\}$, i.e., T is injective. □

1.13 Composition and Inversion of Linear Transformations

Given any linear transformations $S : U \to V$ and $T : V \to W$ of vector spaces U, V, and W, we may define the **composite function** $T \circ S : U \to W$ by declaring that $(T \circ S)(u) = T(S(u))$ holds for all vectors $u \in U$, where $S(u)$ is by definition a vector of V; it is a linear transformation.

Proposition 1.13.1. *If $S : U \to V$ and $T : V \to W$ are linear transformations of vector spaces U, V, and W, then the composite function $T \circ S : U \to W$ is a linear transformation.*

Proof. We must establish that $(T \circ S)(\alpha u + v) = \alpha(T \circ S)(u) + (T \circ S)(v)$ for all vectors $u, v \in U$ and all scalars α. By definition, we have that $(T \circ S)(\alpha u + v) = T(S(\alpha u + v))$. Considering that $S : U \to V$ is a linear transformation, it follows by definition that $S(\alpha u + v) = S(\alpha u) + S(v) = \alpha S(u) + S(v)$. Consequently, we find that $(T \circ S)(\alpha u + v) = T(\alpha S(u) + S(v)$. By the linearity of T, we conclude that $(T \circ S)(\alpha u + v) = T(\alpha S(u)) + T(S(v)) = \alpha T(S(u)) + T(S(v)) = \alpha(T \circ S)(u) + (T \circ S)(v)$. □

Example 1.13.2. Consider the linear transformations $S : \mathbb{R}^{1\times 3} \to \mathbb{R}^{1\times 2}$ and $T : \mathbb{R}^{1\times 2} \to \mathbb{R}^{1\times 1}$ defined by $S(x, y, z) = (x, y)$ and $T(x, y) = (x)$. Put another way, S is the projection of a point in three-space into the xy-plane and T is the projection of a point in the Cartesian plane onto the x-axis. We have that $(T \circ S)(x, y, z) = T(S(x, y, z)) = T(x, y) = (x)$, hence $T \circ S$ can be viewed as the projection of a point in three-space onto the x-axis.

Example 1.13.3. Consider the differentiation transformation $D : P_3(x) \to P_2(x)$ from the real polynomials of degree at most three to the real polynomials of degree at most two that sends a polynomial to its first derivative. Explicitly, we have that $D(ax^3 + bx^2 + cx + d) = 3ax^2 + 2bx + c$. We know from Calculus I (or Example 1.10.5) that differentiation is a linear transformation because

$$\frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx}f(x) + \frac{d}{dx}g(x) \text{ and } \frac{d}{dx}[\alpha f(x)] = \alpha\frac{d}{dx}f(x)$$

for all real functions $f(x)$ and $g(x)$ and all real numbers α. Observe that

$$\begin{aligned}
(D \circ D)(ax^3 + bx^2 + cx + d) &= D(3ax^2 + 2bx + c) = 6ax + 2b, \\
(D \circ D \circ D)(ax^3 + bx^2 + cx + d) &= D(6ax + 2b) = 6a, \text{ and} \\
(D \circ D \circ D \circ D)(ax^3 + bx^2 + cx + d) &= D(6a) = 0,
\end{aligned}$$

hence $D \circ D$ yields the second derivative; $D \circ D \circ D$ yields the third derivative; and so on. Conventionally, we will use D^n to denote the composite function of D with itself n times. Using this notation, it follows that D^2 produces the second derivative; D^3 produces the third derivative; and so on. Considering that the $(n+1)$th derivative of a polynomial of degree n is always zero, it follows that $\ker(D^2) = \{ax + b \mid a, b \in \mathbb{R}\}$, $\ker(D^3) = \{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\}$, and so on.

By Example 1.10.7, we may also define the linear transformation $R_0 : P_2(x) \to P_3(x)$ such that $R_0(ax^2+bx+c) = \int_0^x (at^2+bt+c)\,dt$. Observe that the composite functions $R_0 \circ D : P_3(x) \to P_3(x)$ and $D \circ R_0 : P_2(x) \to P_2(x)$ are linear transformations that satisfy the following identities.

$$(R_0 \circ D)(ax^3 + bx^2 + cx + d) = R_0(3ax^2 + 2bx + c) = \int_0^x (3at^2 + 2bt + c)\,dt = ax^3 + bx^2 + cx$$

$$(D \circ R_0)(ax^2 + bx + c) = D\left(\frac{a}{3}x^3 + \frac{b}{2}x^2 + cx + d\right) = ax^2 + bx + c$$

Consequently, we have that $\ker(R_0 \circ D) = \{d \mid d \in \mathbb{R}\} = \operatorname{span}\{1\}$ and $\operatorname{range}(R_0 \circ D) = \{ax^3+bx^2+cx \mid a, b, c \in \mathbb{R}\} = \operatorname{span}\{x, x^2, x^3\}$ so that $4 = \dim(P_3(x)) = \operatorname{rank}(R_0 \circ D) + \operatorname{nullity}(R_0 \circ D)$. On the other hand, we have that $\ker(D \circ R_0) = \{0\}$ and $\operatorname{range}(D \circ R_0) = \{ax^2+bx+c \mid a, b, c \in \mathbb{R}\} = P_2(x)$.

Corollary 1.13.4. *Composition of linear transformations is not commutative in general. Explicitly, if $S : V \to W$ and $T : W \to V$ are linear transformations of vectors spaces V and W, then it is not necessarily true that $T \circ S : V \to V$ and $S \circ T : W \to W$ satisfy that $T \circ S = S \circ T$.*

Example 1.13.3 gives rise to four important notions in the theory of linear transformations. First, if $T : V \to V$ is a linear transformation from a vector space V to itself, then we will say that T is a **linear operator**. We will henceforth adopt the notation that if n is a positive integer, then T^n is the composite function of T with itself n times, e.g., $T^2 = T \circ T$ and $T^3 = T \circ T \circ T$. Observe that if U is a subspace of V, then the composite function T^n for a positive integer n is well-defined for any linear transformation $T : V \to U$ because the codomain U is a subset of the domain V. Last, we will denote by $I : V \to V$ the **identity operator** defined by $I(v) = v$ for all vectors $v \in V$. If $T : V \to W$ is a linear transformation of vector spaces V and W, then we say that $S : W \to V$ is a **left inverse** of T (or T is a **right inverse** of S) if $S \circ T : V \to V$ satisfies that $S \circ T = I$.

Proposition 1.13.5. *Let $T : V \to W$ be a linear transformation of vector spaces V and W.*

1.) *T admits a left inverse if and only if T is injective.*

2.) *T admits a right inverse if and only if T is surjective.*

Proof. We will assume first that T is injective. We must provide a linear transformation $S : W \to V$ such that $(S \circ T)(v) = v$ for every vector $v \in V$. By Proposition 1.11.17, it suffices to specify $S(w_i)$ for some basis vectors $w_i \in W$. We achieve this as follows. Begin with a basis $\mathscr{B}$ for V. By Proposition 1.11.9, the images $T(v_i)$ of the basis vectors $v_i \in \mathscr{B}$ form a collection $T(\mathscr{B})$ of linearly independent vectors of W. By Theorem 1.8.10, we may extend $T(\mathscr{B})$ to a basis for W. We define the linear transformation $S : W \to V$ by declaring that $S(T(v_i)) = v_i$ for all basis vectors $T(v_i)$ of W and $S(w_i) = O_V$ for all other basis vectors w_i of W. Crucially, observe that $(S \circ T)(v_i) = v_i$. Every element of V can be written uniquely as $\alpha_1 v_1 + \cdots + \alpha_n v_n$ for some scalars $\alpha_1, \ldots, \alpha_n$ and basis vectors $v_1, \ldots, v_n$ and $\alpha_1 v_1 + \cdots + \alpha_n v_n = \alpha_1 (S \circ T)(v_1) + \cdots + \alpha_n (S \circ T)(v_n) = (S \circ T)(\alpha_1 v_1 + \cdots + \alpha_n v_n)$.

Conversely, if T admits a left inverse $S : W \to V$, then for any vector $v \in \ker(T)$, we have that $v = I(v) = (S \circ T)(v) = S(T(v)) = S(O_W) = O_V$. We conclude that T is injective.

Likewise, if T is surjective, we construct a right inverse $S : W \to V$ in an analogous manner as the first paragraph above; we need only recognize that if T is surjective, then Proposition 1.11.15 and Theorem 1.8.10 imply that a basis $\mathscr{B}$ for V gives rise to a spanning set $T(\mathscr{B})$ for W that can be reduced to a basis for W. Consequently, define the linear transformation $S : W \to V$ by declaring that $S(T(v_i)) = v_i$ for all basis vector $T(v_i)$ of W. Every element of W can be written uniquely as $\alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)$ for some scalars $\alpha_1, \dots, \alpha_n$ and basis vectors $T(v_1), \dots, T(v_n)$ and

$$\alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) = \alpha_1 (T \circ S)(T(v_1)) + \cdots + \alpha_n (T \circ S)(T(v_n))$$

$$= (T \circ S)(\alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)).$$

Last, if we assume that T admits a right inverse $S : W \to V$, then for any vector $w \in W$, we have that $w = (T \circ S)(w) = T(S(w))$, hence there exists a vector $S(w) \in V$ such that $w = T(S(w))$. We conclude therefore that $\operatorname{range}(T) = W$, hence T is surjective. $\square$

We say that a linear transformation $T : V \to W$ admits a (two-sided) **inverse** transformation $S : W \to V$ if $S \circ T$ is the identity operator on V and $T \circ S$ is the identity operator on W.

Proposition 1.13.6. *Every left inverse of a linear transformation $T : V \to W$ of vector spaces V and W is a right inverse of T and vice-versa (provided that both a left inverse and a right inverse of T exist). Consequently, if T admits a two-sided inverse, then it is unique.*

Proof. Observe that if there exist linear transformations $L : W \to V$ and $R : W \to V$ satisfying that $L \circ T$ is the identity operator on V and $T \circ R$ is the identity operator on W, then it follows that $L(w) = L((T \circ R)(w)) = (L \circ T)(R(w)) = R(w)$ for all vectors $w \in W$. We conclude that $L = R$; the second statement follows because any two-sided inverse of T is both a left and right inverse. $\square$

Generally, a linear transformation $T : V \to W$ is **invertible** if it admits both a left inverse and a right inverse; the previous proposition implies that this two-sided inverse is unique, denoted by $T^{-1} : W \to V$. By definition, we have that $T^{-1} \circ T$ is the identity operator on V and $T \circ T^{-1}$ is the identity operator on W. We provide necessary and sufficient conditions for the existence of inverses.

Corollary 1.13.7. *If $T : V \to W$ is a linear transformation of vector spaces V and W, then T is invertible if and only if T is bijective, i.e., it is both injective and surjective. Even more, if V is finite-dimensional, then T is invertible if and only if T is injective if and only if T is surjective.*

Proof. By definition, T is invertible if and only if it admits a left inverse and a right inverse if and only if it is injective and surjective by Proposition 1.13.5. Consequently, if V is finite-dimensional, then by the Rank-Nullity Theorem, we have that T is injective if and only if T is surjective, so it suffices to prove that T is invertible if and only if T is injective. If T is invertible, then there exists a unique linear operator $T^{-1} : V \to W$ such that $T^{-1} \circ T = I$. Given any vector $v \in \ker(T)$, we have therefore that $v = I(v) = (T^{-1} \circ T)(v) = T^{-1}(T(v)) = T^{-1}(O_V) = O_V$, hence T is injective. Conversely, if T is injective, then by Proposition 1.13.5, it admits a left inverse; likewise, T admits a right inverse because it is surjective, hence it admits a two-sided inverse by Proposition 1.13.6. $\square$

Example 1.13.8. Consider the real vector space $F(\mathbb{R}, \mathbb{R})$ consisting of all functions $f : \mathbb{R} \to \mathbb{R}$. Given any real number c, we may define a linear transformation $T_c : F(\mathbb{R}, \mathbb{R}) \to F(\mathbb{R}, \mathbb{R})$ by declaring that $T_c(f) = cf$. Observe that if $c = 0$, then $T_c(f) = 0$ for all functions $f : \mathbb{R} \to \mathbb{R}$; however, if c is nonzero, then T is invertible. Explicitly, the linear transformation $T_{c^{-1}} : F(\mathbb{R}, \mathbb{R}) \to F(\mathbb{R}, \mathbb{R})$ satisfies that $(T_{c^{-1}} \circ T_c)(f) = T_{c^{-1}}(cf) = c^{-1}(cf) = f = c(c^{-1}f) = T_c(c^{-1}f) = (T_c \circ T_{c^{-1}})(f)$.

Example 1.13.9. Consider the real vector space $\mathbb{R}^{n \times r}$ of real $n \times r$ matrices. Given any invertible real $n \times n$ matrix A, the linear transformation $T_A : \mathbb{R}^{n \times r} \to \mathbb{R}^{n \times r}$ defined by $T_A(B) = AB$ is invertible. Explicitly, the linear transformation $T_{A^{-1}} : \mathbb{R}^{n \times r} \to \mathbb{R}^{n \times r}$ satisfies that

$$(T_{A^{-1}} \circ T_A)(B) = T_{A^{-1}}(AB) = A^{-1}(AB) = B = A(A^{-1}B) = T_A(A^{-1}B) = (T_A \circ T_{A^{-1}})(B).$$

Example 1.13.10. Consider the real vector space $\mathbb{R}[x]$ of real polynomials in indeterminate x. We may define a function $T_x : \mathbb{R}[x] \to \mathbb{R}[x]$ by $T_x(p(x)) = xp(x)$. Observe that T_x is a linear operator: indeed, it holds that $T_x(\alpha p(x) + q(x)) = x(\alpha p(x) + q(x)) = \alpha(xp(x)) + xq(x) = \alpha T_x(p(x)) + T_x(q(x))$ for all real numbers α and all real polynomials $p(x)$ and $q(x)$. Even more, T_x is injective: if $xp(x) = T_x(p(x)) = T_x(q(x)) = xq(x)$, then we may cancel x from both sides to find that $p(x) = q(x)$. On the other hand, T_x is not surjective because no constant polynomial can be written as $xp(x)$ for any polynomial $p(x)$. We conclude that $T_x : \mathbb{R}[x] \to \mathbb{R}[x]$ is not invertible.

Conversely, let us restrict our attention to the set $W = \{p(x) \mid p(0) = 0\}$ of real polynomials in indeterminate x whose constant term is 0. By the Three-Step Subspace Test, we find that W is a subspace of $\mathbb{R}[x]$. Even more, $T_x : \mathbb{R}[x] \to W$ is surjective because every polynomial with constant term 0 is divisible by x, i.e., if $p(0) = 0$, then there exists a polynomial $q(x)$ such that $p(x) = xq(x)$. By Proposition 1.13.5, it follows that T_x admits a right inverse $S_x : W \to \mathbb{R}[x]$. Explicitly, this linear transformation is defined by $S_x(p(x)) = q(x)$, where $q(x)$ is the polynomial satisfying $p(x) = xq(x)$. On the other hand, $\ker(T_x)$ is the infinite-dimensional vector space consisting of all polynomials that are divisible by x, hence T_x does not admit a left inverse by the same proposition as before.

Remark 1.13.11. Example 1.13.10 exhibits the important and often overlooked fact that a function (and hence a linear transformation) consists of a rule, a domain, and a codomain. Explicitly, if $T : V \to W$ is a linear transformation of vector spaces, the rule is T; the domain is V; and the codomain is W. Each of these three aspects of $T : V \to W$ determines its properties, i.e., none of the information in the definition of T is extraneous. Particularly, it is possible that $T : V \to W$ fails to be surjective; however, it is always true that $T : V \to \operatorname{range}(T)$ is surjective.

One of the primary motivations to study linear transformations of vector spaces is to classify distinct vector spaces up to **isomorphism**. We say that two vectors spaces V and W are **isomorphic** and we write $V \cong W$ if there exists a bijective linear transformation $T : V \to W$. Consequently, by Corollary 1.13.7, the isomorphisms between the vector spaces V and W are precisely the invertible linear transformations $T : V \to W$. Even more, if $T : V \to W$ is an isomorphism, then the inverse transformation $T^{-1} : W \to V$ is also an isomorphism because T is a two-sided inverse for T^{-1}.

Essentially, an isomorphism between the vector spaces V and W can be viewed as a unique relabelling of the vectors of W in terms of the vectors of V: indeed, if $T : V \to W$ is an isomorphism, then T is surjective, hence for every vector $w \in W$, there exists a vector $v \in V$ such that $w = T(v)$. Even more, T is injective, hence the vector v for which $w = T(v)$ is unique to w. Consequently, we may view the vector v for which $w = T(v)$ as the unique relabelling of w in terms of the vector v.

Theorem 1.13.12. *Every real vector space of dimension n is isomorphic to the vector space $\mathbb{R}^{1\times n}$ of real $1\times n$ matrices. Particularly, the real vector space $\mathbb{R}^n$ of real n-tuples is isomorphic to $\mathbb{R}^{1\times n}$.*

Proof. Let $E_1, \dots, E_n$ denote the standard basis vectors of $\mathbb{R}^{1\times n}$, i.e., suppose that E_i is the $1\times n$ matrix consisting of 1 in the ith column and zeros elsewhere. Given any real vector space V of dimension n, there exist linearly independent vectors $v_1, \dots, v_n$ that span V. By Proposition 1.11.17, we may define the **coordinatization** linear transformation $T : V \to \mathbb{R}^{1\times n}$ by declaring that $T(v_i) = E_i$. Given any real $1\times n$ matrix $\begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$ in $\mathbb{R}^{1\times n}$, we have that

$$\begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} = a_1E_1 + \cdots + a_nE_n = a_1T(v_1) + \cdots + a_nT(v_n) = T(a_1v_1 + \cdots + a_nv_n).$$

Consequently, the transformation T is surjective; it is injective by the Rank-Nullity Theorem. □

1.14 Matrix Representations of Linear Transformations

We conclude our chapter on matrices and vector spaces by bringing our discussion full circle. Explicitly, we will demonstrate that every $m\times n$ matrix can be represented (not necessarily uniquely) by a linear transformation $T : V \to W$ from a vector space V of dimension n to a vector space W of dimension m. Conversely, and more importantly, every linear transformation $T : V \to W$ from an n-dimensional vector space V to an m-dimensional vector space W can be represented (not necessarily uniquely) as an $m\times n$ matrix A. Consequently, to understand linear transformations between finite-dimensional vector spaces, it suffices to study matrices and vice-versa.

Consider a vector space V of dimension n for some non-negative integer n. Occasionally, it is possible to find a "canonical" ordered basis for V. We have already encountered this situation.

Example 1.14.1. Consider the real vector space $\mathbb{R}^{1\times 3}$ of real 1×3 matrices. By Example 1.7.11, the real 1×3 matrices $E_1 = (1,0,0)$, $E_2 = (0,1,0)$, and $E_3 = (0,0,1)$ form an ordered basis for $\mathbb{R}^{1\times 3}$. We refer to this ordered basis as the **standard basis** of $\mathbb{R}^{1\times 3}$ because we have that

$$(a_1, a_2, a_3) = (a_1, 0, 0) + (0, a_2, 0) + (0, 0, a_3) = a_1E_1 + a_2E_2 + a_3E_3,$$

hence it is clear that E_1, E_2, E_3 is the canonical choice for an ordered basis of $\mathbb{R}^{1\times 3}$.

Example 1.14.2. Consider the real vector space $\mathbb{R}^{1\times n}$ of real $1\times n$ matrices. Observe that the real $1\times n$ matrices $E_1, E_2, \dots, E_n$ for which E_i consists of 1 in the ith column and zeros elsewhere form the standard basis for $\mathbb{R}^{1\times n}$. Like before, for any real $1\times n$ matrix $(a_1, a_2, \dots, a_n)$, we have that

$$(a_1, a_2, \dots, a_n) = (a_1, 0, \dots, 0) + (0, a_2, \dots, 0) + \cdots + (0, 0, \dots, a_n) = a_1E_1 + a_2E_2 + \cdots + a_nE_n,$$

hence the basis $E_1, E_2, \dots, E_n$ is the canonical choice for an ordered basis of $\mathbb{R}^{1\times n}$.

Example 1.14.3. Consider the real vector space $\mathbb{R}^{2\times 2}$ of real 2×2 matrices. We note that

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \text{ and } E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

form the standard basis for $\mathbb{R}^{2\times 2}$. By definition, every element of $\mathbb{R}^{2\times 2}$ can be written uniquely as

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Even more, the coordinates this 2×2 matrix with respect to this ordered basis are (a, b, c, d).

Example 1.14.4. Consider the real vector space $\mathbb{R}^{m \times n}$ of real $m \times n$ matrices equipped with the usual basis of $m \times n$ matrices $E_{11}, E_{12}, \dots, E_{1n}, \dots, E_{mn}$ for which the (i, j)th component of E_{ij} is 1 and all other components are zero. Every element of $\mathbb{R}^{m \times n}$ can be written uniquely as

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = a_{11}E_{11} + a_{12}E_{12} + \cdots + a_{1n}E_{1n} + \cdots + a_{m1}E_{m1} + a_{m2}E_{m2} + \cdots + a_{mn}E_{mn}.$$

Consequently, the $m \times n$ matrices $E_{11}, E_{12}, \dots, E_{1n}, \dots, E_{mn}$ form the standard basis for $\mathbb{R}^{m \times n}$; the standard coordinates of the displayed $m \times n$ matrix are $(a_{11}, a_{12}, \dots, a_{1n}, \dots, a_{m1}, a_{m2}, \dots, a_{mn})$.

Example 1.14.5. Consider the real vector space $P_n(x)$ of real polynomials in indeterminate x of degree at most n. By definition, every element of $P_n(x)$ can be written uniquely as

$$a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

for some real numbers $a_0, a_1, a_2, \dots, a_n$. Consequently, the polynomials $1, x, x^2, \dots, x^n$ form the standard basis for the real vector space of polynomials of degree at most n. Observe that the coordinates of such a polynomial with respect to this ordered basis are $(a_0, a_1, a_2, \dots, a_n)$.

Our first order of business is to establish that for every real $m \times n$ matrix A, there exists a linear transformation $T : V \to W$ from a real vector space V of dimension n to a real vector space W of dimension m that behaves in the same way as A. Given any real $m \times n$ matrix A, we may define a linear transformation $T_A : \mathbb{R}^{n \times 1} \to \mathbb{R}^{m \times 1}$ by declaring that for any real $n \times 1$ matrix X, we have that $T_A(X) = AX$. Consequently, under this assignment, the linear transformation T_A has the effect of multiplying a real $n \times 1$ column vector X by the $m \times n$ matrix A to product an $m \times 1$ column vector AX. By Proposition 1.11.17, it holds that T_A is the unique linear transformation from $\mathbb{R}^{n \times 1}$ to $\mathbb{R}^{m \times 1}$ that represents A because T_A and A behave the same way with respect to a basis of $\mathbb{R}^{n \times 1}$.

Proposition 1.14.6. *Every real $m \times n$ matrix A can be represented (not necessarily uniquely) by the linear transformation $T_A : \mathbb{R}^{n \times 1} \to \mathbb{R}^{m \times 1}$ defined by $T_A(X) = AX$.*

Example 1.14.7. Be aware that it is possible to represent a real $m \times n$ matrix A by a different linear transformation than $T_A : \mathbb{R}^{n \times 1} \to \mathbb{R}^{m \times 1}$. Consider the following real 2×2 matrix.

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

Certainly, the matrix A is represented by the linear transformation $T_A : \mathbb{R}^{2 \times 1} \to \mathbb{R}^{2 \times 1}$ defined by $T_A(X) = AX$. Consider the real vector space $P_1(x)$ of real polynomials in indeterminate x of degree at most one. By Example 1.14.3, the standard basis of $P_1(x)$ is the ordered basis consisting of 1 and x. Every element of $P_1(x)$ can be written as $a + bx = a \cdot 1 + b \cdot x$, hence the coordinates of a real polynomial of degree at most one with respect to this ordered basis are (a, b). Consider the linear transformation $\frac{d}{dx} : P_1(x) \to P_1(x)$ defined by $\frac{d}{dx}(a + bx) = b$. Observe that the coordinates of $\frac{d}{dx}(a + bx)$ with respect to the standard basis of $P_1(x)$ are $(b, 0)$ because $b = b \cdot 1 + 0 \cdot x$. On the other hand, if we view the polynomial $a + bx$ with respect to its coordinates (a, b), then

$$\begin{bmatrix} b \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}.$$

Consequently, the linear transformation $\frac{d}{dx} : P_1(x) \to P_1(x)$ behaves in the same way as the matrix A with respect to the standard basis of the two-dimensional real vector space $P_1(x)$.

Our previous example is indicative of a more general phenomenon. Consider a linear transformation $T : V \to W$ from a vector space V of dimension n to a vector space W of dimension m. By Proposition 1.11.17, the linear transformation T is uniquely determined by its images on a basis of V. Explicitly, if $v_1, \dots, v_n$ form a basis for V, then the vectors $T(v_1), \dots, T(v_n)$ in W provide all of the information to determine $T(v)$ for any vector $v \in V$. Consequently, for each basis vector v_j, we may write $T(v_j) = a_{1j}w_1 + a_{2j}w_2 + \cdots + a_{mj}w_m$ for some basis $w_1, \dots, w_m$ of W as follows.

$$\begin{aligned} T(v_1) &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\ T(v_2) &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m \\ &\vdots \\ T(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m \end{aligned}$$

Collecting the coefficients of the vectors $T(v_j)$ with respect to the ordered basis vectors $w_1, \dots, w_n$ as the jth column of an $m \times n$ matrix, we obtain the following $m \times n$ matrix.

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

Observe that the coordinates of a vector $v \in V$ with respect to the ordered basis vectors $v_1, \dots, v_n$ of V are uniquely determined by the scalars $\alpha_1, \dots, \alpha_n$ such that $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$. Particularly, the **coordinate vector** of v_j of respect to these ordered basis vectors is simply the standard basis vector E_j of $\mathbb{R}^{1 \times n}$. Consequently, it follows that left-multiplication of each $n \times 1$ column vector E_j^t by A yields $AE_j^t = (a_{1j}, a_{2j}, \dots, a_{mj})$, i.e., the coordinate vector of $T(v_j)$ with respect to the ordered basis vectors $w_1, \dots, w_m$ of W. Unravelling these observations demonstrates that the matrix A acts on the coordinate vector $(\alpha_1, \dots, \alpha_n)^t$ of $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ as the linear transformation T acts on the vector v itself. Consequently, we refer to A as the **matrix representation** of the linear transformation T with respect to the ordered bases $v_1, \dots, v_n$ of V and $w_1, \dots, w_n$ of W.

Algorithm 1.14.8 (Matrix Representation Algorithm)**.** Given a linear transformation $T : V \to W$ between a vector space V of dimension n and a vector space W of dimension m and ordered bases $\mathscr{B}_V = \{v_1, \dots, v_n\}$ and $\mathscr{B}_W = \{w_1, \dots, w_m\}$ of V and W, respectively, use the following algorithm to find the matrix representation of T with respect to the ordered bases $\mathscr{B}_V$ and $\mathscr{B}_w$.

1.) Compute the vector $T(v_1)$ of W; then, find the unique coefficients $a_{11}, a_{21}, \dots, a_{m1}$ for which $T(v_1) = a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m$. Use the method of Gaussian Elimination, if necessary.

2.) Compute the vector $T(v_2)$ of W; then, find the unique coefficients $a_{12}, a_{22}, \dots, a_{m2}$ for which $T(v_2) = a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m$. Use the method of Gaussian Elimination, if necessary.

3.) Continue in this manner for each of the remaining basis vectors of V.

One will ultimately arrive at the $m \times n$ matrix

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

whose jth column consists of the unique coefficients of $T(v_j)$ with respect to the ordered basis $w_1, w_2, \dots, w_m$ of W for each integer $1 \leq i \leq n$; this is the matrix representation of the linear transformation $T : V \to W$ with respect to the ordered bases $\mathscr{B}_V$ and $\mathscr{B}_W$.

Example 1.14.9. Consider the function $T : \mathbb{R}^{1\times 3} \to \mathbb{R}^{1\times 3}$ defined by

$$T(x, y, z) = (x + 2y + 3z, 2x + 3y + 4z, 3x + 4y + 5z).$$

Each of the components of $T(x, y, z)$ is a linear function of x, y, and z, hence all together, T is a linear transformation. We will compute the matrix representation of T with respect to the standard basis $E_1 = (1, 0, 0)$, $E_2 = (0, 1, 0)$, and $E_3 = (0, 0, 1)$ of $\mathbb{R}^{1\times 3}$. We achieve this as follows.

$$\begin{aligned} T(E_1) &= T(1,0,0) = (1,2,3) = 1 \cdot E_1 + 2 \cdot E_2 + 3 \cdot E_3 \\ T(E_2) &= T(0,1,0) = (2,3,4) = 2 \cdot E_1 + 3 \cdot E_2 + 4 \cdot E_3 \\ T(E_3) &= T(0,0,1) = (3,4,5) = 3 \cdot E_1 + 4 \cdot E_2 + 5 \cdot E_3 \end{aligned}$$

Consequently, we obtain the matrix representation of T with respect to the standard basis of $\mathbb{R}^{1\times 3}$.

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

We can verify that this indeed behaves in the same way as the linear transformation T as follows.

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x + 2y + 3z \\ 2x + 3y + 4z \\ 3x + 4y + 5z \end{bmatrix}$$

Example 1.14.10. Consider the function $T : \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ defined by

$$T\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} c & d \\ a & b \end{bmatrix}.$$

One can readily verify that T is a linear transformation because applying T to a real 2×2 matrix B swaps the rows of B, so it preserves linear combinations of matrices. We will compute the matrix representation of T with respect to the standard basis E_{11}, E_{12}, E_{21}, and E_{22} of $\mathbb{R}^{2\times 2}$. We achieve this by expressing $T(E_{11})$, $T(E_{12})$, $T(E_{21})$, and $T(E_{22})$ in terms of the standard basis for $\mathbb{R}^{2\times 2}$.

$$T(E_{11}) = T\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = 0 \cdot E_{11} + 0 \cdot E_{12} + 1 \cdot E_{21} + 0 \cdot E_{22}$$

$$T(E_{12}) = T\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = 0 \cdot E_{11} + 0 \cdot E_{12} + 0 \cdot E_{21} + 1 \cdot E_{22}$$

$$T(E_{21}) = T\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = 1 \cdot E_{11} + 0 \cdot E_{12} + 0 \cdot E_{21} + 0 \cdot E_{22}$$

$$T(E_{22}) = T\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = 0 \cdot E_{11} + 1 \cdot E_{12} + 0 \cdot E_{21} + 0 \cdot E_{22}$$

Consequently, we obtain the matrix representation of T with respect to the standard basis of $\mathbb{R}^{2\times 2}$.

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

We can verify that this indeed behaves in the same way as the linear transformation T as follows.

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} c \\ b \\ a \\ d \end{bmatrix}$$

Even more, composition and inversion of linear transformations are compatible with matrix multiplication and matrix inversion of the matrix representations of linear transformations.

Proposition 1.14.11. *Let $T : V \to W$ be a linear transformation from an n-dimensional vector space V to an m-dimensional vector space W. Let A be the $m \times n$ matrix representation of T with respect to some ordered bases $\mathscr{B}_V$ of V and $\mathscr{B}_W$ of W, respectively.*

1.) *If $S : W \to V$ is a linear transformation and B is the $n \times m$ matrix representation of S with respect to the ordered bases $\mathscr{B}_W$ of W and $\mathscr{B}_V$ of V, then AB is the matrix representation of $T \circ S$ and BA is the matrix representation of $S \circ T$. Put another way, composition of linear transformation corresponds to matrix multiplication of the matrix representations.*

2.) *We have that T is invertible if and only if A is invertible. Even more, the inverse transformation $T^{-1} : W \to V$ of T is represented by the matrix inverse A^{-1} of A with respect to the specified ordered bases $\mathscr{B}_V$ of V and $\mathscr{B}_W$ of W, respectively.*

Proof. (1.) We will assume that $\mathscr{B}_V = \{v_1, \ldots, v_n\}$ and $\mathscr{B}_W = \{w_1, \ldots, w_m\}$. By definition of the matrix representation of T, the ith row of A consists of the scalars $a_{i1}, \ldots, a_{in}$ such that a_{ij} is the coefficient of w_i in the unique expression of $T(v_j)$ with respect to the basis vectors $w_1, \dots, w_m$. Likewise, the jth column of B consists of the scalars $b_{1j}, \dots, b_{nj}$ such that b_{ij} is the coefficient of v_i in the unique expression of $S(w_j)$ with respect to the basis vector $v_1, \dots, v_n$. By Definition 1.2.1, the (i, j)th component of the matrix product AB is given by $\sum_{k=1}^n a_{ik} b_{kj}$. Once we verify that this is indeed the coefficient of w_i in the unique expression of $(T \circ S)(w_j)$ with respect to the basis vectors $w_1, \dots, w_m$, our first claim will be established. By our previous work, we have that

$$\begin{aligned}
(T \circ S)(w_j) = T(S(w_j)) &= T(b_{1j} v_1 + \cdots + b_{nj} v_n) \\
&= b_{1j} T(v_1) + \cdots + b_{nj} T(v_n) \\
&= b_{1j}(a_{11} w_1 + \cdots + a_{m1} w_m) + \cdots + b_{nj}(a_{1n} w_1 + \cdots + a_{mn} w_m) \\
&= (a_{11} b_{1j} + \cdots + a_{1n} b_{nj}) w_1 + \cdots + (a_{m1} b_{1j} + \cdots + a_{mn} b_{nj}) w_m.
\end{aligned}$$

Consequently, the coefficient of w_i in the unique expression of $(T \circ S)(w_j)$ with respect to the basis vectors $w_1, \dots, w_m$ is $a_{i1} b_{1j} + \cdots + a_{in} b_{nj} = \sum_{k=1}^n a_{ik} b_{kj}$, as desired.

(2.) We note that T is invertible if and only if there exists a unique linear transformation $T^{-1} : W \to V$ such that $T \circ T^{-1}$ is the identity operator on W and $T^{-1} \circ T$ is the identity operator on V. Considering that the unique matrix representation of the identity operator on an m-dimensional vector space (with respect to any basis) is the $m \times m$ identity matrix, we conclude that if T is invertible, then the matrix representation B for T^{-1} with respect to the ordered basis $\mathscr{B}_W$ of W satisfies that $AB = I_{n \times n}$ and $BA = I_{n \times n}$, hence A is invertible. Conversely, if the matrix representation A of T with respect to the ordered basis $\mathscr{B}_V$ is invertible, then there exists an $n \times n$ matrix B such that $AB = I_{n \times n}$ and $BA = I_{n \times n}$. Consider the linear transformation $S : W \to V$ defined by $S(w_j) = b_{1j} v_1 + \cdots + b_{nj} v_n$ for each basis vector $w_1, \dots, w_n$ of W. We have that

$$\begin{aligned}
(T \circ S)(w_j) = T(S(w_j)) &= T(b_{1j} v_1 + \cdots + b_{nj} v_n) \\
&= b_{1j} T(v_1) + \cdots + b_{nj} T(v_n) \\
&= b_{1j}(a_{11} w_1 + \cdots + a_{m1} w_m) + \cdots + b_{nj}(a_{1n} w_1 + \cdots + a_{mn} w_m) \\
&= (a_{11} b_{1j} + \cdots + a_{1n} b_{nj}) w_1 + \cdots + (a_{m1} b_{1j} + \cdots + a_{mn} b_{nj}) w_m.
\end{aligned}$$

By the previous paragraph, the coefficient of w_i is equal to the (i, j)th component of the matrix product $AB = I_{n \times n}$, hence the coefficient of w_i is zero unless $i = j$, in which case it is one. We conclude therefore that $(T \circ S)(w_j) = w_j$, hence $T \circ S$ is the identity operator on W. By an analogous argument, it follows that $S \circ T$ is the identity operator on V, hence T is invertible. $\square$

Example 1.14.12. Consider the function $T : \mathbb{R}^{1 \times 2} \to \mathbb{R}^{1 \times 2}$ defined by $T(x, y) = (x + y, 2y)$. Each of the components of $T(x, y)$ is a linear function of x and y, hence T is a linear transformation. We compute the matrix representation of T with respect to the standard basis $(1, 0)$ and $(0, 1)$ of $\mathbb{R}^{1 \times 2}$.

Considering that $T(1,0) = (1,0) = 1 \cdot (1,0) + 0 \cdot (0,1)$ and $T(0,1) = (1,2) = 1 \cdot (1,0) + 2 \cdot (0,1)$, the matrix representation of T with respect to the standard basis of $\mathbb{R}^{1 \times 2}$ is as follows.

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

We note that A is a 2×2 matrix with two pivots, hence it is invertible. By Proposition 1.14.11, it follows that T is invertible. We compute the inverse A^{-1} of A and use it to construct T^{-1}.

$$\left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{array}\right] \overset{(1.)}{\sim} \left[\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & \frac{1}{2} \end{array}\right] \overset{(2.)}{\sim} \left[\begin{array}{cc|cc} 1 & 0 & 1 & -\frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{2} \end{array}\right]$$

(1.) We employed the elementary row operation $\frac{1}{2}R_2 \mapsto R_2$.

(2.) We employed the elementary row operation $R_1 - R_2 \mapsto R_2$.

Using the scalars belonging to the rows of A^{-1}, we construct T^{-1} as follows.

$$T^{-1}(x,y) = \left(1x + -\frac{1}{2}y, 0x + \frac{1}{2}y\right) = \left(x - \frac{1}{2}y, \frac{1}{2}y\right)$$

One can verify that $(T \circ T^{-1})(x,y) = (x,y)$ and that $(T^{-1} \circ T)(x,y) = (x,y)$ as follows.

$$(T \circ T^{-1})(x,y) = T\left(x - \frac{1}{2}y, \frac{1}{2}y\right) = \left(x - \frac{1}{2}y + \frac{1}{2}y, 2 \cdot \frac{1}{2}y\right) = (x,y)$$

$$(T^{-1} \circ T)(x,y) = T^{-1}(x+y, 2y) = \left(x + y - \frac{1}{2} \cdot 2y, \frac{1}{2} \cdot 2y\right) = (x,y)$$

Example 1.14.13. We adapt this example from [Str06, Problem 31] on page 152. Consider the linear transformation $T : \mathbb{R}^{2 \times 2} \to \mathbb{R}^{2 \times 2}$ defined by $T(B) = AB$ for the following real 2×2 matrix.

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

We compute the matrix representation of T with respect to the standard basis of $\mathbb{R}^{2 \times 2}$. We must first find the coordinates of $T(E_{11}), T(E_{12}), T(E_{21})$, and $T(E_{22})$ the standard basis of $\mathbb{R}^{2 \times 2}$.

$$T(E_{11}) = T\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 0 \end{bmatrix} = 1 \cdot E_{11} + 0 \cdot E_{12} + 3 \cdot E_{21} + 0 \cdot E_{22}$$

$$T(E_{12}) = T\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 3 \end{bmatrix} = 0 \cdot E_{11} + 1 \cdot E_{12} + 0 \cdot E_{21} + 3 \cdot E_{22}$$

$$T(E_{21}) = T\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 4 & 0 \end{bmatrix} = 2 \cdot E_{11} + 0 \cdot E_{12} + 4 \cdot E_{21} + 0 \cdot E_{22}$$

$$T(E_{22}) = T\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 0 & 4 \end{bmatrix} = 0 \cdot E_{11} + 2 \cdot E_{12} + 0 \cdot E_{21} + 4 \cdot E_{22}$$

We arrive at the matrix representation for T by forming the following 4×4 matrix.

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

Even though it is not immediately clear that the matrix R is invertible, we suspect that it is because A is an invertible matrix, hence T should be an invertible transformation. Explicitly, we have that $1 \cdot 4 - 2 \cdot 3 = 4 - 6 = -2$ is nonzero, hence A is invertible by Example 1.5.9. Either way, we may perform elementary row operations to convert R to its reduced row echelon form.

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 3 & 0 & 4 & 0 & 0 & 0 & 1 & 0 \\ 0 & 3 & 0 & 4 & 0 & 0 & 0 & 1 \end{array}\right] \overset{(1.)}{\sim} \left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 & -3 & 0 & 1 & 0 \\ 0 & 3 & 0 & 4 & 0 & 0 & 0 & 1 \end{array}\right] \overset{(2.)}{\sim} \left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 & -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & -2 & 0 & -3 & 0 & 1 \end{array}\right]$$

$$\overset{(3.)}{\sim} \left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 & -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & \frac{3}{2} & 0 & -\frac{1}{2} \end{array}\right]$$

$$\overset{(4.)}{\sim} \left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -2 & 0 & 1 \\ 0 & 0 & -2 & 0 & -3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & \frac{3}{2} & 0 & -\frac{1}{2} \end{array}\right]$$

$$\overset{(5.)}{\sim} \left[\begin{array}{cccc|cccc} 1 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -2 & 0 & 1 \\ 0 & 0 & 1 & 0 & \frac{3}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & \frac{3}{2} & 0 & -\frac{1}{2} \end{array}\right]$$

$$\overset{(6.)}{\sim} \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & -2 & 0 & 1 \\ 0 & 0 & 1 & 0 & \frac{3}{2} & 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 & 0 & \frac{3}{2} & 0 & -\frac{1}{2} \end{array}\right]$$

(1.) We employed the elementary row operation $R_3 - 3R_1 \mapsto R_3$.

(2.) We employed the elementary row operation $R_4 - 3R_2 \mapsto R_4$.

(3.) We employed the elementary row operation $-\frac{1}{2}R_4 \mapsto R_4$.

(4.) We employed the elementary row operation $R_2 - 2R_4 \mapsto R_2$.

(5.) We employed the elementary row operation $-\frac{1}{2}R_3 \mapsto R_3$.

(6.) We employed the elementary row operation $R_1 - 2R_3 \mapsto R_3$.

Consequently, we find that R is invertible, hence the linear transformation T that it represents is invertible. We compute T^{-1} by taking the rows of R^{-1} as the coefficients of E_{11}, E_{12}, E_{21}, and E_{22}.

$$T^{-1}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} -2a + c & -2b + d \\ \frac{3}{2}a - \frac{1}{2}c & \frac{3}{2}b - \frac{1}{2}d \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Consequently, we find that $T^{-1}(B) = CB$, where C is the following real 2×2 matrix.

$$C = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}$$

One can readily verify that $C = A^{-1}$, but this agrees with our intuition: because A is invertible, there exists a real 2×2 matrix A^{-1} such that $A^{-1}A = I_{2\times 2} = AA^{-1}$. Consequently, the linear transformation $S : \mathbb{R}^{2\times 2} \to \mathbb{R}^{2\times 2}$ defined by $S(B) = A^{-1}B$ satisfies that

$$(T \circ S)(B) = T(A^{-1}B) = A(A^{-1}B) = B = A^{-1}(AB) = S(AB) = (S \circ T)(B).$$

1.15 Chapter 1 Overview

This section is currently under construction.

Chapter 2

Canonical Forms for Matrices

We introduced in the first chapter the notion of matrices, their arithmetic, and numerous important properties of them. Essentially, the theory of matrices vastly simplifies the algebra of large sets of data. We demonstrated that the collection of all real $m \times n$ matrices forms an algebraic structure called a vector space; vector spaces are ubiquitous throughout mathematics, so it is critical to understand their properties. We defined functions (linear transformations) between vector spaces, and we studied certain vector spaces called the kernel and the range associated to a linear transformation. Ultimately, we established that linear transformations and matrices are intimately connected in a rigorous sense: explicitly, every linear transformation induces a matrix that is uniquely determined by specifying a basis for the domain and codomain spaces of the linear transformation. Consequently, we are motivated to return to further develop the theory of matrices in this chapter.

2.1 Determinants of $n \times n$ Matrices

Back in Example 1.10.6, we defined the **determinant** of a real 2×2 matrix as

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}.$$

Explicitly, the determinant of a real 2×2 matrix is a function $\det : \mathbb{R}^{2\times 2} \to \mathbb{R}$ that sends a real 2×2 matrix to the difference of the product $a_{11}a_{22}$ of its diagonal elements and the product $a_{12}a_{21}$ of its **anti-diagonal** elements. Generally, the determinant can be defined **recursively** for an $n \times n$ matrix for any positive integer n. We will not concern ourselves with determinants of matrices of size exceeding three, so it suffices to define the determinant of a real 3×3 matrix. Out of desire for notational convenience, we will seldom use the $\det(-)$ notation for a matrix whose components we wish to display explicitly; rather, we will denote the determinant using vertical bars as follows.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Under this identification, the determinant of a 3×3 matrix can be defined as follows.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{21} \\ a_{31} & a_{32} \end{vmatrix}$$

Explicitly, we take the product of the $(1, 1)$th component a_{11} of the matrix with the determinant of the 2×2 submatrix obtained by deleting row one and column one; then, we subtract from that the product of the $(1, 2)$th component a_{12} of the matrix with the determinant of the 2×2 submatrix obtained by deleting row one and column two; and we add to that the product of the $(1, 3)$th component of the matrix with the determinant of the 2×2 submatrix obtained by deleting row one and column three. Using the determinant of a 2×2 matrix, we obtain the following formula.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

One naturally wonders the purpose of defining the determinant of a 3×3 matrix by **expanding along** the first row, i.e., using the first row of the matrix as the coefficients of the determinants of the attendant 2×2 submatrices instead of using the second row or even some column of the matrix. Out of curiosity and for illustrative purposes, let us compute the determinant using the second row of the matrix. Essentially, we must rearrange the above displayed equation to obtain an alternating sum of $a_{21}(a_{12}a_{33} - a_{13}a_{32})$, $a_{22}(a_{11}a_{33} - a_{13}a_{31})$, and $a_{23}(a_{11}a_{32} - a_{12}a_{31})$; the differences are obtained as the determinants of the 2×2 submatrices obtained by deleting the second row and jth column for each integer $1 \leq j \leq 3$. By finding each of these terms in the above displayed equation and determining the appropriate signs, we obtain the following description of the determinant.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = -a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{22}(a_{11}a_{33} - a_{13}a_{31}) - a_{23}(a_{11}a_{32} - a_{12}a_{31})$$

Generally, we may define the determinant of an $n \times n$ matrix as follows.

Definition 2.1.1. Given any $n \times n$ matrix A, let A_{ij} denote the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the ith row and jth column of A. We define the **determinant** of A by

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Example 2.1.2. By the recursive definition of the determinant, we obtain the following.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = 1(5 \cdot 9 - 6 \cdot 8) - 2(4 \cdot 9 - 6 \cdot 7) + 3(4 \cdot 8 - 5 \cdot 7) = -3 - 2(-6) + 3(-3) = 0$$

Example 2.1.3. By the recursive definition of the determinant, we obtain the following.

$$\begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} = 1(0 \cdot 1 - 1 \cdot 1) - 1(1 \cdot 1 - 1 \cdot 0) + 0(1 \cdot 1 - 0 \cdot 0) = -1 - 1 + 0 = -2$$

Last chapter, we discussed the importance of the three elementary row operations for matrices. Explicitly, the method of Gaussian Elimination can be used to convert a real $m \times n$ matrix to its unique reduced row echelon form, from which many important properties of a matrix (e.g., rank and invertibility) can be deduced. Consequently, it is natural to consider the behavior of the determinant of a matrix with respect to the elementary row operations. We achieve this as follows.

Proposition 2.1.4. *Given any $n \times n$ matrix A and any scalar α, consider the $n \times n$ matrix B obtained from A by multiplying any row of A by α. We have that* $\det(B) = \alpha \det(A)$.

Proof. We will assume that B is obtained from A by multiplying the ith row of A by α. Consider the $(n-1) \times (n-1)$ matrix A_{ij} obtained from A by deleting the ith row and jth column of A. By hypothesis, we have that $b_{ij} = \alpha a_{ij}$ and $B_{ij} = A_{ij}$ for each integer $1 \leq j \leq n$. By Definition 2.1.1, we conclude that $\det(B) = \sum_{j=1}^n (-1)^{i+j} b_{ij} \det(B_{ij}) = \alpha \Big(\sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}) \Big) = \alpha \det(A)$. $\square$

Corollary 2.1.5. *Given any $n \times n$ matrix A with a zero row, we have that* $\det(A) = 0$.

Proof. We will assume that the ith row of A is zero. Considering that A is obtained from some $n \times n$ matrix B by multiplying the ith row of B by zero, we conclude that $\det(A) = 0 \det(B) = 0$. $\square$

Corollary 2.1.6. *Given any $n \times n$ matrix A and any scalar α, we have that* $\det(\alpha A) = \alpha^n \det(A)$.

Proof. By definition, the $n \times n$ matrix αA is obtained from the matrix A by scaling each of the n rows of A by α. Consequently, we have that $\det(\alpha A) = \alpha^n \det(A)$ by repeatedly factoring α. $\square$

Proposition 2.1.7. *Given any $n \times n$ matrices A and B that are equal except in one row, consider the $n \times n$ matrix C obtained from A and B by adding the two rows of A and B that are distinct and including all of the rows of A and B that are equal. We have that* $\det(C) = \det(A) + \det(B)$.

Proof. We will assume that the ith row of A is distinct from the ith row of B for some integer $1 \leq i \leq n$. By definition, the $n \times n$ matrix C satisfies that $c_{jk} = a_{jk} = b_{jk}$ for all integers $1 \leq j \leq n$ with $j \neq i$ and $c_{ik} = a_{ik} + b_{ik}$ for all integers $1 \leq k \leq n$. Consequently, the $(n-1) \times (n-1)$ matrix C_{ik} obtained from C by deleting the ith row and the kth column of C satisfies that $C_{ik} = A_{ik} = B_{ik}$ so that $\det(C_{ik}) = \det(A_{ik}) = \det(B_{ik})$ for all integers $1 \leq i \leq k$. We conclude the result as follows.

$$\begin{aligned}
\det(C) = \sum_{k=1}^n (-1)^{i+k} c_{ik} \det(C_{ik}) &= \sum_{k=1}^n (-1)^{i+k} (a_{ik} + b_{ik}) \det(C_{ik}) \\
&= \sum_{k=1}^n (-1)^{i+k} a_{ik} \det(C_{ik}) + \sum_{k=1}^n (-1)^{i+k} b_{ik} \det(C_{ik}) \\
&= \sum_{k=1}^n (-1)^{i+k} a_{ik} \det(A_{ik}) + \sum_{k=1}^n (-1)^{i+k} b_{ik} \det(B_{ik}) \\
&= \det(A) + \det(B)
\end{aligned}$$

$\square$

Proposition 2.1.8. *Given any $n \times n$ matrix A with two equal rows, we have that* $\det(A) = 0$.

Proof. We will proceed by induction on the integer $n \geq 2$. Certainly, if there are only two rows of A, then they must be equal to one another, hence the result holds in the case that $n = 2$ as follows.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{11} & a_{12} \end{vmatrix} = a_{11}a_{12} - a_{12}a_{11} = 0$$

Consequently, we may assume inductively that the result holds for some integer $n \geq 3$. We may assume that the ith row of A and the jth row of A are equal for some integers $1 \leq i < j \leq n$. Consider the $n \times n$ matrix $A_{k\ell}$ obtained from A by deleting the kth row and ℓth column of A for some integer $1 \leq k \leq n$ that is distinct from both i and j. We may find such an integer k by assumption that $n \geq 3$. Crucially, we note that the ith row of $A_{k\ell}$ and the jth row of $A_{k\ell}$ are equal for all integers $1 \leq \ell \leq n$, hence by induction, it follows that $\det(A_{k\ell}) = 0$ for all integers $1 \leq \ell \leq n$. By Definition 2.1.1, we conclude the desired result that $\det(A) = \sum_{\ell=1}^n (-1)^{k+\ell} a_{k\ell} \det(A_{k\ell}) = 0$. □

Proposition 2.1.9. *Given any $n \times n$ matrix A, any scalar α, and any integers $1 \leq i < j \leq n$, consider the $n \times n$ matrix B obtained from A by replacing the jth row of A with the sum of α times the ith row and the jth row of A. We have that* $\det(B) = \det(A)$. *Put another way, if we add any scalar multiple of a row of an $n \times n$ matrix to any other row, the determinant does not change.*

Proof. By definition of B, we have that $b_{k\ell} = a_{k\ell}$ for all integers $1 \leq k \leq n$ such that $k \neq j$ and $b_{j\ell} = \alpha a_{i\ell} + a_{j\ell}$ for all integers $1 \leq \ell \leq n$. Consider the $n \times n$ matrix C obtained from A by replacing the jth row of A with α times the ith row of A. Crucially, observe that B is obtained from A and C by including all common rows of A and C and taking the sum of the jth rows of A and C as the jth row of B. Consequently, by Proposition 2.1.7, we have that $\det(B) = \det(A) + \det(C)$. Consider the $n \times n$ matrix D obtained from A by replacing the jth row of A with the ith row of A. Explicitly, we note that C is obtained from D by multiplying the jth row of D by α. By Proposition 2.1.4, we have that $\det(C) = \alpha \det(D)$. Considering that the ith and jth rows of D are equal, it follows from Proposition 2.1.8 that $\det(D) = 0$ so that $\det(B) = \det(A) + \det(C) = \det(A) + \alpha \det(D) = \det(A)$. □

Corollary 2.1.10. *Given any $n \times n$ matrix A, if some row of A can be written as a linear combination of some other rows of A, then we have that* $\det(A) = 0$.

Proof. We will denote by A_i the ith row of A. Consider the case that $A_i = \alpha_1 A_{i_1} + \cdots + \alpha_k A_{i_k}$ for some integers $1 \leq i_1 < \cdots < i_k \leq n$ and some scalars $\alpha_1, \ldots, \alpha_k$. By rearranging the terms of the above identity, we find that $-\alpha_1 A_{i_1} - \cdots - \alpha_k A_{i_k} + A_i = O$. Consequently, we may subtract α_j times the i_jth row of A from the ith row of A for each integer $1 \leq j \leq k$ to reduce the ith row of A to zero. By Proposition 2.1.9, this process does not change the determinant of A; on the other hand, the determinant of the resulting matrix is zero by Corollary 2.1.5 so that $\det(A) = 0$. □

Proposition 2.1.11. *Given any $n \times n$ matrix A, consider the $n \times n$ matrix B obtained from A by interchanging any pair of rows of A. We have that* $\det(B) = -\det(A)$. *Put another way, swapping any pair of rows of an $n \times n$ matrix alters the sign of the determinant.*

Proof. Certainly, if any pair of rows of A are equal, then we have that $\det(B) = 0 = -0 = -\det(A)$. Consequently, we may assume that all rows of A are distinct. Crucially, we may obtain B from A by a sequence of operations that alter the determinant in exactly the manner claimed. Begin with the matrix C that is obtained from A by replacing the ith row of A with the sum of the ith and jth rows of A. By Propositions 2.1.7 and 2.1.8, it follows that $\det(C) = \det(A)$. Consider next the matrix D that is obtained from C by subtracting the ith row of C from the jth row of C so that the jth row of D is the ith row of A with the opposite sign. By Proposition 2.1.9, it follows that $\det(D) = \det(C) = \det(A)$. Last, we notice that B can be obtained from D by multiplying the jth row of D by -1; then, Proposition 2.1.4 yields that $\det(B) = -\det(D) = -\det(A)$. □

By the previous laundry list of properties of the determinant, we have fully described the behavior of the determinant with respect to the elementary row operations on matrices. We demonstrate next these properties also hold for the columns, and we summarize in the following corollary.

Proposition 2.1.12. *Given any $n \times n$ matrix A, we have that* $\det(A^t) = \det(A)$.

Proof. Unlike usual, we will prove the proposition only in the case that $n = 2$ or $n = 3$; the proof of the general case is beyond the scope of this class at the moment. Observe that the following hold.

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} = a_{11}a_{22} - a_{21}a_{12} = \begin{vmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{vmatrix}$$

Considering that the left-hand side is an arbitrary 2×2 matrix and the right-hand side is the transpose of this matrix, the result holds for $n = 2$. Likewise, the following identities hold.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

$$\begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{13}a_{22})$$

Once again, the result holds as soon as we recognize that the right-hand sides are equal. □

Corollary 2.1.13. *Given any $n \times n$ matrix A, the following properties hold.*

1.) *We may compute* $\det(A)$ *by expanding along any row of A.*

2.) *By multiplying any row of A by α, we multiply* $\det(A)$ *by α.*

3.) *By adding a scalar multiple of one row of A to another row, we do not change* $\det(A)$.

4.) *By swapping two rows of A, we change the sign of* $\det(A)$.

5.) *We have that* $\det(A) = 0$ *if any row of A is zero.*

6.) *We have that* $\det(A) = 0$ *if any pair of rows of A are equal.*

7.) *We have that* $\det(A) = 0$ *if any row of A is a linear combination of other rows of A.*

8.) *We have that* $\det(A) = \alpha \det(\mathrm{RREF}(A))$ *for some scalar α.*

Each of the above statements also holds if we use columns instead of rows.

Example 2.1.14. Consider the following real 3×3 matrix.

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

Considering that the second row of A is equal to twice the first row of A, it follows by Proposition 2.1.10 that $\det(A) = 0$. One could make a similar argument with the first and third rows of A.

Example 2.1.15. Consider the following real 3×3 matrix.

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

By employing the elementary row operations $R_2 - R_1 \mapsto R_2$ and $R_3 - R_1 \mapsto R_3$, according to Proposition 2.1.9, we do not alter $\det(A)$. Consequently, obtain the following 3×3 matrix.

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

By employing the elementary row operation $R_2 \leftrightarrow R_3$, we obtain the following 3×3 matrix.

$$C = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

By Example 2.1.3 and Proposition 2.1.11, we conclude the following.

$$\det(A) = -\det(C) = -\begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} = 2$$

Example 2.1.16. Consider the following real 3×3 matrix.

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

By employing the elementary column operation $C_1 \leftrightarrow C_3$, we obtain the 3×3 identity matrix. Consequently, by Corollary 2.1.13, we have that $\det(A) = -\det(I_{3\times 3})$. Last, observe the following.

$$\det(A) = -\det(I_{3\times 3}) = -\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -1\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = -1$$

2.2 The Adjugate of a Matrix

Every square matrix possesses a numerical invariant called a determinant. We will gradually come to understand throughout this chapter that the determinant of a matrix contains a wealth of information about the properties of a matrix, e.g., we have already seen that a matrix has determinant zero if it possesses a pair of linearly dependent rows or columns. Computing the determinant of a square matrix amounts to recursively expanding the matrix about some row or column by multiplying each subsequent entry a_{ij} of the specified row or column of the matrix by the determinant of the submatrix obtained by deleting the ith row and column jth column of the matrix.

One other way to obtain the determinant of an $n \times n$ matrix A is as the coefficient of the **scalar matrix** $\det(A)I$. We achieve this by taking the product of A with its **adjugate matrix** $\operatorname{adj}(A)$. We note that the adjugate matrix can also be encountered under the name of the **classical adjoint** (cf. [HK71, Exercise 5.2.3]); however, we will not adopt such terminology here because it is often associated with another object related to linear transformations. Like before, the adjugate matrix is defined recursively beginning with the case of 2×2 matrices as follows.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ and } \operatorname{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Explicitly, the adjugate matrix of any 2×2 matrix is obtained by swapping the elements on the main diagonal and changing the signs of the elements on the anti-diagonal. Observe the following.

$$\operatorname{adj}(A)A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} = \begin{bmatrix} \det(A) & 0 \\ 0 & \det(A) \end{bmatrix} = \det(A)I_{2\times 2}$$

Consequently, if $\det(A)$ is nonzero, then A is an invertible 2×2 matrix with $A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$.

We will soon verify that this rationale is much more general and applies to square matrices of all sizes. Before we are able to do this, we must define the adjugate of any $n \times n$ matrix.

Definition 2.2.1. Given any $n \times n$ matrix A, let A_{ij} denote the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the ith row and jth column of A. We refer to the (real) number $\mu_{ij} = \det(A_{ij})$ used in the definition of the determinant of A as the (i, j)th **minor** of the matrix A.

Definition 2.2.2. Given any $n \times n$ matrix A, let μ_{ij} denote the (i, j)th minor of A, i.e., μ_{ij} is the determinant of the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the ith row and jth column of A. We refer to the (real) number $\gamma_{ij} = (-1)^{i+j}\mu_{ij}$ as the (i, j)th **cofactor** of the matrix A.

Definition 2.2.3. Given any $n \times n$ matrix A, let γ_{ij} denote the (i, j)th cofactor of A, i.e., suppose that $\gamma_{ij} = (-1)^{i+j}\mu_{ij} = (-1)^{i+j}\det(A_{ij})$, where A_{ij} is the matrix obtained from A by deleting its ith row and jth column. We refer to the matrix $\Gamma = \begin{bmatrix}\gamma_{ij}\end{bmatrix}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$ as the **cofactor matrix** of A.

Definition 2.2.4. Given any $n \times n$ matrix A, let Γ denote the $n \times n$ cofactor matrix of A. We refer to the $n \times n$ matrix $\operatorname{adj}(A) = \Gamma^t$ as the **adjugate** (or adjugate matrix) of A.

One thing to notice is that the adjugate matrix can be defined for any square matrix over any ring because it only involves the operations of multiplication and subtraction; we will see that this provides a drastic improvement to the method of Gaussian Elimination we used previously to detect if a matrix is invertible. Explicitly, the process of Gaussian Elimination is only defined for matrices over fields because division is sometimes necessary to find the reduced row echelon form of a matrix.

Example 2.2.5. Let us compute the adjugate of the following real 3×3 matrix.

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

By Example 2.1.3, we have that $\det(A) = -2$. We will verify that $\operatorname{adj}(A)A = -2I_{3\times 3} = \det(A)I_{3\times 3}$. By Definition 2.2.4, we note that $\operatorname{adj}(A)$ is given by the transpose of the cofactor matrix Γ of A. By

Definition 2.2.3, the (i, j)th component of the cofactor matrix Γ is the (i, j)th cofactor γ_{ij} of A. By Definition 2.2.2, the cofactors of A are the signed 2×2 minors μ_{ij} of A. Ultimately, we must begin by finding the 2×2 minors μ_{ij} of A. Considering that A is a 3×3 matrix, there are $9 = 3 \cdot 3$ minors. By Definition 2.2.1, each minor μ_{ij} is given by the determinant of the 2×2 matrix A_{ij} obtained from A by deleting its ith row and jth column. Consequently, we find the following minors.

$$\mu_{11} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1 \qquad \mu_{21} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 \qquad \mu_{31} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

$$\mu_{12} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \qquad \mu_{22} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \qquad \mu_{32} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

$$\mu_{13} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \qquad \mu_{23} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \qquad \mu_{33} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

Continuing from this point, we find the $9 = 3 \cdot 3$ cofactors $\gamma_{ij} = (-1)^{i+j} \mu_{ij}$.

$$\gamma_{11} = (-1)^{1+1} \mu_{11} = -1 \qquad \gamma_{21} = (-1)^{2+1} \mu_{21} = -1 \qquad \gamma_{31} = (-1)^{3+1} \mu_{31} = 1$$

$$\gamma_{12} = (-1)^{1+2} \mu_{12} = -1 \qquad \gamma_{22} = (-1)^{2+2} \mu_{22} = 1 \qquad \gamma_{32} = (-1)^{3+2} \mu_{32} = -1$$

$$\gamma_{13} = (-1)^{1+3} \mu_{13} = 1 \qquad \gamma_{23} = (-1)^{2+3} \mu_{23} = -1 \qquad \gamma_{33} = (-1)^{3+3} \mu_{33} = -1$$

We are now in position to form the 3×3 cofactor matrix Γ as follows.

$$\Gamma = \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$$

Observe that in this case, Γ is a symmetric matrix because each row of Γ is equal to the corresponding column of Γ. Consequently, we have that $\operatorname{adj}(A) = \Gamma^t = \Gamma$. Even more, the following holds.

$$\operatorname{adj}(A)A = \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} = -2I = \det(A)I$$

Observe that if we divide both sides by $\det(A) = -2$, then we find the following.

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A) = -\frac{1}{2} \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Example 2.2.6. Let us compute the adjugate of the following real 3×3 matrix.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

By Definition 2.2.4, we have that $\operatorname{adj}(A)$ is equal to the transpose of the cofactor matrix Γ of A. By Definition 2.2.3, we construct the cofactor matrix Γ by finding each of the cofactors γ_{ij} of A. By Definition 2.2.2, the cofactors of A are the signed 2×2 minors μ_{ij} of A. Ultimately, we must begin by finding the 2×2 minors μ_{ij} of A. Considering that A is a 3×3 matrix, there are $9 = 3 \cdot 3$ minors. By Definition 2.2.1, each minor μ_{ij} is given by the determinant of the 2×2 matrix A_{ij} obtained from A by deleting its ith row and jth column. Consequently, we find the following minors.

$$\mu_{11} = \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} = -3 \qquad \mu_{21} = \begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} = -6 \qquad \mu_{31} = \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} = -3$$

$$\mu_{12} = \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = -6 \qquad \mu_{22} = \begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} = -12 \qquad \mu_{32} = \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} = -6$$

$$\mu_{13} = \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = -3 \qquad \mu_{23} = \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} = -6 \qquad \mu_{33} = \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} = -3$$

Continuing from this point, we find the $9 = 3 \cdot 3$ cofactors $\gamma_{ij} = (-1)^{i+j} \mu_{ij}$.

$$\gamma_{11} = (-1)^{1+1} \mu_{11} = -3 \qquad \gamma_{21} = (-1)^{2+1} \mu_{21} = 6 \qquad \gamma_{31} = (-1)^{3+1} \mu_{31} = -3$$

$$\gamma_{12} = (-1)^{1+2} \mu_{12} = 6 \qquad \gamma_{22} = (-1)^{2+2} \mu_{22} = -12 \qquad \gamma_{32} = (-1)^{3+2} \mu_{32} = 6$$

$$\gamma_{13} = (-1)^{1+3} \mu_{13} = -3 \qquad \gamma_{23} = (-1)^{2+3} \mu_{23} = 6 \qquad \gamma_{33} = (-1)^{3+3} \mu_{33} = -3$$

We are now in position to form the 3×3 cofactor matrix Γ as follows.

$$\Gamma = \begin{bmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{bmatrix}$$

Observe that in this case, Γ is a symmetric matrix because each row of Γ is equal to the corresponding column of Γ. Consequently, we have that $\operatorname{adj}(A) = \Gamma^t = \Gamma$. Even more, the following holds.

$$\operatorname{adj}(A)A = \begin{bmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = O_{3\times 3} = 0 I_{3 \times 3} = \det(A) I_{3 \times 3}$$

We will demonstrate next that the observations and patterns that have held across our examples are indicative of a general relationship between a square matrix and its adjugate.

Proposition 2.2.7. *Given any $n \times n$ matrix A, we have that* $\operatorname{adj}(A)A = \det(A) I_{n \times n}$.

Proof. By Definition 2.2.4, we have that $\operatorname{adj}(A) = \Gamma^t$, where Γ is the cofactor matrix of A. By Definition 2.2.3, the (i,j)th component of Γ is the (i,j)th cofactor γ_{ij} of A. By Definition 2.2.2, it follows that $\gamma_{ij} = (-1)^{i+j} \det(A_{ij})$, where A_{ij} is the $(n-1) \times (n-1)$ submatrix of A obtained from A

by deleting the ith row and jth column of A. Consequently, the (i,j)th component of $\operatorname{adj}(A)$ is the (j,i)th component of Γ, i.e., the (j,i)th cofactor $\gamma_{ji} = (-1)^{i+j}\det(A_{ji})$ of A. By Definition 1.2.1, we note that the (i,j)th component of $\operatorname{adj}(A)A$ is the sum of the products of the (i,k)th component of $\operatorname{adj}(A)$ and the (k,j)th component of A for each integer $1 \le k \le n$, i.e., the (i,j)th component of $\operatorname{adj}(A)A$ is $\sum_{k=1}^n (-1)^{i+k} a_{kj}\det(A_{ki})$. By Definition 2.1.1, we conclude that the (i,i)th components of $\operatorname{adj}(A)A$ are exactly $\det(A)$ because these are obtained from the aforementioned sum by setting $i = j$. Consequently, it suffices to prove that $\sum_{k=1}^n (-1)^{i+k} a_{kj}\det(A_{ki}) = 0$ whenever $i \neq j$.

Consider the $n \times n$ matrix B obtained from A by replacing the ith column of A with the jth column of A. Observe that for each integer $1 \le k \le n$, we have that $b_{ki} = a_{kj}$ because the ith column of B is equal to the jth column of A. Even more, we have that $B_{ki} = A_{ki}$ for all integers $1 \le k \le n$ because A and B only differ in the ith column. By Corollary 2.1.13, we have that

$$0 = \det(B) = \sum_{k=1}^n (-1)^{i+k} b_{ki}\det(B_{ki}) = \sum_{k=1}^n (-1)^{i+k} a_{kj}\det(A_{ki}).$$

We conclude therefore that the non-diagonal components of $\operatorname{adj}(A)A$ are zero, as desired. □

Proposition 2.2.8. *Given any $n \times n$ matrix A, we have that $\operatorname{adj}(A^t) = \operatorname{adj}(A)^t$. Put another way, the adjugate of the transpose is the transpose of the adjugate.*

Proof. Crucially, observe that deleting the ith row and jth column of A^t is the same as deleting the ith column and jth row of A and taking its transpose because the ith row of A^t is the ith column of A and the jth column of A^t is the jth row of A. Consequently, we have that $(A^t)_{ij} = (A_{ji})^t$. By the underlying definitions of the adjugate, the (i,j)th component of $\operatorname{adj}(A^t)$ is $(-1)^{i+j}\det((A^t)_{ij})$, hence by our opening remarks, the (i,j)th component of $\operatorname{adj}(A^t)$ is exactly $(-1)^{i+j}\det((A_{ji})^t)$. By Proposition 2.1.12, it follows that the (i,j)th component of $\operatorname{adj}(A^t)$ is $(-1)^{i+j}\det(A_{ji})$. Considering that this is the (j,i)th component of $\operatorname{adj}(A)$ by definition, we conclude that the (i,j)th component of $\operatorname{adj}(A^t)$ is the (i,j)th component of $\operatorname{adj}(A)^t$, hence the two matrices in consideration are equal. □

Corollary 2.2.9. *Given any $n \times n$ matrix A, we have that $A\operatorname{adj}(A) = \det(A)I_{n\times n}$.*

Proof. By Proposition 2.2.7, we have that $\operatorname{adj}(A^t)A^t = \det(A^t)I_{n\times n}$. By Proposition 2.1.12, we have that $\det(A^t) = \det(A)$ so that $\operatorname{adj}(A^t)A^t = \det(A)I_{n\times n}$. By Proposition 2.2.8, we have that $\operatorname{adj}(A^t) = \operatorname{adj}(A)^t$ so that $\operatorname{adj}(A)^t A^t = \det(A)I_{n\times n}$. Last, by Proposition 1.2.8, we conclude that

$$\det(A)I_{n\times n} = \det(A)I_{n\times n}^t = (\det(A)I_{n\times n})^t = (\operatorname{adj}(A)^t A^t)^t = (A^t)^t(\operatorname{adj}(A)^t)^t = A\operatorname{adj}(A). \qquad \square$$

Theorem 2.2.10. *Given any $n \times n$ matrix A, we have that A is invertible if and only if $\det(A) \neq 0$.*

Proof. Certainly, if the determinant of A is nonzero, then Propositions 2.2.7 and 2.2.9 imply that

$$\left(\frac{1}{\det(A)}\operatorname{adj}(A)\right)A = I_{n\times n} = A\left(\frac{1}{\det(A)}\operatorname{adj}(A)\right)$$

and $A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A)$. Conversely, if $\det(A) = 0$, then $\operatorname{adj}(A)A = \det(A)I_{n\times n} = O_{n\times n}$. Consequently, there is no $n \times n$ matrix B such that $AB = I_{n\times n} = BA$, i.e., A is not invertible. □

Example 2.2.11. By Example 2.1.2, the following 3×3 matrix is not invertible.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Example 2.2.12. By Example 2.1.3, the following 3×3 matrix is invertible.

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

Example 2.2.13. By Example 2.1.14, the following 3×3 matrix is not invertible.

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

We could have also noticed that A is row equivalent to a matrix with a zero row.

Example 2.2.14. By Example 2.1.15, the following 3×3 matrix is invertible.

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

Example 2.2.15. By Example 2.1.16, the following 3×3 matrix is invertible.

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

We could have also noticed that it is row equivalent to the 3×3 identity matrix.

Before we conclude this section, we state a critically important property of determinants.

Theorem 2.2.16. *Given any $n \times n$ matrices A and B, we have that* $\det(AB) = \det(A)\det(B)$.

Proof. Consider the unique reduced row echelon form $R = \operatorname{RREF}(A)$ for A. By Corollary 2.1.13, there exists a scalar α that is uniquely determined by the elementary row operations $E_1, \dots, E_k$ that are used to convert R to A such that $\det(A) = \alpha \det(R)$ and $E_k \cdots E_1 R = A$. Either R has a row consisting of zeros, or it is the $n \times n$ identity matrix. By the aforementioned corollary, if R has a row consisting of zeros, then $\det(R) = 0$ so that $\det(A) = \alpha \det(R) = 0$ and $\det(A)\det(B) = 0$. By Theorem 2.2.10, we have that $\det(AB)$ is nonzero if and only if AB is invertible if and only if RB is invertible. By assumption that R has a row consisting of zeros, it follows that RB is not invertible because it has a column consisting of zeros, and we conclude that $\det(AB) = 0$. Conversely, if R is the $n \times n$ identity matrix, then $\det(A) = \alpha \det(R) = \alpha$ and $A = E_k \cdots E_1 R = E_k \cdots E_1$, from which we conclude that $\det(A)\det(B) = \alpha \det(B) = \det(E_k \cdots E_1 B) = \det(E_k \cdots E_1 RB) = \det(AB)$. $\square$

2.3 Polynomials Associated to Matrices

We introduce in this section two polynomial invariants of an $n \times n$ matrix. Both of these polynomials are related to the determinant of a matrix associated with the given square matrix. Explicitly, suppose that A is any $n \times n$ matrix. We will adopt the shorthand I for the $n \times n$ identity matrix. Given any indeterminate x, we refer to the matrix $xI - A$ as the **characteristic matrix** of A. Both A and I are by assumption $n \times n$ matrices, hence the characteristic matrix $xI - A$ is likewise an $n \times n$ matrix. Even more, we note that diagonal of $xI - A$ consists of $x - a_{ii}$ for each integer $1 \leq i \leq n$ and the off-diagonal components of $xI - A$ are the off-diagonal components of A with the opposite sign. Explicitly, we have that $xI - A = \begin{bmatrix} x\delta_{ij} - a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$ for the Kronecker delta δ_{ij}.

Example 2.3.1. Consider the following 2×2 matrix A and its characteristic matrix $xI - A$.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad xI - A = \begin{bmatrix} x-1 & -2 \\ -2 & x-1 \end{bmatrix}$$

We note that $\det(xI - A) = (x-1)(x-1) - (-2)(-2) = x^2 - 2x - 3 = (x-3)(x+1)$.

Example 2.3.2. Consider the following 3×3 matrix A and its characteristic matrix $xI - A$.

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \qquad xI - A = \begin{bmatrix} x-1 & -1 & 0 \\ -1 & x & -1 \\ 0 & -1 & x-1 \end{bmatrix}$$

We note that $\det(xI - A) = (x-1)[x(x-1) - (-1)(-1)] - (-1)[(-1)(x-1) - (-1)(0)]$. By simplifying this, we obtain that $\det(xI - A) = (x-1)(x^2 - x - 1) - (x-1)$, hence factoring by grouping yields that $\det(xI - A) = (x-1)(x^2 - x - 1 - 1) = (x-1)(x^2 - x - 2) = (x-1)(x-2)(x+1)$.

Considering that we may always expand the determinant of the $n \times n$ characteristic matrix $xI - A$ along the first row, it follows that $\chi_A(x) = \det(xI - A)$ must be a polynomial in indeterminate x of degree n because the product of the diagonal elements of $xI - A$ form a polynomial in indeterminate x of degree n. (Concretely, one can prove this by induction.) Consequently, we refer to the determinant $\det(xI - A)$ of the characteristic matrix of A as the **characteristic polynomial** of A. One of the first observations that we can make regarding the characteristic polynomial is the following.

Proposition 2.3.3. *Given any $n \times n$ matrix A with characteristic polynomial $\chi(x)$, we have that $\det(A) = (-1)^n \chi(0)$. Put another way, the constant term of $\chi(x)$ is $(-1)^n \det(A)$.*

Proof. By definition of the characteristic polynomial, we have that $\chi(0) = \det(0I - A) = \det(-A)$. Consequently, by Proposition 2.1.6, it follows that $\chi(0) = (-1)^n \det(A)$, hence the result can be obtained by multiplying both sides of this identity by $(-1)^n$ and using the fact that $(-1)^{2n} = 1$. □

Example 2.3.4. Given any 2×2 matrix A with characteristic polynomial $\chi(x) = x^2 - 2x + 1$, we must have that $\det(A) = (-1)^2(0^2 - 2(0) + 1) = 1$.

Example 2.3.5. Given any 3×3 matrix A with characteristic polynomial $\chi(x) = x^3 - ex^2 + \pi$, we must have that $\det(A) = (-1)^3(0^3 - e(0)^2 + \pi) = -\pi$.

Given any polynomial $p(x) = c_k x^k + \cdots + c_1 x + c_0$, we can "plug in" any $n \times n$ matrix A to the polynomial $p(x)$ to obtain a **matrix polynomial** $p(A) = c_k A^k + \cdots + c_1 A + c_0 I$. Explicitly, the matrices A^i for each integer $1 \leq i \leq k$ are given by the i-fold product of the matrix A with itself, and the constant term c_0 of $p(x)$ becomes the scalar matrix $c_0 I$ in the matrix polynomial $p(A)$.

Example 2.3.6. Consider the 2×2 matrix A from Example 2.3.1. Observe that the following hold.

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

$$A + I = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

$$(A - 3I)(A + I) = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Consequently, the matrix polynomial $\chi(A) = (A - 3I)(A + I)$ yields the 2×2 zero matrix.

Example 2.3.7. Consider the 3×3 matrix A from Example 2.3.2. Observe that the following hold.

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

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

$$A + I = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

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

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

Consequently, the matrix polynomial $\chi(A) = (A - I)(A - 2I)(A + I)$ yields the 3×3 zero matrix.

Our next theorem demonstrates that these examples are indicative of a general phenomenon.

Theorem 2.3.8 (Cayley-Hamilton Theorem). *Given any $n \times n$ matrix A with characteristic polynomial $\chi(x)$, it holds that $\chi(A) = O$, i.e., the characteristic polynomial of A annihilates A.*

Proof. Because we have the adjugate matrix at our disposal from our discussion in the previous section, we will incorporate it into this proof; however, there are a wealth of interesting proofs of this fact that the interested reader is encouraged to discover. Considering that the characteristic matrix $xI - A$ of A is an $n \times n$ matrix whose coefficients lie in a polynomial ring, it admits an adjugate matrix $\operatorname{adj}(xI - A)$ such that $\operatorname{adj}(xI - A)(xI - A) = \det(xI - A)I = \chi(x)I$ by Proposition 2.2.7 and the definition of the characteristic polynomial $\chi(x)$. On the other hand, the components of the $n \times n$ matrices $xI - A$, $\operatorname{adj}(xI - A)$, and $\chi(x)I$ are polynomials in indeterminate x, hence these matrices can be written uniquely as formal polynomials with matrix coefficients: we must simply determine the part of the matrices corresponding to each monomial x^i for each integer $0 \leq i \leq n$. Explicitly, the characteristic matrix $xI - A$ is already written as a formal polynomial with matrix coefficients: indeed, the degree-one "coefficient" is the identity matrix I, and the "constant term" is the matrix A. Even more, if we write $\chi(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ for some scalars $c_{n-1}, \dots, c_1, c_0$, then the unique expression of $\chi(x)I$ as a formal polynomial with matrix coefficients is $\chi(x)I = x^n I + c_{n-1}x^{n-1}I + \cdots + c_1 xI + c_0 I$. Consider the unique $n \times n$ matrices $B_{n-1}, \dots, B_1, B_0$ such that $\operatorname{adj}(xI - A) = x^{n-1}B_{n-1} + \cdots + xB_1 + B_0$. Expanding the left- and right-hand sides of the identity $\operatorname{adj}(xI - A)(xI - A) = \chi(x)I$ according to our formal polynomial factorizations, we find that $(x^{n-1}B_{n-1} + \cdots + xB_1 + B_0)(xI - A) = x^n I + c_{n-1}x^{n-1}I + \cdots + c_1 xI + c_0 I$. Expanding the product on the left-hand side and comparing the terms with x^i, we obtain the following.

$$\begin{aligned} B_{n-1} &= I && \text{(the coefficient of } x^n\text{)} \\ B_{n-2} - B_{n-1}A &= c_{n-1}I && \text{(the coefficient of } x^{n-1}\text{)} \\ &\vdots \\ B_0 - B_1 A &= c_1 I && \text{(the coefficient of } x\text{)} \\ -B_0 A &= c_0 I && \text{(the constant term)} \end{aligned}$$

Crucially, we may now multiply each subsequent identity from bottom to top by A^i for the integer $0 \leq i \leq n$ corresponding to the monomial x^i to find the following identities.

$$\begin{aligned} B_{n-1}A^n &= A^n && \text{(the coefficient of } x^n\text{)} \\ B_{n-2}A^{n-1} - B_{n-1}A^n &= c_{n-1}A^{n-1} && \text{(the coefficient of } x^{n-1}\text{)} \\ &\vdots \\ B_0 A - B_1 A^2 &= c_1 A && \text{(the coefficient of } x\text{)} \\ -B_0 A &= c_0 I && \text{(the constant term)} \end{aligned}$$

Last, summing the left-hand column yields a telescoping sum that results in the zero matrix; however, the right-hand sums to the $n \times n$ matrix $A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = \chi(A)$. $\square$

One immediate consequence of the Cayley-Hamilton Theorem is that for every $n \times n$ matrix A, there exists a unique **monic** polynomial $\mu_A(x)$ of least degree such that $\mu_A(A) = O$. We refer to this polynomial as the **minimal polynomial** of A. Explicitly, a monic polynomial is one whose leading coefficient is one. By the Cayley-Hamilton Theorem, the characteristic polynomial $\chi_A(x)$ of A is a monic polynomial satisfying that $\chi_A(A) = O$, hence there exists a monic polynomial with the desired property. Consequently, we can find a monic polynomial of least degree that annihilates A by the **Well-Ordering Principle** applied to the nonempty set of positive integers corresponding to the degree of monic polynomials that annihilate A. Even more, the uniqueness of the minimal polynomial comes from the fact that if we take two monic polynomials of least degree that both annihilate A, then each of the polynomials will divide the other, hence they must be equal.

Even if this line of argument is not immediately clear, what matters is the following.

Proposition 2.3.9. *Given any $n \times n$ matrix A, its minimal polynomial $\mu(x)$ divides every polynomial $p(x)$ such that $p(A) = O$. Consequently, the minimal polynomial of A must divide the characteristic polynomial of A, so it is either the characteristic polynomial of A or a proper factor of it.*

Proof. By the Division Algorithm for polynomials, there exist unique polynomials $q(x)$ and $r(x)$ such that $p(x) = q(x)\mu(x) + r(x)$ and the degree of $r(x)$ is strictly smaller than the degree of $\mu(x)$. By assumption, we have that $p(A) = O$. By definition of $\mu(x)$, we have that $\mu(A) = O$. Combined, these observations imply that $O = p(A) = q(A)\mu(A) + r(A) = q(A)O + r(A) = r(A)$. Consequently, we have found a polynomial $r(x)$ of lesser degree than $\mu(x)$ that annihilates A. Even more, if $r(x)$ is nonzero, then we may multiply by the multiplicative inverse of its leading coefficient to obtain a monic polynomial of lesser degree than $\mu(x)$ that annihilates A. Because this is impossible by the definition of $\mu(x)$, we conclude that $r(x)$ must be the zero polynomial so that $\mu(x)$ divides $p(x)$.

By the Cayley-Hamilton Theorem, the characteristic polynomial of A annihilates A, so it must be divisible by the minimal polynomial of A by the argument of the preceding paragraph. $\square$

Example 2.3.10. Consider the 2×2 matrix A from Examples 2.3.1 and 2.3.6. We proved previously that the characteristic polynomial of A is $\chi(x) = (x-3)(x+1)$; neither the polynomial $x-3$ nor $x+1$ annihilates A by the previous example, hence we conclude by Proposition 2.3.9 that $\mu(x) = \chi(x)$.

Example 2.3.11. Consider the 3×3 matrix A from Examples 2.3.2 and 2.3.7. We proved previously that the characteristic polynomial of A is $\chi(x) = (x-1)(x-2)(x+1)$. Observe that none of the linear polynomials $x-1$ or $x-2$ or $x+1$ annihilate A by the previous example. Even more, the quadratic polynomials $(x-1)(x-2)$ and $(x-1)(x+1)$ and $(x-2)(x+1)$ do not annihilate A. Consequently, we conclude by Proposition 2.3.9 that $\mu(x) = \chi(x)$.

Example 2.3.12. Consider the 3×3 zero matrix O. Observe that the characteristic polynomial of O is given by $\chi(x) = \det(xI - O) = \det(xI) = x^3 \det(I) = x^3$; however, the minimal polynomial of O is simply $\mu(x) = x$. Generally, this is similarly the case for all $n \times n$ zero matrices.

Even though the minimal polynomial of a matrix is not necessarily the characteristic polynomial of the matrix, we know by Proposition 2.3.9 that the minimal polynomial is always a factor of the characteristic polynomial. Consequently, the **roots** of the minimal polynomial are always among the roots of the characteristic polynomial. Explicitly, for any scalar c such that $\mu(c) = 0$, we must have that $\chi(c) = 0$. We refer to such a scalar c such that $\chi_A(c) = 0$ as a **characteristic value** of A. We note that the characteristic values of A are precisely those scalars such that $\det(cI - A) = 0$.

Proposition 2.3.13. *Given any $n \times n$ matrix A, the following are equivalent.*

1.) *We have that $\chi_A(c) = 0$.*

2.) *We have that* $\det(cI - A) = 0$.

3.) *We have that* $cI - A$ *is not invertible.*

Proof. By definition of the characteristic polynomial $\chi_A(x)$ of A, it follows that the first two statements are equivalent. By Proposition 2.2.10, the second and third statements are equivalent. □

Under this identification, we can drastically narrow down the possibilities for $\mu_A(x)$.

Proposition 2.3.14. *Given any* $n \times n$ *matrix* A*, the characteristic polynomial of* A *and the minimal polynomial of* A *have the same roots. Particularly, the minimal polynomial of* A *is divisible by every irreducible polynomial factor of the characteristic polynomial of* A*.*

Proof. We will prove that $\mu_A(c) = 0$ if and only if c is a characteristic value of A. By the Factor Theorem, if we assume that $\mu_A(c) = 0$, then $\mu_A(x) = (x-c)q(x)$ for some polynomial $q(x)$ of strictly lesser degree than $\mu_A(x)$. By definition of $\mu_A(x)$, we must have that $q(A)$ is nonzero. Consequently, we have that $O = \mu_A(A) = (A - cI)q(A)$, hence $cI - A$ cannot be invertible because its product with the nonzero matrix $-q(A)$ is the zero matrix. We conclude by Proposition 2.3.13 that c is a characteristic value of A. Conversely, if c is a characteristic value of A, then $cI - A$ is not invertible, hence there exists a nonzero $n \times n$ matrix B such that $(cI - A)B = O$ or $cB = AB$. Crucially, for any integer $1 \leq k \leq n$, we have that $A^k B = A^{k-1}(AB) = A^{k-1}(cB) = c(A^{k-1}B) = \cdots = c^k B$ by Propositions 1.2.5 and 1.2.6. Consequently, it follows that $O = OB = \mu_A(A)B = \mu_A(c)B$. Considering that $\mu_A(c)$ is a scalar and B is a nonzero matrix, this is only possible if $\mu_A(c) = 0$. □

Example 2.3.15. Consider the following 3×3 matrix A and its characteristic matrix $xI - A$.

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad xI - A = \begin{bmatrix} x+1 & 0 & 0 \\ 0 & x-1 & 0 \\ 0 & 0 & x+1 \end{bmatrix}$$

One can readily verify that $\chi(x) = (x+1)^2(x-1)$ is the characteristic polynomial of A. Consequently, by Proposition 2.3.14, we must have that $\mu(x) = \chi(x)$ or $\mu(x) = (x+1)(x-1) = x^2 - 1$. Considering that $A^2 = I$, it follows that $A^2 - I = O$ so that $\mu(x) = x^2 - 1$.

Example 2.3.16. Consider the following 3×3 matrix A and its characteristic matrix $xI - A$.

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

By definition, the characteristic polynomial of A is found by computing the following.

$$\begin{aligned} \chi(x) = \det(xI - A) &= (x-1)[(x-2)(x-3) - 6] + [-2(x-3) - 6] + [6 + 3(x-2)] \\ &= (x-1)(x^2 - 5x + 6 - 6) - (2x - 6 + 6) + (6 - 3x - 6) \\ &= (x-1)(x^2 - 5x) - 5x \\ &= x^3 - 6x^2 \end{aligned}$$

Considering that $\chi(x) = x^3 - 6x^2 = x^2(x-6)$, it follows that $\mu(x) = \chi(x)$ or $\mu(x) = x(x-6)$. We conclude that $\mu(x) = x(x-6)$ because $A(A-6I) = O$, as the following calculation shows.

$$A(A-6I) = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} -5 & 1 & 1 \\ 2 & -4 & 2 \\ 3 & 3 & -3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Explicitly, one need only check that the first row is zero because the second and third rows of $A(A-6I)$ are merely a scalar multiple of the first row of $A(A-6I)$ by definition of A.

2.4 Eigenvalues and Eigenvectors

Given any linear transformation $T : V \to V$ from any vector space V to itself, we refer to a vector $v \in V$ as an **eigenvector** of T corresponding to a scalar α if and only if we have that $T(v) = \alpha v$.

Example 2.4.1. Consider the real vector space $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$. We have seen many times already (and we know from calculus) that the derivative $\frac{d}{dx}$ defines a linear transformation from the vector space $\mathcal{C}^1(\mathbb{R})$ of continuously differentiable functions to itself. Particularly, observe that for any real number α, the function $f(x) = e^{\alpha x}$ is continuously differentiable and satisfies that

$$\frac{d}{dx} e^{\alpha x} = \alpha e^{\alpha x}.$$

Consequently, $e^{\alpha x}$ is an eigenvector of $\mathcal{C}^1(\mathbb{R})$ corresponding to the real number α.

Example 2.4.2. Consider the real vector space $\mathbb{R}^{3\times 1}$ of real 3×1 matrices. Given any real 3×3 matrix A, we may define a linear transformation $T_A : \mathbb{R}^{3\times 1} \to \mathbb{R}^{3\times 1}$ by declaring that $T_A(X) = AX$. Particularly, if A is a diagonal real 3×3 matrix, then the standard basis vectors $E_1 = (1, 0, 0)$, $E_2 = (0, 1, 0)$, and $E_3 = (0, 0, 1)$ are three examples of eigenvectors of the linear transformation T_A.

$$T_A(E_1) = AE_1 = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ 0 \\ 0 \end{bmatrix} = a_{11} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = a_{11} E_1$$

$$T_A(E_2) = AE_2 = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ a_{22} \\ 0 \end{bmatrix} = a_{22} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = a_{22} E_2$$

$$T_A(E_3) = AE_3 = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ a_{33} \end{bmatrix} = a_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = a_{33} E_3$$

Explicitly, we have that E_i is an eigenvector of T_A corresponding to the scalar a_{ii}.

Certainly, the zero vector is an eigenvector of every linear transformation that corresponds to every scalar α: indeed, we have that $T(O) = O = \alpha O$ for all scalars α. Consequently, we will restrict our attention to those nonzero vectors that are eigenvectors of T. Given any nonzero vector $v \in V$ such that $T(v) = \alpha v$ for some scalar α, we will say that α is an **eigenvalue** of T corresponding to the eigenvector v of T. Crucially, the following uniqueness property of eigenvalues holds.

Proposition 2.4.3. *Given any linear transformation $T : V \to V$ from a vector space V to itself, if v is an eigenvector of T corresponding to an eigenvalue α, then the scalar α is uniquely determined by its eigenvector v in the sense that if $T(v) = \beta v$ for any scalar β, we must have that $\beta = \alpha$.*

Proof. On the contrary, we will assume that β and α are distinct. Consequently, we have that $\alpha - \beta$ is a nonzero scalar. By assumption that $T(v) = \beta v$ and by hypothesis that v is an eigenvector of T corresponding to the eigenvalue α, we have that $\alpha v = T(v) = \beta v$ so that $\alpha v - \beta v = O$ and $(\alpha - \beta)v = O$. Considering that $\alpha - \beta$ is a nonzero scalar, we can multiply both sides of this identity by its inverse to obtain that $v = O$. But this is impossible: by hypothesis that v admits an eigenvalue α, we must have that v is a nonzero vector by definition of an eigenvalue. $\square$

Consequently, if a nonzero vector $v \in V$ admits an eigenvalue α, then that scalar α is uniquely determined by v, and there cannot exist another scalar β such that $T(v) = \beta v$.

Equivalently, we can define the eigenvectors of T as the vectors of V that lie in the kernel of some linear transformation from V to itself that can be obtained from T.

Proposition 2.4.4. *Given any linear transformation $T : V \to V$ from a vector space V to itself, the following statements are equivalent.*

1.) *We have that v is an eigenvector of T corresponding to some scalar α.*

2.) *We have that $v \in \ker(\alpha I - T)$ for some scalar α and the identity transformation $I : V \to V$.*

Proof. By definition, if v is an eigenvector of T corresponding to some scalar α, then we must have that $T(v) = \alpha v = \alpha I(v)$ for the identity transformation $I : V \to V$, from which it follows by subtraction that $\alpha I(v) - T(v) = O$ and $(\alpha I - T)(v) = O$ so that v lies in the kernel of the linear transformation $\alpha I - T$. Conversely, if $v \in \ker(\alpha I - T)$ for some scalar α, then by definition of the kernel of $\alpha I - T$, we have that $O = (\alpha I - T)(v) = \alpha I(v) - T(v) = \alpha v - T(v)$ so that $T(v) = \alpha v$. $\square$

Likewise, we can equivalently define the eigenvalues of T as the scalars α for which the linear transformation $\alpha I - T$ from the vector space V to itself is not invertible.

Proposition 2.4.5. *Given any linear transformation $T : V \to V$ from a vector space V to itself, the following statements are equivalent.*

1.) *We have that α is an eigenvalue of T corresponding to some nonzero vector $v \in V$.*

2.) *We have that $\alpha I - T$ is not invertible for the identity transformation $I : V \to V$.*

Proof. By definition, if α is an eigenvalue of T corresponding to some nonzero vector $v \in V$, then we must have that $T(v) = \alpha v = \alpha I(v)$ for the identity transformation $I : V \to V$, from which it follows by subtraction that $\alpha I(v) - T(v) = O$ and $(\alpha I - T)(v) = O$ so that v lies in the kernel of the linear transformation $\alpha I - T$. Considering that v is a nonzero vector of V, we conclude by Proposition 1.11.6 and Corollary 1.13.7 that $\alpha I - T$ is not invertible. Conversely, by the same proposition and corollary as before, if $\alpha I - T$ is not invertible, then there exists a nonzero vector $v \in V$ such that $O = (\alpha I - T)(v) = \alpha I(v) - T(v) = \alpha v - T(v)$ so that $T(v) = \alpha v$ and α is an eigenvalue of v. $\square$

Based on the previous two propositions, we are in a very profitable position to relate our study of the eigenvectors and eigenvalues of a linear transformation back to our previous foray into determinants of square matrices and their characteristic polynomials. Explicitly, for any linear transformation $T : V \to V$ from a vector space V of dimension n to itself and any ordered basis $v_1, \dots, v_n$ of V, we obtain an $n \times n$ matrix A that behaves as the linear transformation T on the coordinate vectors of V with respect to the chosen ordered basis vectors. Consequently, we may identify the linear transformation T and the $n \times n$ matrix A in this sense. We will soon see that the specification of the ordered basis of V is merely a choice that we can make to simplify A as much as possible. Particularly, if we were to pick another ordered basis of V, then the matrix representation A of T with respect to the original (convenient) ordered basis and the matrix representation B of T with respect to this new ordered basis would possess the same properties. Even more, for any scalar α, the matrix representation for the linear transformation $\alpha I - T$ is given by the $n \times n$ matrix $\alpha I - A$ for the $n \times n$ identity matrix I. Based on this observation, we define the **determinant** $\det(\alpha I - T)$ of the linear transformation $\alpha I - T$ as the determinant $\det(\alpha I - A)$ of the $n \times n$ matrix $\alpha I - A$. We will also define the **characteristic polynomial** of T as $\det(xI - T)$. By the previous sentence, this is nothing more than the characteristic polynomial of the matrix A that represents T.

Conversely, we may also translate our present terminology about eigenvalues and eigenvectors of linear transformations into meaningful statements involving matrices. We will say that a real $n \times 1$ vector X is an **eigenvector** of a real $n \times n$ matrix A if it holds that $AX = cX$ for some real number c. Like before, this is equivalent to the property that $(cI - A)X = O$ for the $n \times n$ identity matrix I and the $n \times 1$ zero vector O. Consequently, the $n \times 1$ zero vector O is an eigenvector of any real $n \times n$ matrix. Even more, we will say that c is an **eigenvalue** of a real $n \times n$ matrix A if and only if there exists a nonzero real $n \times 1$ matrix X such that $AX = cX$; such a nonzero real $n \times 1$ matrix X is called an eigenvector of A corresponding to the eigenvalue c of A.

Our next proposition summarizes the relationship between eigenvalues of a linear transformation T from a finite-dimensional vector space to itself and the eigenvalues of any $n \times n$ matrix representing T. Crucially, it also provides a necessary and sufficient condition for the existence of eigenvalues.

Proposition 2.4.6. *Given any linear transformation $T : V \to V$ from a (real) vector space V of dimension n to itself and any (real) $n \times n$ matrix A that represents T with respect to some fixed ordered basis of V, the following statements are equivalent.*

1.) *We have that c is an eigenvalue of T corresponding to some nonzero vector $v \in V$.*

2.) *We have that c is an eigenvalue of A corresponding to some nonzero (real) $n \times 1$ matrix X.*

3.) *We have that $\det(cI - A) = 0$, i.e., we have that c is a root of $\chi(x) = \det(xI - A)$.*

Put another way, the eigenvalues of T are precisely the eigenvalues of any matrix A that represents T, and these eigenvalues are exactly the roots of the characteristic polynomial $\det(xI - A)$ of A.

Proof. We have that c is an eigenvalue of T corresponding to some nonzero vector $v \in V$ if and only if v lies in the kernel of the linear transformation $cI - T$ if and only if the coordinate vector X of v with respect to the fixed ordered basis of V satisfies that $(cI - A)X = O$ if and only if $\det(cI - A) = 0$. Explicitly, the first equivalence holds by Proposition 2.4.4; the second equivalence holds by definition of the matrix representation A of T; and the third equivalence holds by Proposition 2.2.10. ☐

Example 2.4.7. Consider the following diagonal real 3×3 matrix.

$$A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

Observe that the characteristic matrix $xI - A$ is a diagonal matrix with diagonal components $x - a$, $x - b$, and $x - c$. By [Lan86, Exercise 3] on page 208, the determinant of a diagonal matrix is the product of its diagonal components, hence we have that $\chi(x) = \det(xI - A) = (x - a)(x - b)(x - c)$. Consequently, the eigenvalues of A are simply the diagonal components a, b, and c.

Example 2.4.8. Consider the 2×2 matrix A from Example 2.3.1.

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

We showed that $\chi(x) = \det(xI - A) = (x - 3)(x + 1)$, hence the eigenvalues of A are -1 and 3.

Example 2.4.9. Consider the 3×3 matrix A from Example 2.3.2.

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

Considering that $\chi(x) = \det(xI - A) = (x - 1)(x - 2)(x + 1)$, the eigenvalues of A are -1, 1, and 2.

Example 2.4.10. Consider the 3×3 matrix A from Example 2.3.16.

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

We demonstrated previously that $\chi(x) = \det(xI - A) = x^2(x - 6)$, hence the eigenvalues of A are 0 (with multiplicity two) and 6. We will soon return to this notion of multiplicity.

Once we have found the eigenvalues of a matrix by computing the roots of its characteristic polynomial, the hunt is on to determine the eigenvectors of A corresponding to these eigenvalues. We note that if c is an eigenvalue of an $n \times n$ matrix A, then by the proof of Proposition 2.4.6, an eigenvector of A corresponding to the eigenvalue c of A is simply an $n \times 1$ matrix X such that $(cI - A)X = O$. Consequently, in practice, the way to find the eigenvectors of an $n \times n$ matrix A corresponding to an eigenvalue c of A is to solve the matrix equation $(cI - A)X = O$.

Example 2.4.11. Consider the following diagonal real 3×3 matrix with eigenvalues 1, 2, and 3.

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

By definition, an eigenvector of A corresponding to the eigenvalue 1 of A is a real 3×1 matrix X such that $(I - A)X = O$. By interpreting this in the present context, we obtain the following.

$$\begin{bmatrix} 0 \\ -2y \\ -3z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Consequently, we must have that $-2y = 0$ and $-3z = 0$ so that $y = 0$ and $z = 0$, hence every real 3×1 vector $X = (x, 0, 0)^t$ is an eigenvector of A corresponding to the eigenvalue 1.

Example 2.4.12. Consider the 2×2 matrix A from Example 2.4.8 with eigenvalues -1 and 3.

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

We have that $X = (x, y)^t$ is an eigenvector of A corresponding to the eigenvalue -1 if and only if $(-I - A)X = O$ if and only if $(-I - A)(x, y)^t = (0, 0)^t$ if and only if

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

if and only if $x + y = 0$ if and only if $y = -x$. Consequently, the eigenvectors of A corresponding to the eigenvalue -1 of A are precisely the real 2×1 matrices $X = (x, -x)^t$ for some real number x.

Example 2.4.13. Consider the 3×3 matrix A from Example 2.4.9 with eigenvalues -1, 1, and 2.

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

We have that $X = (x, y, z)^t$ is an eigenvector of A corresponding to the eigenvalue 2 if and only if $(2I - A)X = O$ if and only if $(2I - A)(x, y, z)^t = (0, 0, 0)^t$ if and only if

$$\begin{bmatrix} x - y \\ -x + 2y - z \\ -y + z \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

if and only if $x - y = 0$ and $-x + 2y - z = 0$ and $-y + z = 0$ if and only if $y = x$ and $z = y$. Consequently, the eigenvectors of A corresponding to the eigenvalue 2 of A are precisely the real 3×1 matrices $X = (x, x, x)^t$ for some real number x.

Example 2.4.14. Consider the 3×3 matrix A from Example 2.4.10 with eigenvalues 0 and 6.

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

We have that $X = (x, y, z)^t$ is an eigenvector of A corresponding to the eigenvalue 0 if and only if $-AX = O$ if and only if $AX = O$ if and only if $A(x, y, z)^t = (0, 0, 0)^t$ if and only if

$$\begin{bmatrix} x + y + z \\ 2x + 2y + 2z \\ 3x + 3y + 3z \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

if and only if $x + y + z = 0$ if and only if $z = -x - y$. Consequently, the eigenvectors of A corresponding to the eigenvalue 0 of A are precisely the real 3×1 matrices

$$X = \begin{bmatrix} x \\ y \\ -x - y \end{bmatrix} = \begin{bmatrix} x \\ 0 \\ -x \end{bmatrix} + \begin{bmatrix} 0 \\ y \\ -y \end{bmatrix}$$

for some real numbers x and y. Crucially, we note that in this example, the multiplicity of the root 0 in the characteristic polynomial of A is two, and two distinct variables appeared in the eigenvectors of A corresponding to the eigenvalue 0. Once again, we will soon investigate further.

2.5 Eigenspaces

Given any linear transformation $T : V \to V$ from a vector space V to itself, we have previously distinguished a vector $v \in V$ as an eigenvector of T if there exists a scalar α such that $T(v) = \alpha v$. Consequently, the zero vector is an eigenvector of any linear transformation because it holds that $T(O) = O = \alpha O$ for any scalar α. Even more, if v is a nonzero vector, then we say that v is an eigenvector of T corresponding to the eigenvalue α of T if $T(v) = \alpha v$. By the linearity of T, if v and w are any eigenvectors of T corresponding to an eigenvalue α of T, then we have that

$$T(v+w) = T(v) + T(w) = \alpha v + \alpha w = \alpha(v+w),$$

hence $v + w$ is an eigenvector of T corresponding to the eigenvalue α. Likewise, for any scalar β, we have that $T(\beta v) = \beta T(v) = \beta(\alpha v) = \alpha(\beta v)$, from which it follows that βv is an eigenvector of T corresponding to the eigenvalue α. Combined, these two observations prove the following.

Proposition 2.5.1. *Given any linear transformation $T : V \to V$ from a vector space V to itself, the collection $W_\alpha = \{v \in V \mid T(v) = \alpha v\}$ of eigenvectors of V corresponding to an eigenvalue α of T is a vector subspace of V that is called the* **eigenspace** *of T with respect to the eigenvalue α.*

Remark 2.5.2. By Proposition 2.4.4, we may identify the eigenspace W_α and $\ker(\alpha I - T)$.

Example 2.5.3. Consider the following 3×3 matrix of Example 2.4.11.

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

We showed previously that the eigenvectors of A corresponding to the eigenvalue 1 are $(x, 0, 0)^t$ for some real number x, hence we have that $W_1 = \{X \in \mathbb{R}^{3\times 1} \mid AX = X\} = \operatorname{span}\{(1, 0, 0)^t\}$.

Example 2.5.4. Consider the following 2×2 matrix A from Example 2.4.12.

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

We proved in that example that the eigenvectors of A corresponding to the eigenvalue -1 are $(x, -x)^t$ for some real number x, hence we have that $W_{-1} = \{X \in \mathbb{R}^{2\times 1} \mid AX = -X\} = \operatorname{span}\{(1, -1)^t\}$.

Example 2.5.5. Consider the following 3×3 matrix A from Example 2.4.13.

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

Last section, we illustrated that the eigenvalues of A corresponding to the eigenvalue 2 are $(x, x, x)^t$ for some real number x so that $W_2 = \{X \in \mathbb{R}^{3\times 1} \mid AX = 2X\} = \operatorname{span}\{(1, 1, 1)^t\}$.

Example 2.5.6. Consider the following 3×3 matrix A from Example 2.4.10.

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

We demonstrated previously that the eigenvalues of A corresponding to the eigenvalue 0 are of the form $(x, 0, -x)^t + (0, y, -y)^t$ for some real numbers x and y. Consequently, the eigenspace of A corresponding to the eigenvalue 0 is $W_0 = \{X \in \mathbb{R}^{3 \times 1} \mid AX = O\} = \operatorname{span}\{(1, 0, -1)^t, (0, 1, -1)^t\}$.

Our ultimate objective throughout this chapter is to study the **canonical forms** for a linear transformation from a vector space to itself (or equivalently of an $n \times n$ matrix). Put simply, these are representations of linear transformations (or matrices) by matrices that are (in a strict sense) in "simplest form." One of the most delightful examples of this occurs when the underlying vector space on which the linear transformation is defined admits a basis of eigenvectors for the linear transformation. Explicitly, let us assume that some vectors $v_1, \dots, v_n$ form a basis for the n-dimensional vector space V on which a linear transformation $T : V \to V$ is defined. Certainly, the best case scenario is that the vectors $v_1, \dots, v_n$ are actually eigenvectors of T corresponding to distinct eigenvalues $\alpha_1, \dots, \alpha_n$, respectively: indeed, if this is the case, then the following hold.

$$\begin{aligned} T(v_1) &= \alpha_1 v_1 = \alpha_1 v_1 + 0 v_2 + \cdots + 0 v_n \\ T(v_2) &= \alpha_2 v_2 = 0 v_1 + \alpha_2 v_2 + \cdots + 0 v_n \\ &\vdots \\ T(v_n) &= \alpha_n v_n = 0 v_1 + 0 v_2 + \cdots + \alpha_n v_n \end{aligned}$$

Consequently, the $n \times n$ matrix A that represents T with respect to this ordered basis $v_1, \dots, v_n$ is a diagonal matrix! Explicitly, the jth column of A consists of zeros in all rows except the jth row, and the component of the jth row of A is the eigenvalue α_j corresponding to the eigenvector v_j.

$$A = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{bmatrix}$$

Definition 2.5.7. We say that a linear transformation $T : V \to V$ from a vector space of dimension n to itself is **diagonalizable** if there exists an ordered basis $v_1, \dots, v_n$ of V such that $T(v_i) = \alpha_i v_i$ for some scalars $\alpha_1, \dots, \alpha_n$. Put another way, a linear transformation from a finite-dimensional vector space V to itself is diagonalizable if and only if V admits a basis of eigenvectors for T if and only if T can be represented by a diagonal matrix with respect to some ordered basis of V.

Our first order of business is to provide a necessary and sufficient condition for the diagonalizability of a linear transformation (or a square matrix). We must first demonstrate that the eigenvectors of a linear transformation corresponding to distinct eigenvalues are linearly independent.

Proposition 2.5.8. *Given any linear transformation $T : V \to V$ from a vector space V to itself, if $v_1, \dots, v_n$ are any eigenvectors of T corresponding respectively to the pairwise distinct eigenvalues $\alpha_1, \dots, \alpha_n$ of T, then the collection of eigenvectors $\{v_1, \dots, v_n\}$ is linearly independent.*

Proof. We proceed by induction on the number n of eigenvectors present. By definition, if α_1 is an eigenvalue of T corresponding to the eigenvector v_1 of T, then v_1 is a nonzero vector, hence v_1 is linearly independent. Consider any eigenvectors $v_1, \dots, v_n$ of T corresponding respectively to the pairwise distinct eigenvalues $\alpha_1, \dots, \alpha_n$ of T. We must show that if $\beta_1 v_1 + \cdots + \beta_n v_n = O$, then $\beta_1 = \cdots = \beta_n = 0$. Observe that if we apply T to the above relation of linear dependence, then

$$O = T(O) = T(\beta_1 v_1 + \cdots + \beta_n v_n) = \beta_1 T(v_1) + \cdots + \beta_n T(v_n) = \beta_1 \alpha_1 v_1 + \cdots + \beta_n \alpha_n v_n$$

by assumption that v_i is an eigenvector of T corresponding to the eigenvalue α_i of T. On the other hand, if we multiply our original relation of linear dependence by α_1, then we find that

$$O = \alpha_1 O = \alpha_1(\beta_1 v_1 + \cdots + \beta_n v_n) = \beta_1 \alpha_1 v_1 + \cdots + \beta_n \alpha_1 v_n.$$

By subtracting the second identity above from the first, we obtain a third identity

$$O = \beta_2(\alpha_1 - \alpha_2)v_2 + \cdots + \beta_n(\alpha_1 - \alpha_n)v_n.$$

By induction, these $n-1$ vectors are linearly independent, hence we conclude that $\beta_i(\alpha_1 - \alpha_i) = 0$ for each integer $2 \leq i \leq n$. Considering that α_1 and α_i are distinct eigenvalues for each integer $2 \leq i \leq n$, we must have that $\alpha_1 - \alpha_i$ is nonzero. Cancelling the factor of $\alpha_1 - \alpha_i$ from each identity $\beta_i(\alpha_1 - \alpha_i) = 0$ yields that $\beta_2 = \cdots = \beta_n = 0$, so our original relation of linear independence now states that $\beta_1 v_1 = O$. But this implies that $\beta_1 = 0$ because v_1 is nonzero by hypothesis. □

Corollary 2.5.9. *Given any linear transformation $T : V \to V$ from a vector space V to itself, if T admits $\dim(V)$ eigenvectors corresponding to $\dim(V)$ distinct eigenvalues, then the collection of eigenvectors of T form a basis for V. Consequently, in this case, we have that T is diagonalizable. Particularly, if T admits $\dim(V)$ distinct eigenvalues, then T must be diagonalizable.*

Proof. By the fourth part of Theorem 1.8.10, every collection of $\dim(V)$ linearly independent vectors of V form a basis for V. By Proposition 2.5.8, eigenvectors corresponding to distinct eigenvalues are linearly independent, hence any collection of $\dim(V)$ eigenvectors corresponding to $\dim(V)$ distinct eigenvalues form a basis for V. By Definition 2.5.7, we conclude that T is diagonalizable: its matrix representation with respect to any ordered basis of eigenvectors of T corresponding to distinct eigenvalues of T is a diagonal matrix. Last, if T admits $\dim(V)$ distinct eigenvalues, then T must be diagonalizable because in this case, each of the $\dim(V)$ distinct eigenvalues of T corresponds to an eigenvector of T, i.e., there are $\dim(V)$ linearly independent eigenvectors of T. □

Example 2.5.10. Consider any linear transformation $T : V \to V$ from a vector space V of dimension three to itself that is represented by the following 3×3 matrix from Example 2.4.7.

$$A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

We demonstrated previously that the eigenvalues of T are a, b, and c corresponding to the respective eigenvectors E_1, E_2, and E_3. Consequently, T is diagonalizable. Of course, we did not need Corollary 2.5.9 to deduce this fact; we could have simply looked at the diagonal matrix A representing T.

Example 2.5.11. Consider any linear transformation $T : V \to V$ from a vector space V of dimension two to itself that is represented by the following 2×2 matrix A from Example 2.4.8.

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

We proved in that example that the eigenvalues of T are -1 and 3, hence by Corollary 2.5.9, we conclude that T is diagonalizable because it admits $\dim(V) = 2$ distinct eigenvalues.

Example 2.5.12. Consider any linear transformation $T : V \to V$ from a vector space V of dimension three to itself that is represented by the following 3×3 matrix A from Example 2.4.9.

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

We demonstrated in that example that T admits the $\dim(V) = 3$ distinct eigenvalues -1, 1, and 2, hence by Corollary 2.5.9, it follows that T is diagonalizable.

Example 2.5.13. Consider any linear transformation $T : V \to V$ from a vector space V of dimension three to itself that is represented by the following 3×3 matrix A from Example 2.4.10.

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

Even though we showed that T admits only two distinct eigenvalues 0 and 6, it turns out that T is diagonalizable. Explicitly, by Example 2.5.6, the eigenspace of A corresponding to the eigenvalue 0 has dimension two: indeed, we have that $W_0 = \{X \in \mathbb{R}^{3 \times 1} \mid AX = O\} = \operatorname{span}\{(1, 0, -1)^t, (0, 1, -1)^t\}$. Consequently, for any eigenvector X of A corresponding to the eigenvalue 6, it follows by Proposition 2.5.8 and Corollary 2.5.9 that $\{(1, 0, -1)^t, (0, 1, -1)^t, X\}$ is an ordered basis for $\mathbb{R}^{3 \times 1}$ consisting of eigenvectors for A; thus, we conclude that the vectors v_1, v_2, and v_3 of V corresponding to these coordinate vectors in $\mathbb{R}^{3 \times 1}$ form an ordered basis of V consisting of eigenvectors of T.

Example 2.5.13 illustrates that the condition of Corollary 2.5.9 is sufficient but not necessary for the diagonalizability of T. Consequently, we seek more restrictive properties of T under which T is diagonalizable and for which T is not diagonalizable if the properties are not satisfied. Before we are able to state such properties explicitly, we need the following lemmas.

Lemma 2.5.14. *Let V be a vector space with vector subspaces U and W. If U and W have finite dimension, then $U + W$ has finite dimension* $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$.

Proof. We must first recall the definitions of the attendant objects at hand. By Proposition 1.6.17, we have that $U + W = \{u + w \mid u \in U \text{ and } w \in W\}$ is the vector subspace of V consisting of all sums of a vector $u \in U$ and a vector $w \in W$. Likewise, we have that $U \cap W = \{v \in V \mid v \in U \cap W\}$ is the vector subspace of V consisting of all vectors $v \in V$ that lie in both U and W.

By the fifth part of Theorem 1.8.10, the vector subspace $U \cap W$ of the finite-dimensional vector spaces U and W admits a basis $v_1, \dots, v_k$. Even more, by Proposition 1.8.8, we may extend this to a basis $v_1, \dots, v_k, u_{k+1}, \dots, u_\ell$ of U and a basis $v_1, \dots, v_k, w_{k+1}, \dots, w_m$ of W. We claim that the vectors $v_1, \dots, v_k, u_{k+1}, \dots, u_\ell, w_{k+1}, \dots, w_m$ form a basis for $U + W$. Observe that in this case, we have that $\dim(U+W) = k + (\ell - k) + (m - k) = \ell + m - k = \dim(U) + \dim(W) - \dim(U \cap W)$, as desired. By definition, every vector of $U + W$ can be written as $u + w$ for some vectors $u \in U$ and $w \in W$. Consequently, by our proposed basis, there exist unique scalars $\alpha_1, \dots, \alpha_\ell$ such that

$$u = \alpha_1 v_1 + \cdots + \alpha_k v_k \alpha_{k+1} u_{k+1}, \dots, \alpha_\ell u_\ell.$$

Likewise, there exist unique scalars $\beta_1, \dots, \beta_m$ such that

$$w = \beta_1 v_1 + \cdots + \beta_k v_k + \beta_{k+1} w_{k+1} + \cdots + \beta_\ell w_\ell.$$

Combined, these two observations demonstrate that every vector of $U + W$ can be written as

$$u + w = (\alpha_1 + \beta_1) v_1 + \cdots + (\alpha_k + \beta_k) v_k + \alpha_{k+1} u_{k+1} + \cdots + \alpha_\ell u_\ell + \beta_{k+1} w_{k+1} + \cdots + \beta_m w_m.$$

We conclude that $U + W = \operatorname{span}\{v_1, \dots, v_k, u_{k+1}, \dots, u_\ell, w_{k+1}, \dots, w_m\}$, hence it remains to be seen that these vectors are linearly independent. Consider any scalars $\alpha_1, \dots, \alpha_\ell, \beta_{k+1}, \dots, \beta_m$ such that

$$\alpha_1 v_1 + \cdots + \alpha_k v_k + \alpha_{k+1} u_{k+1} + \cdots + \alpha_\ell u_\ell + \beta_{k+1} w_{k+1} + \cdots + \beta_m w_m = O.$$

By subtracting the linear combination of $w_{k+1}, \dots, w_m$ from both sides, we find that

$$-\beta_{k+1} w_{k+1} - \cdots - \beta_m w_m = \alpha_1 v_1 + \cdots + \alpha_k v_k + \alpha_{k+1} u_{k+1} + \cdots + \alpha_\ell u_\ell$$

so that $-\beta_{k+1} w_{k+1} - \cdots - \beta_m w_m$ lies in U. By assumption that the vectors $w_{k+1}, \dots, w_m$ belong to the vector space W in the first space, we must have that $-\beta_{k+1} w_{k+1} - \cdots - \beta_m w_m$ lies in W, from which we conclude that $-\beta_{k+1} w_{k+1} - \cdots - \beta_m w_m$ lies in $U \cap W$. By appealing to our basis $v_1, \dots, v_k$ for $U \cap W$, we may find (unique) scalars $\gamma_1, \dots, \gamma_k$ such that

$$-\beta_{k+1} w_{k+1} - \cdots - \beta_m w_m = \gamma_1 v_1 + \cdots + \gamma_k v_k.$$

Ultimately, we obtain a relation of linear dependence among the vectors $v_1, \dots, v_k, w_{k+1}, \dots, w_m$.

$$\gamma_1 v_1 + \cdots + \gamma_k v_k + \beta_{k+1} w_{k+1} + \cdots + \beta_m w_m = O$$

By construction, these vectors are linearly independent, and we conclude that $\beta_{k+1} = \cdots = \beta_m = 0$. By returning to our fourth displayed equation above, we find that

$$\alpha_1 v_1 + \cdots + \alpha_k v_k + \alpha_{k+1} u_{k+1} + \cdots + \alpha_\ell u_\ell = O,$$

and the linear independence of $v_1, \dots, v_k, u_{k+1}, \dots, u_\ell$ implies that $\alpha_1 = \cdots = \alpha_\ell = 0$. □

Given any subspaces U and W of a vector space V, we say that the sum $U + W$ is **direct** if it holds that $U \cap W = \{O\}$; in this case, there are no relations among the vectors of U and W.

Corollary 2.5.15. *Let V be a vector space with vector subspaces U and W. If U and W have finite dimension and $U \cap W = \{O\}$, then $U + W$ has finite dimension* $\dim(U + W) = \dim(U) + \dim(W)$.

Proof. By Lemma 2.5.14, it suffices to note that $\dim(U \cap W) = 0$. □

Lemma 2.5.16. *Given any linear transformation $T : V \to V$ from a finite-dimensional vector space V to itself, if $\alpha_1, \dots, \alpha_k$ are the distinct eigenvalues of T and $W_{\alpha_1}, \dots, W_{\alpha_k}$ are the respective eigenspaces of V, then $\dim(W_{\alpha_1} + \cdots + W_{\alpha_k}) = \dim(W_{\alpha_1}) + \cdots + \dim(W_{\alpha_k})$. Particularly, an ordered basis for $W_{\alpha_1} + \cdots + W_{\alpha_k}$ consists of consecutive ordered bases for W_{α_i} for each integer $1 \leq i \leq k$.*

Proof. We proceed by induction on the number k of distinct eigenvalues of T. By Proposition 2.4.3, we have that $W_{\alpha_1} \cap W_{\alpha_2} = \{O\}$ for any pair of distinct eigenvalues α_1 and α_2 of T, hence the claim follows from Corollary 2.5.15 in the case that $k = 2$. Given distinct eigenvalues $\alpha_1, \dots, \alpha_k$ of T, consider the vector spaces $U = W_{\alpha_1}$ and $W = W_{\alpha_2} + \cdots + W_{\alpha_k}$ By Lemma 2.5.14, we have that

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

Consider any vector $v \in U \cap W$. By definition, such a vector has the property that $T(v) = \alpha_1 v$, and it can be written as $v_2 + \cdots + v_k$ for some eigenvectors $v_i \in W_{\alpha_i}$, hence we have that

$$\alpha_1 v = T(v) = T(v_2 + \cdots + v_k) = T(v_2) + \cdots + T(v_k) = \alpha_2 v_2 + \cdots + \alpha_k v_k.$$

By subtracting $\alpha_1 v_1$ from both sides, we obtain an expression of linear dependence

$$-\alpha_1 v + \alpha_2 v_2 + \cdots + \alpha_k v_k = O.$$

By Proposition 2.5.8, we must have that $\alpha_1 = \cdots = \alpha_k = 0$. But this is impossible: $\alpha_1, \dots, \alpha_k$ are distinct eigenvalues of T. Consequently, we must have that $v = O$. We conclude that the sum $U + W$ is direct, hence we have that $\dim(U + W) = \dim(U) + \dim(W)$. By induction, the sum $W = W_{\alpha_2} + \cdots + W_{\alpha_k}$ is direct, i.e., $\dim(W_{\alpha_1} + \cdots + W_{\alpha_k}) = \dim(W_{\alpha_1}) + \cdots + \dim(W_{\alpha_k})$. □

Theorem 2.5.17. *Given any linear transformation $T : V \to V$ from a finite-dimensional vector space V to itself with distinct eigenvalues $\alpha_1, \dots, \alpha_k$, the following conditions are equivalent.*

1.) *We have that T is diagonalizable.*

2.) *We have that $\chi_T(x) = (x - \alpha_1)^{e_1} \cdots (x - \alpha_k)^{e_k}$ and $\dim(W_{\alpha_i}) = e_i$ for each integer $1 \leq i \leq k$.*

3.) *We have that $\dim(V) = \dim(W_{\alpha_1}) + \cdots + \dim(W_{\alpha_k})$.*

Proof. By Definition 2.5.7, we have that T is diagonalizable if and only if there exists a basis of V consisting of eigenvectors of T. Order these basis vectors such that the first e_1 of them correspond to the eigenvalue α_1 and the next e_2 of them correspond to the eigenvalue α_2 and so on for each integer up to and including k. Observe that the matrix representation A of T with respect to this ordered basis is the diagonal matrix in which α_i appears e_i times along the diagonal. Consequently, the characteristic matrix $xI - A$ is the diagonal matrix in which $x - \alpha_i$ appears e_i times along the diagonal, hence we have that $\chi_T(x) = \det(xI - T) = \det(xI - A) = (x - \alpha_1)^{e_1} \cdots (x - \alpha_k)^{e_k}$. We claim that $\dim(W_{\alpha_i}) = e_i$ for each integer $1 \leq i \leq k$. Considering that W_{α_i} can be identified with $\ker(\alpha_i I - A)$ by Proposition 2.4.4 and the construction of the matrix representation of T, it follows that $\dim(W_{\alpha_i}) = \operatorname{nullity}(\alpha_i I - A)$, i.e., $\dim(W_{\alpha_i})$ is equal to the number of zero rows in the square matrix $\alpha_i I - A$. But by construction of A, there are exactly e_i zero rows of $\alpha_i I - A$.

Even more, if we assume that the second condition holds, then the third condition holds because the dimension of V is equal the degree of the characteristic polynomial of T, and that degree is exactly the sum of the exponents of the irreducible factors of the characteristic polynomial of T.

Last, if the third condition holds, then $W_{\alpha_1} + \cdots + W_{\alpha_k}$ is a vector subspace of V of dimension $\dim(W_{\alpha_1} + \cdots + W_{\alpha_k}) = \dim(W_{\alpha_1}) + \cdots + \dim(W_{\alpha_k}) = \dim(V)$ by Lemma 2.5.16. We conclude by the sixth part of Theorem 1.8.10 that $V = W_{\alpha_1} + \cdots + W_{\alpha_k}$ is spanned by the set of eigenvectors of T; thus, by Proposition 1.8.6, it follows that V admits a basis of eigenvectors for T. $\square$

We have encountered diagonalizable matrices so far in this section; however, it is unfortunately not true that every matrix is diagonalizable. We conclude this section with an example.

Example 2.5.18. Consider the following real 3×3 matrix.

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

Even though this matrix looks quite harmless and inconspicuous, it turns out that it is not diagonalizable. Explicitly, there is no basis of the real vector space $\mathbb{R}^{3\times 1}$ of real 3×1 matrices in which the matrix representation of the linear transformation that is left multiplication by A is diagonal. Observe that the characteristic matrix $xI - A$ is the following upper-triangular matrix.

$$xI - A = \begin{bmatrix} x & -1 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{bmatrix}$$

Consequently, the characteristic polynomial of A is $\chi(x) = x^3$ so that 0 is the only eigenvalue of A. By Theorem 2.5.17, we have that A is diagonalizable if and only if the eigenspace W_0 of $\mathbb{R}^{3\times 1}$ corresponding to the eigenvalue 0 of A has dimension three. By definition, we have that $X \in W_0$ if and only if $-AX = O$ if and only if $-A(x, y, z)^t = (0, 0, 0)^t$ if and only if

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -y \\ 0 \\ 0 \end{bmatrix}$$

if and only if $-y = 0$ if and only if $y = 0$. Consequently, we conclude that $X \in W_0$ if and only if there exist real numbers x and y such that $X = (x, 0, z)^t = x(1, 0, 0)^t + z(0, 0, 1)^t$. Put another way, we have that $W_0 = \operatorname{span}\{(1, 0, 0)^t, (0, 0, 1)^t\}$ so that $\dim(W_0) = 2$ and A is not diagonalizable.

We will therefore benefit from the development of tools to understand matrices that are not diagonalizable. One natural question is whether a matrix that is not diagonalizable admits some other adequately nice property. Even though the matrix of Example 2.5.18 is not diagonalizable, it is at least upper-triangular, so perhaps there is some hope. We will soon focus our attention there.

2.6 The Spectral Theorem

Before we move away from our study of diagonalizable matrices, we reserve this section to discuss and prove (as we much as we can of) the following fundamental theorem of real symmetric matrices.

Theorem 2.6.1 (Spectral Theorem)**.** *Every real symmetric matrix is orthogonally diagonalizable. Conversely, every real orthogonally diagonalizable matrix is symmetric.*

Example 2.6.2. Consider the following real 3×3 matrix.

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

Certainly, we have that A is symmetric: it is straightforward to verify that $A^t = A$. Consequently, the Spectral Theorem implies that this matrix is orthogonally diagonalizable. By definition, a real 3×3 matrix is diagonalizable if and only if there exists a basis for $\mathbb{R}^{3\times 1}$ consisting of eigenvectors for A; naturally, this leads us to determine the eigenvectors for A. Before this, of course, we must find the eigenvalues of A. By Proposition 2.4.6, these are simply the roots of the polynomial $\det(xI - A)$.

$$\begin{aligned}
\det(xI - A) &= \begin{vmatrix} x-1 & -2 & -3 \\ -2 & x-1 & -3 \\ -3 & -3 & x \end{vmatrix} \\
&= (x-1)\begin{vmatrix} x-1 & -3 \\ -3 & x \end{vmatrix} - (-2)\begin{vmatrix} -2 & -3 \\ -3 & x \end{vmatrix} + (-3)\begin{vmatrix} -2 & x-1 \\ -3 & -3 \end{vmatrix} \\
&= (x-1)[(x-1)x - (-3)(-3)] + 2[-2x - (-3)(-3)] - 3[(-2)(-3) - (x-1)(-3)] \\
&= (x-1)(x^2 - x - 9) - (4x + 18) - (18 + 9x - 9) \\
&= x^3 - 2x^2 - 8x + 9 - 4x - 18 - 9 - 9x \\
&= x^3 - 2x^2 - 21x - 18
\end{aligned}$$

By inspection, we notice that -1 is a root of this cubic polynomial, hence by the Factor Theorem, it follows that the linear polynomial $x+1$ divides $x^3 - 2x^2 - 21x - 18$. By polynomial long division, we find that $x^3 - 2x^2 - 21x - 18 = (x+1)(x^2 - 3x - 18) = (x+1)(x-6)(x+3)$. Consequently, the eigenvalues of A are -3, -1, and 6. We determine the eigenvectors corresponding to these eigenvalues by solving the matrix equations $AX = -3X$, $AX = -X$, and $AX = 6X$. We have that

$$AX = -3X \text{ if and only if } \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 3 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -3x \\ -3y \\ -3z \end{bmatrix} \text{ if and only if } \begin{bmatrix} x + 2y + 3z \\ 2x + y + 3z \\ 3x + 3y \end{bmatrix} = \begin{bmatrix} -3x \\ -3y \\ -3z \end{bmatrix}$$

if and only if $4x + 2y + 3z = 0$ and $2x + 4y + 3z = 0$ and $3x + 3y + 3z = 0$. By subtracting the second equation from the first equation, we find that $2x - 2y = 0$ so that $y = x$; then, by substituting $y = x$ in the third equation and solving for z, we obtain that $z = -2x$. Consequently, the eigenvectors of A corresponding to the eigenvalue -3 are all of the form $(x, x, -2x)^t$ for some real number x. Choosing to set $x = 1$ gives us an eigenvector $X_{-3} = (1, 1, -2)^t$ for A corresponding to the eigenvalue -3. Likewise, turning our attention to the eigenvalue -1, we have that

$$AX = -X \text{ if and only if } \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 3 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -x \\ -y \\ -z \end{bmatrix} \text{ if and only if } \begin{bmatrix} x + 2y + 3z \\ 2x + y + 3z \\ 3x + 3y \end{bmatrix} = \begin{bmatrix} -x \\ -y \\ -z \end{bmatrix}$$

if and only if $2x + 2y + 3z = 0$ and $3x + 3y + z = 0$. By subtracting the first equation from the second, it follows that $x + y - 2z = 0$; then, by subtracting twice this equation from the first, we find that $7z = 0$ so that $z = 0$ and $y = -x$. We conclude that the eigenvectors of A corresponding to the eigenvalue -1 are of the form $(x, -x, 0)^t$ for some real number x. By taking $x = 1$, we obtain an eigenvector $X_{-1} = (1, -1, 0)^t$ for A corresponding to the eigenvalue -1. Last, we have that

$$AX = 6X \text{ if and only if } \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 3 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 6x \\ 6y \\ 6z \end{bmatrix} \text{ if and only if } \begin{bmatrix} x + 2y + 3z \\ 2x + y + 3z \\ 3x + 3y \end{bmatrix} = \begin{bmatrix} 6x \\ 6y \\ 6z \end{bmatrix}$$

if and only if $-5x + 2y + 3z = 0$ and $2x - 5y + 3z = 0$ and $3x + 3y - 6z = 0$. By dividing this last equation by 3, we find that $x + y - 2z = 0$; now, we may subtract twice this equation from the second equation to find that $-7y + 7z = 0$ or $y = z$. Likewise, we may subtract twice the third equation from the first equation to obtain that $-7x + 7z = 0$ or $x = z$. Ultimately, it follows that the eigenvectors of A corresponding to the eigenvalue 6 are $(x, x, x)^t$ for some real number x; thus, if we substitute $x = 1$, we obtain an eigenvector $X_6 = (1, 1, 1)^t$ for A corresponding to the eigenvalue 6. By Proposition 2.5.8, the eigenvectors X_{-3}, X_{-1}, and X_6 are linearly independent, hence they form a basis for $\mathbb{R}^{3\times 1}$ and the matrix representation for A with respect to this basis is diagonal by the paragraph preceding Definition 2.5.7, so there is essentially nothing new here; however, by the Spectral Theorem, moreover, it is guaranteed that A is orthogonally diagonalizable.

We will return to the above example to finish our verification of the Spectral Theorem for the real symmetric 3×3 matrix at hand, but we must at this point digress to discuss the property of orthogonality of vectors. Earlier in these notes, we defined the **vector dot product** $X \cdot Y$ between real $n \times 1$ column vectors X and Y by declaring that $X \cdot Y = X^t Y$. Observe that the dot product of such vectors results in a real 1×1 vector whose only component is equal to the sum of the products of each row of X and each row of Y. Explicitly, if $X = (x_1, \dots, x_n)^t$ and $Y = (y_1, \dots, y_n)^t$, then

$$X \cdot Y = X^t Y = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 y_1 + \cdots + x_n y_n \end{bmatrix}$$

is just found by doing the ordinary matrix multiplication of a $1 \times n$ matrix and and $n \times 1$ matrix. One important property of the dot product is that it allows us to determine an equivalent condition for an $n \times n$ matrix A to be symmetric in terms of how it behaves with respect to the dot product.

Proposition 2.6.3. *Given any $n \times n$ matrix A, we have that A is symmetric if and only if it holds that $(AX) \cdot Y = X \cdot (AY)$ for all $n \times 1$ column vectors X and Y.*

Proof. Observe that if A is a symmetric matrix, then by definition, we have that $A^t = A$. Consequently, it follows that $(AX) \cdot Y = (AX)^t Y = X^t A^t Y = X^t AY = X \cdot (AY)$ for all $n \times 1$ column vectors X and Y by definition of the dot product. Conversely, we will assume that $(AX) \cdot Y = X \cdot (AY)$ for all $n \times 1$ column vectors X and Y. By definition of the dot product and by hypothesis, we have that $X^t A^t Y = (AX)^t Y = (AX) \cdot Y = X \cdot (AY) = X^t AY$. Considering that this holds for all $n \times 1$ column vectors Y, we may substitute the standard basis vectors in place of Y to conclude that $X^t A^t = X^t A$ for all $n \times 1$ column vectors X. By performing the same substitution with the standard basis vectors in place of X, we conclude that $A^t = A$. □

We will say that two real $n \times 1$ column vectors are **orthogonal** if and only if $X \cdot Y = O$. We have already tacitly encountered examples of orthogonal vectors such as the standard basis vectors E_i and E_j for any distinct positive integers i and j. Let us return to the example for a moment.

Example 2.6.4. (Example 2.6.2, Cont'd) We found previously the following eigenvectors.

$$X_{-3} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \qquad X_{-1} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \qquad X_6 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

By definition of the dot product, in order to find the lone component of the 1×1 matrix $X_i \cdot X_j$, we simply take the product of each row of X_i by the row of X_j, and we add up all of these values over all rows. Consequently, we have that $X_{-3} \cdot X_{-1} = \begin{bmatrix}(1)(1) + (1)(-1) + (-2)(0)\end{bmatrix} = \begin{bmatrix}0\end{bmatrix}$ and $X_{-3} \cdot X_6 = \begin{bmatrix}(1)(1) + (1)(1) + (-2)(1)\end{bmatrix} = \begin{bmatrix}0\end{bmatrix}$ and $X_{-1} \cdot X_6 = \begin{bmatrix}(1)(1) + (1)(-1) + (0)(1)\end{bmatrix} = \begin{bmatrix}0\end{bmatrix}$. Consequently, the eigenvectors of the real symmetric 3×3 matrix A are orthogonal.

Our next proposition demonstrates that this is a general property of symmetric matrices.

Proposition 2.6.5. *Given any $n \times n$ symmetric matrix and any pair of eigenvectors X_1 and X_2 corresponding respectively to the distinct eigenvalues α_1 and α_2 of A, we have that $X_1 \cdot X_2 = O$. Put another way, eigenvectors belonging to distinct eigenvalues of a symmetric matrix are orthogonal.*

Proof. By assumption that α_1 and α_2 are distinct eigenvalues, it follows that $\alpha_1 - \alpha_2$ is nonzero. Consequently, if we can establish that $(\alpha_1 - \alpha_2)(X_1 \cdot X_2) = O$, then we must have that $X_1 \cdot X_2 = O$. By Proposition 1.2.6, it follows that matrix multiplication is distributive, and by [Lan86, Exercise 6] on page 47, the transpose commutes with scalar multiplication, hence we have the following.

$$\begin{aligned} \alpha_1(X_1 \cdot X_2) &= \alpha_1 X_1^t X_2 = (\alpha_1 X)^t X_2 = (AX_1)^t X_2 = (AX_1) \cdot X_2 \\ \alpha_2(X_1 \cdot X_2) &= \alpha_2 X_1^t X_2 = X_1^t(\alpha_2 X_2) = X_1^t(AX_2) = X_1 \cdot (AX_2) \end{aligned}$$

By hypothesis that A is symmetric, we may apply Proposition 2.6.3 to conclude that $(AX_1) \cdot X_2 = X_1 \cdot (AX_2)$, from which it follows that $\alpha_1(X_1 \cdot X_2) = \alpha_2(X_1 \cdot X_2)$ and $(\alpha_1 - \alpha_2)(X_1 \cdot X_2) = O$. □

Consequently, if we can demonstrate that any real symmetric matrix A is diagonalizable, then it will follow from Proposition 2.6.5 that A is **orthogonally diagonalizable**, by which we mean that there exists a basis of $\mathbb{R}^{n \times 1}$ consisting of eigenvectors for A that are pairwise orthogonal. Equivalently, one can define the condition of orthogonal diagonalizability in terms of matrices. We say that a real $n \times n$ matrix Q is **orthogonal** if it holds that $QQ^t = I = Q^t Q$. Put another way, a real $n \times n$ matrix Q is orthogonal if and only if it is invertible and its transpose Q^t is its matrix inverse. Consequently, we will say that a real $n \times n$ matrix A is orthogonally diagonalizable if and only if there exists a real orthogonal $n \times n$ matrix Q such that QAQ^t is a diagonal matrix. By this identification, we can prove one direction of the Spectral Theorem.

Proposition 2.6.6. *Every real orthogonally diagonalizable matrix is symmetric.*

Proof. By definition, if A is a real $n \times n$ matrix that is orthogonally diagonalizable, then there exists a real orthogonal $n \times n$ matrix Q such that QAQ^t is a diagonal matrix. Observe that the transpose of a diagonal matrix is itself, hence we have that $QAQ^t = (QAQ^t)^t = (Q^t)^t A^t Q^t = QA^tQ^t$. Considering that Q and Q^t are invertible matrices, we may "cancel" them on the left- and right-hand sides of this identity (by multiplying by their matrix inverses) to conclude that $A = A^t$. $\square$

Only the implication of the Spectral Theorem remains to be seen. We will not prove this here, but we will prove a necessary lemma that the eigenvalues of a real symmetric $n \times n$ matrix are always real numbers. We should point out that this is an inextricable property of real symmetric matrices that is an important fact in its own right and stands in sharp contrast to the general situation with non-symmetric real matrices: the following 2×2 matrix does not have a real eigenvalue!

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

Explicitly, the characteristic polynomial of this matrix is $x^2 + 1$, and we have that c is a root of $x^2 + 1$ if and only if $c^2 + 1 = 0$ if and only if $c^2 = -1$ if and only if $c = \pm\sqrt{-1}$. Conventionally, we write $i = \sqrt{-1}$, and we point out that this is not a real number because the square of every real number is a non-negative real number; in fact, we say that i is an **imaginary number**. We refer to the set $\mathbb{C} = \{a + bi \mid a \text{ and } b \text{ are real numbers and } i = \sqrt{-1}\}$ as the **complex numbers**. Consequently, we may view i itself as a complex number. We distinguish the real number a of the complex number $a + bi$ as the **real part** of $a + bi$, and the real number b is the **imaginary part** of $a + bi$. Complex numbers admit a notion of addition that allow us to view $\mathbb{C}$ as the two-dimensional real vector space $\mathbb{C} = \operatorname{span}\{1, i\}$. Explicitly, we define $(a + bi) + (c + di) = (a + b) + (c + d)i$ as per the usual addition of vectors with respect to a basis. Consequently, the zero vector of $\mathbb{C}$ is $0 + 0i$. We define multiplication of complex numbers by "foiling" a product of complex numbers as follows.

$$(a + bi)(c + di) = ac + adi + bci + (bi)(di) = (ac - bd) + (ad + bc)i$$

Even more, if a and b are nonzero real numbers, then $a + bi$ and $a - bi$ are nonzero complex numbers, and we have that $(a + bi)(a - bi) = a^2 + b^2$ is a nonzero real number. We refer to the complex number $a - bi$ as the **complex conjugate** of $a + bi$, and the real number $\sqrt{a^2 + b^2} = (a + bi)(a - bi)$ is called the **modulus** of $a + bi$. Often, authors throughout the literature will denote $z = a + bi$; its complex conjugate $\bar{z} = a - bi$; and its modulus $|z| = \sqrt{a^2 + b^2}$. Crucially, we have that $|z|^2 = a^2 + b^2 = z\bar{z}$, and for any pair of complex numbers z_1 and z_2, it follows that $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$.

Recall that a **root** of a polynomial $\alpha_n x^n + \cdots + \alpha_1 x + \alpha_0$ with complex coefficients $\alpha_0, \alpha_1, \ldots, \alpha_n$ is a complex number z such that $\alpha_n z^n + \cdots + \alpha_1 z + \alpha_0 = 0$. Even though it is a classical theorem of algebra, the following is typically proved using complex analysis. Consequently, we will not attempt in this course to supply any justification ourselves; we will simply take it for granted.

Theorem 2.6.7 (Fundamental Theorem of Algebra)**.** *Let n be a positive integer. Every polynomial $p(x)$ of degree n with complex coefficients has exactly n (not necessarily distinct) roots.*

Consequently, the polynomial equation $z^3 = 1$ has exactly three solutions over the complex numbers. Certainly, one solution is simply $z = 1$; however, the other two solutions have nonzero imaginary part. Explicitly, we may factor $x^3 - 1 = (x-1)(x^2+x+1)$ such that x^2+x+1 has no real roots because the discriminant $b^2 - 4ac$ of the Quadratic Formula is negative.

We are now ready to prove the following indispensable fact about real symmetric matrices.

Theorem 2.6.8. *Every eigenvalue of a real symmetric matrix is a real number.*

Proof. Converting the above statement into symbols, we need to prove that if A is a real symmetric $n \times n$ matrix and α is an eigenvalue of A corresponding to an eigenvector X for A, then α is a real number. Unfortunately, it is not clear a priori that the eigenvector X for A corresponding to α is a real $n \times 1$ column vector; however, we may assume that its entries are all complex numbers. Explicitly, we will assume for the moment that $X = (z_1, \dots, z_n)^t$ for some complex numbers $z_1, \dots, z_n$. Consider the column vector $\bar{X} = (\bar{z}_1, \dots, \bar{z}_n)^t$ consisting of the complex conjugates of the components of X. By definition of the dot product, it follows that $X \cdot \bar{X} = X^t \bar{X} = \begin{bmatrix} z_1\bar{z}_1 + \cdots + z_n\bar{z}_n \end{bmatrix}$. Each of the products $z_i\bar{z}_i$ is a non-negative real number, hence $X \cdot \bar{X}$ is a nonzero 1×1 matrix. Complex multiplication is commutative, hence we have that $z_i\bar{z}_i = \bar{z}_i z_i$ for all integers $1 \leq i \leq n$, and the same argument used to compute the dot product as before shows that $\bar{X} \cdot X = X \cdot \bar{X}$. Considering that X is an eigenvector for A corresponding to the eigenvalue α, it holds that

$$(AX) \cdot \bar{X} = (\alpha X) \cdot \bar{X} = (\alpha X)^t \bar{X} = \alpha X^t \bar{X} = \alpha(X \cdot \bar{X}).$$

By assumption that A is a real matrix, complex conjugation does not affect its entries. Put another way, if we denote by $\bar{A}$ the matrix obtained from A by taking the complex conjugate of each of its entries, then we have that $\bar{A} = A$. Observe that $\overline{AX}$ is by definition the $n \times 1$ column vector obtained from the $n \times 1$ column vector AX by taking the complex conjugate of each of its entries. Complex conjugates satisfies that $\overline{z_1 z_2} = \bar{z}_1\bar{z}_2$ for any pair of complex numbers z_1 and z_2, hence we find that $A\bar{X} = \bar{A}\bar{X} = \overline{AX} = \overline{\alpha X} = \bar{\alpha}\bar{X}$. On the level of the dot product, this gives the following.

$$X \cdot (A\bar{X}) = X \cdot (\bar{\alpha}\bar{X}) = X^t(\bar{\alpha}\bar{X}) = \bar{\alpha}(X^t\bar{X}) = \bar{\alpha}(X \cdot \bar{X})$$

By Proposition 2.6.3, we conclude that $\alpha(X \cdot \bar{X}) = (AX) \cdot \bar{X} = X \cdot (A\bar{X}) = \bar{\alpha}(X \cdot \bar{X})$ by assumption that A is a symmetric matrix. Consequently, we find that $(\alpha - \bar{\alpha})(X \cdot \bar{X}) = O$; then, using the fact that $X \cdot \bar{X}$ is a nonzero matrix, we conclude that $\alpha - \bar{\alpha} = 0$ so that $\alpha = \bar{\alpha}$. Last, if we write $\alpha = a + bi$, then we have shown that $a + bi = \alpha = \bar{\alpha} = a - bi$ so that $2bi = 0$ and $b = 0$. □

Often, the implication of the Spectral Theorem is stated throughout the literature as follows.

Theorem 2.6.9 (Principal Axis Theorem)**.** *Every real symmetric matrix is orthogonally diagonalizable. Explicitly, if A is a real symmetric matrix, then there exists an orthogonal matrix Q such that QAQ^t is the diagonal matrix whose diagonal entries are the eigenvalues of A.*

Even though we will not complete the proof here, we remark that the argument is made by induction on the number of rows and columns of the real symmetric matrix in question. One can find a proof using the ingredients present in these notes in either [McK22] or [Smi17].

2.7 Nilpotent Matrices and Cyclic Subspaces

We have seen previously that diagonal matrices are the simplest kinds of matrices (other than scalar matrices), hence it has been our hope that every linear transformation from a finite-dimensional vector space to itself could be represented by a diagonal matrix. Unfortunately, this is not possible by Example 2.5.18: even a matrix as inconspicuous as the one obtained from the zero matrix by replacing some non-diagonal component by one cannot be converted to a diagonal matrix; however, this matrix is itself **triangular**, hence it is natural to study triangular matrices. Explicitly, an **upper-triangular** matrix is a square matrix whose components below the main diagonal are zero. Conversely, we say that a matrix is **lower-triangular** if it is the transpose of an upper-triangular matrix. Considering that the determinant of a matrix is equal to the determinant of its transpose and that the characteristic polynomial of a matrix is therefore equal to the characteristic matrix of its transpose, we may strictly fix our attention on the upper-triangular matrices.

Our first order of business is to establish that the determinant of an upper-triangular matrix is the product of its diagonal components; this affords us a simple way to compute the characteristic polynomial of an upper-triangular matrix because its characteristic matrix is also upper-triangular.

Proposition 2.7.1. *The determinant of a triangular matrix is the product of its diagonal entries.*

Proof. Considering that a lower-triangular matrix is the transpose of an upper-triangular matrix and the determinant of a matrix is equal to the determinant of its transpose by Proposition 2.1.12, we may prove the claim for upper-triangular matrices. We proceed by induction on the size n of an $n \times n$ upper-triangular matrix A. Every 2×2 diagonal matrix is of the following form.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}$$

Consequently, we have that $\det(A) = a_{11}a_{22}$, as desired. We will assume by induction that the claim holds for $(n-1) \times (n-1)$ upper-triangular matrices. Consider the following $n \times n$ matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

Expanding the determinant along the first column, we obtain the following identity.

$$\det(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix}$$

Considering that the determinant on the right-hand side is taken from an $(n-1) \times (n-1)$ matrix, it follows by our inductive hypothesis that $\det(A) = a_{11}a_{22} \cdots a_{nn}$ is the product of the diagonal. $\square$

Corollary 2.7.2. *Given any triangular $n \times n$ matrix A whose diagonal components are $a_1, \dots, a_n$, the characteristic polynomial of A is given by $\chi_A(x) = (x - a_1) \cdots (x - a_n)$.*

Proof. Considering that xI is a diagonal matrix, it follows that $xI - A$ is a triangular matrix because the difference does not affect any components of A other than those lying on the diagonal of A. Observe that the diagonal components of $xI - A$ are simply the linear polynomials $x - a_1, \dots, x - a_n$, hence by Proposition 2.7.1, we conclude that $\chi_A(x) = \det(xI - A) = (x - a_1) \cdots (x - a_n)$. $\square$

We return our attention to the non-diagonalizable matrix A of Example 2.5.18. Observe that the characteristic polynomial of A is $\chi_A(x) = x^3$; however, its minimal polynomial is $\mu_A(x) = x^2$. Consequently, the matrix A admits an eigenvalue 0 of **algebraic multiplicity** three because the power of the linear factor $x - 0$ in the characteristic polynomial of A is three. On the other hand, the **geometric multiplicity** of the eigenvalue 0 is two, i.e., the dimension of the eigenspace W_0 of V corresponding to the eigenvalue 0 is two. Both of these observations lead us to study square matrices whose powers eventually all result in the zero matrix. We will say that an $n \times n$ matrix A is **nilpotent** if there exists an integer $k \geq 1$ such that $A^k = O$. Likewise, we will say that a linear transformation $T : V \to V$ from a vector space V to itself is nilpotent if there exists an integer $k \geq 1$ such that the k-fold composite transformation $T^k : V \to V$ is the zero transformation.

Proposition 2.7.3. *Given any $n \times n$ matrix A, the following properties are equivalent.*

1.) *We have that $A^k = O$ and $A, A^2, \dots, A^{k-1}$ are nonzero.*

2.) *The minimal polynomial of A is $\mu_A(x) = x^k$.*

3.) *The characteristic polynomial of A is $\chi_A(x) = x^n$.*

Proof. By Proposition 2.3.9, if $A^k = O$ for some positive integer k, then the minimal polynomial $\mu_A(x)$ of A divides x^k. Consequently, we must have that $\mu_A(x)$ is a power of x because the only polynomials that divide x^k are $x, x^2, \dots, x^k$. Even more, by assumption that the matrices $A, A^2, \dots, A^{k-1}$ are nonzero, it follows that $\mu_A(x) = x^k$ by assumption that k is the least positive integer for which $A^k = O$. By Proposition 2.3.14, we conclude that if $\mu_A(x) = x^k$ for some integer $1 \leq k \leq n$, then $\chi_A(x) = x^n$ because $\mu_A(x)$ and $\chi_A(x)$ have the same irreducible factors. Last, if the characteristic polynomial of A is x^n, then A is nilpotent by the Cayley-Hamilton Theorem. $\square$

Example 2.7.4. Consider the following real 2×2 matrix A.

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

Observe that $A^2 = O$, hence A is a nonzero nilpotent matrix.

Example 2.7.5. Consider the following real 3×3 matrix A.

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

Observe that A and A^2 are nonzero and $A^3 = O$, hence A is a nonzero nilpotent matrix.

We refer to the positive integer k defined in the first and second parts of Proposition 2.7.3 as the **index of nilpotency** (or **degree of nilpotency**) of the nilpotent $n \times n$ matrix A. Crucially, observe that the only eigenvalue of a nilpotent matrix is zero. By Definition 2.5.7, a real nilpotent $n \times n$ matrix is diagonalizable if and only if there exists a basis of $\mathbb{R}^{n \times 1}$ consisting of eigenvectors for 0 if and only if we have that $AE_i = O$ for all standard basis vectors $E_1, \dots, E_n$ of $\mathbb{R}^{n \times 1}$ if and only if A is the $n \times n$ zero matrix. Consequently, by this observation, we obtain the following proposition.

Proposition 2.7.6. *The only diagonalizable nilpotent $n \times n$ matrix is the $n \times n$ zero matrix.*

Even though no nonzero nilpotent matrix is diagonalizable, every nilpotent matrix admits an upper-triangular matrix representation for which the diagonal components are all zeros.

Theorem 2.7.7. *Every nilpotent $n \times n$ matrix admits an upper-triangular matrix representation. Even more, the diagonal components of such a matrix representation are all zeros.*

We devote the remainder of this section to the proof of the aforementioned theorem. Consider any nonzero vector v of any nonzero vector space V. Given any linear transformation $T : V \to V$ from V to itself, we refer to the collection of vectors of V of the form $\alpha_n T^n(v) + \cdots + \alpha_1 T(v) + \alpha_0 v$ for some scalars $\alpha_n \dots, \alpha_1, \alpha_0$ as the T-**cyclic subspace** of V generated by the vector v. Explicitly, the T-cyclic subspace $C(T, v)$ of V generated by v is the span of all vectors of the form $T^k(v)$ for some integer $k \geq 0$. Consequently, it is by definition a vector subspace of V.

Example 2.7.8. Observe that the T-cyclic subspace of V generated by the zero vector of V is simply the zero subspace of V: indeed, we have that $T^k(O) = O$ for all integers $k \geq 0$.

Example 2.7.9. Given any eigenvector v of T corresponding to an eigenvalue α of T, we have that $T(v) = \alpha v$. Consequently, for any integer $k \geq 1$, we have that $T^k(v) = \alpha^k v$, and the T-cyclic subspace of V generated by the eigenvector v is nothing more than the span of v.

Example 2.7.10. Let us compute the T-cyclic subspace of $\mathbb{R}^{1 \times 3}$ generated by $X = (1, 1, 1)$ with respect to the linear transformation $T : \mathbb{R}^{1 \times 3} \to \mathbb{R}^{1 \times 3}$ defined by $T(x, y, z) = (x, -y, z)$. Observe that $T(X) = T(1, 1, 1) = (1, -1, 1)$ and $T^2(X) = T(T(X)) = T(1, -1, 1) = (1, 1, 1) = X$. Consequently, the T-cyclic subspace of $\mathbb{R}^{1 \times 3}$ generated by X is given by $C(T, X) = \operatorname{span}\{(1, 1, 1), (1, -1, 1)\}$.

Example 2.7.11. Consider the T-cyclic subspace of $\mathbb{R}^{3 \times 1}$ generated by the vector $X^t = (1, 1, 1)^t$ with respect to the linear transformation $T : \mathbb{R}^{3 \times 1} \to \mathbb{R}^{3 \times 1}$ with the following matrix representation.

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

By definition, the image $T(X^t)$ is given by the matrix product AX^t as follows.

$$AX^t = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$$

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

We claim that the column vectors X^t, AX^t, and A^2X^t are linearly independent, hence they form a basis for the T-cyclic subspace of $\mathbb{R}^{3\times 1}$ generated by X^t. By Algorithm 1.7.9, it suffices to check that the real 3×3 matrix with columns X^t, AX^t, and A^2X^t has three pivots.

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & -2 \\ 1 & 1 & 0 \end{bmatrix} \overset{(1.)}{\sim} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -5 \\ 0 & -1 & -3 \end{bmatrix} \overset{(2.)}{\sim} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -2 & -5 \\ 0 & 0 & -\frac{1}{2} \end{bmatrix}$$

(1.) We employ the elementary row operations $R_2 - R_1 \mapsto R_2$ and $R_3 - R_1 \mapsto R_3$.

(2.) We employ the elementary row operation $R_3 - \frac{1}{2}R_2 \mapsto R_3$.

Consequently, the vectors X^t, AX^t, and A^2X^t are linearly independent, hence the T-cyclic subspace of $\mathbb{R}^{3\times 1}$ generated by X^t is given by $C(T, X^t) = \operatorname{span}\{(1,1,1)^t, (2,0,1)^t, (3,-2,0)^t\}$.

One of the foremost advantages of working with the T-cyclic subspace $C(T, v)$ of a vector space V generated by a nonzero vector v is that the matrix representation of T as a linear transformation from $C(T, v)$ to itself is especially simple to describe, as our next proposition guarantees.

Proposition 2.7.12. *Consider any linear transformation $T : V \to V$ from any nonzero finite-dimensional vector space V to itself.*

1.) *There exists a nonzero vector $v \in V$ and a positive integer n such that the T-cyclic subspace $C(T, v)$ of V generated by v is spanned by the vectors $v, T(v), T^2(v), \dots, T^{n-1}(v)$.*

2.) *Even more, the positive integer n from the previous part can be chosen such that the vectors $v, T(v), T^2(v), \dots, T^{n-1}(v)$ are linearly independent, hence they form a basis for $C(T, v)$.*

3.) *We may view T as a linear transformation from $C(T, v)$ to itself such that the matrix representation of T with respect to the ordered basis $v, T(v), T^2(v), \dots, T^{n-1}(v)$ is as follows.*

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & -\alpha_1 \\ 0 & 1 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\alpha_{n-1} \end{bmatrix}$$

We refer to this as the **companion matrix** *of $x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2x^2 + \alpha_1x + \alpha_0$.*

Proof. (1.) By definition, for any nonzero vector $v \in V$, the T-cyclic subspace $C(T, v)$ of V generated by v is spanned by the vectors $T^k(v)$ for all integers $k \geq 0$; however, by assumption that V has finite dimension, there exists an integer $n \geq 1$ such that $v, T(v), T^2(v), \dots, T^n(v)$ are linearly dependent. Consequently, there exist scalars $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{n-1}$ (not all of which are zero) such that

$$T^n(v) + \alpha_{n-1}T^{n-1}(v) + \cdots + \alpha_2T^2(v) + \alpha_1T(v) + \alpha_0v = O.$$

By rearranging this expression of linear dependence, it follows that

$$T^n(v) = -\alpha_{n-1}T^{n-1}(v) - \cdots - \alpha_2T^2(v) - \alpha_1T(v) - \alpha_0v.$$

Given any integer $k \geq n$, observe that $T^k(v) = T^{k-n}(T^n(v))$ is obtained by applying the linear transformation T to the above expression $k - n$ times. By the linearity of T, every vector $T^k(v)$ with $k \geq n$ can be written as a linear combination of $v, T(v), T^2(v), \dots, T^{n-1}(v)$, as desired.

(2.) We must prove next that the positive integer n from the previous part can be chosen such that the vectors $v, T(v), T^2(v), \dots, T^{n-1}(v)$ are linearly independent. Considering that v is a nonzero vector of V, it follows that v is linearly independent. Consequently, it suffices to check that v and $T(v)$ are linearly independent. Observe that if $T(v) = O$, then $T^k(v) = O$ for all integers $k \geq 1$. Consequently, $C(T, v)$ is spanned by v, hence v forms a basis for $C(T, v)$. Otherwise, we may assume that $T(v)$ is nonzero. We note in this case that if v and $T(v)$ are linearly dependent, then there exists a nonzero scalar α such that $T(v) + \alpha v = O$ or $T(v) = -\alpha v$. Consequently, for each integer $k \geq 1$, we have that $T^k(v) = (-1)^k \alpha v$ so that once again, $C(T, v)$ is spanned by v. Given that neither $T(v) = O$ nor $T(v) = -\alpha v$, it follows that v and $T(v)$ are linearly independent. Continuing in this manner, we may find the smallest positive integer n for which $v, T(v), T^2(v), \dots, T^{n-1}(v)$ are linearly independent; by the previous part, these vectors form a basis for $C(T, v)$.

(3.) By the first part of the proof, we note that T can be regarded as a linear transformation from $C(T, v)$ to itself. Considering that $v, T(v), T^2(v), \dots, T^{n-1}(v)$ form an ordered basis of this vector space, we may consider the matrix representation of T with respect to this ordered basis.

$$\begin{aligned}
T(v) &= 0 \cdot v + 1 \cdot T(v) + 0 \cdot T^2(v) + \cdots + 0 \cdot T^{n-1}(v) \\
T(T(v)) = T^2(v) &= 0 \cdot v + 0 \cdot T(v) + 1 \cdot T^2(v) + \cdots + 0 \cdot T^{n-1}(v) \\
&\ \ \vdots \\
T(T^{n-2}(v)) = T^{n-1}(v) &= 0 \cdot v + 0 \cdot T(v) + 0 \cdot T^2(v) + \cdots + 1 \cdot T^{n-1}(v) \\
T(T^{n-1}(v)) = T^n(v) &= -\alpha_0 v - \alpha_1 T(v) - \alpha_2 T^2(v) - \cdots - \alpha_{n-1} T^{n-1}(v)
\end{aligned}$$

Certainly, this gives rise to the desired matrix representation of T. □

Even more, the characteristic polynomial and minimal polynomial of the companion matrix of the polynomial $p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ are both $p(x)$ itself!

Proposition 2.7.13. *Given any scalars $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{n-1}$, consider the following $n \times n$ matrix.*

$$C = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & -\alpha_1 \\ 0 & 1 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\alpha_{n-1} \end{bmatrix}$$

We have that $\mu_C(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$. Put another way, the characteristic polynomial and the minimal polynomial of the companion matrix of $p(x)$ are both $p(x)$.

We reserve the proof of this fact for later when we discuss companion matrices; however, we note that the result is useful because it allows us to cook up matrices with any desired characteristic and minimal polynomials. Back in our present discussion of nilpotent matrices, we are able to apply the previous two propositions to obtain the desired matrix representation of a nilpotent matrix.

Proposition 2.7.14. *Given any nilpotent linear transformation $T : V \to V$ from any nonzero finite-dimensional vector space V to itself, there exists a nonzero vector $v \in V$ such that the T-cyclic subspace $C(T, v)$ of V generated by v admits an ordered basis $v, T(v), T^2(v), \dots, T^{n-1}(v)$ with respect to which the matrix transformation of $T : C(T, v) \to C(T, v)$ has the following form.*

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Explicitly, this matrix is lower-triangular with diagonal components of zero.

Proof. Considering that T is nilpotent, by Proposition 2.7.3, the minimal polynomial of T is given by $\mu_T(x) = x^k$ for some positive integer k not exceeding the dimension of V. By the proof of Proposition 2.7.12, we may consider T as a linear transformation from $C(T, v)$ to itself; the minimal polynomial $\mu(x)$ of the **restriction** of T to the T-cyclic subspace $C(T, v)$ of V generated by v divides any polynomial that annihilates T by Proposition 2.3.9, from which it follows that $\mu(x) = x^\ell$ for some integer $1 \leq \ell \leq k$. Consequently, we have that $T^\ell = O$ so that $T(T^{\ell-1}(v)) = T^\ell(v) = O$. On the other hand, the vectors $v, T(v), T^2(v), \dots, T^{\ell-1}(v)$ cannot be linearly dependent because this would imply that some polynomial $\alpha_{\ell-1}x^{\ell-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ annihilates $T(v)$ — a contradiction. We conclude that $v, T(v), T^2(v), \dots, T^{\ell-1}(v)$ form an ordered basis for $C(T, v)$, hence the matrix representation of T with respect to this ordered basis has the desired form by Proposition 2.7.12. ☐

2.8 The Smith Normal Form

We turn our attention next to an indispensable tool in the theory of canonical forms for matrices. Explicitly, we will construct a canonical form for the characteristic matrix $xI - A$ of a real $n \times n$ matrix that will allow us to determine the minimal polynomial and characteristic polynomial of A.

Theorem 2.8.1 (Smith Normal Form)**.** *Given any real $n \times n$ matrix A and any indeterminate x, there exist invertible $n \times n$ matrices P and Q and real polynomials $p_1(x), p_2(x), \dots, p_\ell(x)$ such that*

$$P(xI - A)Q = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & p_1(x) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & p_2(x) & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & p_\ell(x) \end{bmatrix}$$

and the polynomials $p_i(x)$ are unique (up to sign) and satisfy that $p_1(x) \mid p_2(x) \mid \cdots \mid p_\ell(x)$. Even more, the non-constant polynomials are called **invariant factors***; the minimal polynomial of A is the largest invariant factor $p_\ell(x)$; and the characteristic polynomial of A is $p_1(x)p_2(x) \cdots p_\ell(x)$.*

Computing the Smith Normal Form for the characteristic matrix $xI - A$ of a real $n \times n$ matrix A amounts to carrying out some elementary row operations and elementary column operations on $xI - A$ to reduce the given matrix to the desired form. Explicitly, we will find that the invertible $n \times n$ matrix P is obtained from the $n \times n$ identity matrix by performing the specified elementary row operations on $xI - A$; likewise, the invertible $n \times n$ matrix Q is obtained from the $n \times n$ identity matrix by performing the specified elementary column operations on $xI - A$. We note that there are three elementary row (or column) operations that are valid in this scenario.

1.) We may multiply any row (or column) of the matrix by a nonzero real number a.

2.) We may add any polynomial multiple of a row (or column) to another row (or column).

3.) We may interchange any pair of rows (or columns) of the matrix.

We continue using the shorthand $R_i \mapsto aR_i$ to denote the operation of multiplying the ith row of the matrix by a; we will use the shorthand $R_j + p(x)R_i \mapsto R_j$ to denote the operation of adding a polynomial multiple $p(x)$ of the ith row of the matrix to the jth row of the matrix (for any distinct indices i and j); and we will use the shorthand $R_i \leftrightarrow R_j$ to denote the operation of interchanging the ith and jth rows of the matrix. Each of these elementary row operations can also be performed with the ith and jth columns C_i and C_j of the matrix for any pair of distinct indices i and j.

Example 2.8.2. Let us compute the Smith Normal Form for $xI - A$ of the following 2×2 matrix.

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

We will keep track of the elementary row operations by performing each such operation on the 2×2 identity matrix; likewise, we will keep track of the column operations by manipulating the columns of the 2×2 identity matrix according to the column operations on $xI - A$.

$$xI - A = \begin{bmatrix} x-1 & 0 \\ -1 & x+1 \end{bmatrix}$$

1.) $C_2 + (x+1)C_1 \mapsto C_2$ $\qquad xI - A \sim \begin{bmatrix} x-1 & (x-1)(x+1) \\ -1 & 0 \end{bmatrix} \qquad Q \sim \begin{bmatrix} 1 & x+1 \\ 0 & 1 \end{bmatrix}$

2.) $R_1 \leftrightarrow R_2$ $\qquad xI - A \sim \begin{bmatrix} -1 & 0 \\ x-1 & (x-1)(x+1) \end{bmatrix} \qquad P \sim \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

3.) $R_2 + (x-1)R_1 \mapsto R_2$ $\qquad xI - A \sim \begin{bmatrix} -1 & 0 \\ 0 & (x-1)(x+1) \end{bmatrix} \qquad P \sim \begin{bmatrix} 0 & 1 \\ 1 & x-1 \end{bmatrix}$

4.) $-R_1 \mapsto R_1$ $\qquad xI - A \sim \begin{bmatrix} 1 & 0 \\ 0 & (x-1)(x+1) \end{bmatrix} \qquad P \sim \begin{bmatrix} 0 & -1 \\ 1 & x-1 \end{bmatrix}$

Consequently, the Smith Normal Form for $xI-A$ and the invertible matrices P and Q are as follows.

$$\text{SNF}(xI-A) = P(xI-A)Q = \begin{bmatrix} 0 & -1 \\ 1 & x-1 \end{bmatrix} \begin{bmatrix} x-1 & 0 \\ -1 & x+1 \end{bmatrix} \begin{bmatrix} 1 & x+1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & (x-1)(x+1) \end{bmatrix}$$

Even more, the only invariant factor of A is $(x-1)(x+1)$, hence we have that $\mu_A(x) = (x-1)(x+1)$ and $\chi_A(x) = (x-1)(x+1)$. Last, the **elementary divisors** of A are $x-1$ and $x+1$.

Example 2.8.3. Let us compute the Smith Normal Form for $xI - A$ of the following 2×2 matrix.

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

We will keep track of the elementary row operations by performing each such operation on the 2×2 identity matrix; likewise, we will keep track of the column operations by manipulating the columns of the 2×2 identity matrix according to the column operations on $xI - A$.

$$xI - A = \begin{bmatrix} x & -1 \\ 0 & x \end{bmatrix}$$

1.) $C_1 \leftrightarrow C_2$ $\quad xI - A \sim \begin{bmatrix} -1 & x \\ x & 0 \end{bmatrix}$ $\quad Q \sim \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

2.) $R_2 + xR_1 \mapsto R_2$ $\quad xI - A \sim \begin{bmatrix} -1 & x \\ 0 & x^2 \end{bmatrix}$ $\quad P \sim \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix}$

3.) $C_2 + xC_1 \mapsto C_2$ $\quad xI - A \sim \begin{bmatrix} -1 & 0 \\ 0 & x^2 \end{bmatrix}$ $\quad Q \sim \begin{bmatrix} 0 & 1 \\ 1 & x \end{bmatrix}$

4.) $-R_1 \mapsto R_1$ $\quad xI - A \sim \begin{bmatrix} 1 & 0 \\ 0 & x^2 \end{bmatrix}$ $\quad P \sim \begin{bmatrix} -1 & 0 \\ x & 1 \end{bmatrix}$

Consequently, the Smith Normal Form for $xI-A$ and the invertible matrices P and Q are as follows.

$$\text{SNF}(xI-A) = P(xI-A)Q = \begin{bmatrix} -1 & 0 \\ x & 1 \end{bmatrix} \begin{bmatrix} x & -1 \\ 0 & x \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & x \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & x^2 \end{bmatrix}$$

Even more, the only invariant factor of A is x^2, hence the minimal polynomial and the characteristic polynomial of A are $\mu_A(x) = x^2$ and $\chi_A(x) = x^2$. Last, the only elementary divisor of A is x^2.

Going forward into the case of 3×3 matrices, out of want for simplicity, we will not concern ourselves with keeping track of the matrices P and Q; however, we note that (somewhat miraculously) in order to determine the invertible matrix P that converts A to its **Rational Canonical Form** or **Jordan Canonical Form**, it suffices to keep track *only* of the elementary row operations.

Example 2.8.4. Let us compute the Smith Normal Form for $xI - A$ of the following 3×3 matrix.

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

We will keep track of the elementary row operations and often abbreviate column operations.

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

1.) $C_1 \leftrightarrow C_2$
$$xI - A \sim \begin{bmatrix} -1 & x-1 & -1 \\ x-2 & -2 & -2 \\ -3 & -3 & x-3 \end{bmatrix}$$

2.) $R_2 + (x-2)R_1 \mapsto R_2$
$$xI - A \sim \begin{bmatrix} -1 & x-1 & -1 \\ 0 & (x-1)(x-2)-2 & -(x-2)-2 \\ -3 & -3 & x-3 \end{bmatrix}$$

3.) $R_3 - 3R_1 \mapsto R_3$
$$xI - A \sim \begin{bmatrix} -1 & x-1 & -1 \\ 0 & (x-1)(x-2)-2 & -(x-2)-2 \\ 0 & -3(x-1)-3 & x \end{bmatrix}$$

Perform column operations and simplify the result.
$$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & x(x-3) & -x \\ 0 & -3x & x \end{bmatrix}$$

4.) $C_2 + (x-3)C_3 \mapsto C_2$
$$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -x \\ 0 & -3x + x(x-3) & x \end{bmatrix}$$

5.) $R_3 + R_2 \mapsto R_3$
$$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -x \\ 0 & x(x-6) & 0 \end{bmatrix}$$

6.) $C_2 \leftrightarrow C_3$, $-C_2 \mapsto C_2$
$$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x(x-6) \end{bmatrix}$$

We note that this last matrix is by definition the Smith Normal Form for $xI - A$. Consequently, the invariant factors of A are x and $x(x-6)$; the elementary divisors of A are x, x, and $x-6$; the minimal polynomial of A is $\mu_A(x) = x(x-6)$; and the characteristic polynomial of A is $\chi_A(x) = x^2(x-6)$.

Example 2.8.5. Let us compute the Smith Normal Form for $xI - A$ of the following 3×3 matrix.

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

We will keep track of the elementary row operations and often abbreviate column operations; however, it is possible here to get away almost entirely with using column operations.

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

1.) $R_3 + \dfrac{1}{2}(x-1)R_1 \mapsto R_3$
$$xI - A \sim \begin{bmatrix} x-1 & 0 & -2 \\ 0 & x-1 & 0 \\ \frac{1}{2}(x-1)^2 & 0 & 0 \end{bmatrix}$$

2.) $C_1 \leftrightarrow C_3$
$$xI - A \sim \begin{bmatrix} -2 & 0 & x-1 \\ 0 & x-1 & 0 \\ 0 & 0 & \frac{1}{2}(x-1)^2 \end{bmatrix}$$

3.) $-\dfrac{1}{2}C_1 \mapsto C_1$
$$xI - A \sim \begin{bmatrix} 1 & 0 & x-1 \\ 0 & x-1 & 0 \\ 0 & 0 & \frac{1}{2}(x-1)^2 \end{bmatrix}$$

4.) $C_3 - (x-1)C_3 \mapsto C_3$
$$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & x-1 & 0 \\ 0 & 0 & \frac{1}{2}(x-1)^2 \end{bmatrix}$$

5.) $2C_3 \mapsto C_3$
$$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & x-1 & 0 \\ 0 & 0 & (x-1)^2 \end{bmatrix}$$

We note that this last matrix is the Smith Normal Form for $xI - A$. Consequently, the invariant factors of A are $x-1$ and $(x-1)^2$; the elementary divisors of A are $x-1$ and $(x-1)^2$; the minimal polynomial of A is $\mu_A(x) = (x-1)^2$; and the characteristic polynomial of A is $\chi_A(x) = (x-1)^3$.

Example 2.8.6. Observe that the characteristic matrix of the $n \times n$ zero matrix O is simply the $n \times n$ matrix xI. Consequently, the Smith Normal Form for the characteristic matrix of the $n \times n$ zero matrix is the diagonal matrix consisting of n copies of x along the main diagonal. Particularly, the invariant factors and the elementary divisors of O are $x, x, \dots, x$ (n times); the minimal polynomial of O is $\mu_O(x) = x$; and the characteristic polynomial of O is $\chi_O(x) = x^n$.

Example 2.8.7. Observe that the characteristic matrix of the $n \times n$ identity matrix I is the matrix $(x-1)I$. Consequently, the Smith Normal Form for the characteristic matrix of the $n \times n$ identity matrix is the diagonal matrix consisting of n copies of $x-1$ along the main diagonal. Particularly, the invariant factors and the elementary divisors of I are $x-1, x-1, \dots, x-1$ (n times); the minimal polynomial of I is $\mu_I(x) = x-1$; and the characteristic polynomial of I is $\chi_I(x) = (x-1)^n$.

We will find that the Rational Canonical Form for A is built out of the invariant factors of A; similarly, the Jordan Canonical Form for A is built out of the **elementary divisors** of A. By definition, the elementary divisors of A are the powers of the irreducible polynomial factors of the invariant factors of A. We have tacitly used this fact already, but let us do some more examples.

Example 2.8.8. Given that the invariant factors of a matrix A are $x-1$ and $(x-1)(x-2)$, the elementary divisors of A must be $x-1$, $x-1$, and $x-2$; this must be a 3×3 matrix with minimal polynomial $\mu_A(x) = (x-1)(x-2)$ and characteristic polynomial $\chi_A(x) = (x-1)^2(x-2)$.

Example 2.8.9. Given that the invariant factors of a matrix A are x, x^2, and $x^3(x+1)^2$, the elementary divisors of A must be x, x^2, x^3, and $(x+1)^2$; this must be an 8×8 matrix with minimal polynomial $\mu_A(x) = x^3(x+1)^2$ and characteristic polynomial $\chi_A(x) = x^6(x+1)^2$.

Example 2.8.10. Observe that there cannot be a matrix with invariant factors $x-1$ and $x+1$ because neither of these linear polynomials divides the other. Explicitly, they have distinct roots.

We provide an algorithm for determining the elementary divisors from the invariant factors.

Algorithm 2.8.11 (Converting Invariant Factors to Elementary Divisors)**.** Let A be a real $n \times n$ matrix whose invariant factors are known. Use the following to find the elementary divisors of A.

1.) Given the invariant factors $p_i(x)$ with $p_1(x) \mid p_2(x) \mid \cdots \mid p_\ell(x)$, express each invariant factor $p_i(x)$ as a product of distinct prime powers of irreducible polynomials.

2.) Construct an upper-triangular array whose ith column consists of the distinct prime powers of irreducible polynomials $q_{i1}(x)^{e_{i1}}, \dots, q_{ik}(x)^{e_{ik}}$ such that $p_i(x) = q_{i1}(x)^{e_{i1}} \cdots q_{ik}(x)^{e_{ik}}$.

3.) We obtain the elementary divisors of A as the components of the upper-triangular array.

Example 2.8.12. By the previous algorithm, if A admits an invariant factor $x(x-1)^2(x^2+1)^3$, then the elementary divisors of A corresponding to this invariant factor are x, $(x-1)^2$, and $(x^2+1)^3$.

Conversely, it is possible to ask for the invariant factors from the elementary divisors. We provide an algorithm for this task; however, we note that it is slightly more delicate than the last.

Algorithm 2.8.13 (Converting Elementary Divisors to Invariant Factors)**.** Let A be a real $n \times n$ matrix whose elementary divisors are known. Use the following to find the invariant factors of A.

1.) Find the irreducible polynomial $p(x)$ that appears the most times among the elementary divisors of A. Choose one arbitrarily if more than one polynomial fits this criterion.

2.) Create an array whose first row consists of all powers of $p(x)$ that appear as elementary divisors of A, listing these powers in non-decreasing order from left to right.

3.) Repeat the second step in the second row with the irreducible polynomial $q(x)$ that appears the second most times among the elementary divisors of A.

4.) Continue this process until all irreducible polynomials appearing as elementary divisors of A have been written in a row. One should end with an upper-triangular array.

5.) By multiplying the elements of each consecutive column, we obtain the invariant factors of A.

Example 2.8.14. Given that the elementary divisors of a matrix A are x, x, x^2, x^3, $x-1$, x^2+1, and x^2+1, the previous algorithm leads us to the following upper-triangular array.

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

Consequently, the invariant factors of A are the products of the columns of this array, i.e., they are x, x, $x^2(x^2+1)$, and $x^3(x-1)(x^2+1)$. We conclude that A is a 12×12 matrix with minimal polynomial $\mu_A(x) = x^3(x-1)(x^2+1)$ and characteristic polynomial $\chi_A(x) = x^7(x-1)(x^2+1)^2$.

Example 2.8.15. Given that the elementary divisors of a matrix A are x^2, x^2, x^2+x+1, and x^2+x+1, the previous algorithm leads us to the following upper-triangular array.

$$\begin{array}{cc} x^2 & x^2 \\ x^2+x+1 & x^2+x+1 \end{array}$$

Consequently, the invariant factors of A are the products of the columns of this array, i.e., they are $x^2(x^2+x+1)$ and $x^2(x^2+x+1)$. We conclude that A is an 8×8 matrix with minimal polynomial $\mu_A(x) = x^2(x^2+x+1)$ and characteristic polynomial $\chi_A(x) = x^4(x^2+x+1)^2$.

Example 2.8.16. Observe that there cannot be a 3×3 matrix with elementary divisors x^2 and x^2 because this would force the characteristic polynomial to be x^4, and this is impossible.

Example 2.8.17. Likewise, there cannot be any 3×3 matrices with elementary divisors x and x because this would force the characteristic polynomial to be x^2, and this is impossible.

2.9 The Rational Canonical Form

Last section, we defined the Smith Normal Form of the characteristic matrix of a real $n \times n$ matrix. Essentially, the Smith Normal Form provides a generalization of the reduced row echelon form of a matrix with entries that do not lie in a **field**. Explicitly, polynomials do not admit multiplicative inverses, so a matrix whose entries consist of polynomials might not admit a typical reduced row echelon form consisting of zeros and ones; however, the Smith Normal Form guarantees that every such matrix can be placed in a unique diagonal form consisting of ones and polynomials along the diagonal in such a manner that each of the non-constant polynomials divides the next. Even more, the Smith Normal Form provides the invariant factors and elementary divisors of a real $n \times n$ matrix. We will see throughout this section and the next that this information leads to canonical forms that are (in a strict sense) "simplest" and from which the properties of a matrix can be easily deduced.

Given any monic polynomial $p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ of degree n, we define the **companion matrix** of the polynomial $p(x)$ as the following $n \times n$ matrix.

$$C_{p(x)} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & -\alpha_1 \\ 0 & 1 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\alpha_{n-1} \end{bmatrix}$$

Example 2.9.1. Observe that the companion matrix of any linear polynomial $x + c$ is $\begin{bmatrix} -c \end{bmatrix}$. Explicitly, the companion matrix of x is $\begin{bmatrix} 0 \end{bmatrix}$, and the companion matrix of $x - 1$ is $\begin{bmatrix} 1 \end{bmatrix}$.

Example 2.9.2. Observe that the companion matrix of any quadratic polynomial $x^2 + ax + b$ is

$$\begin{bmatrix} 0 & -b \\ 1 & -a \end{bmatrix}.$$

Explicitly, the companion matrix of $x^2 + 1$ is given as follows.

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

Likewise, the companion matrix of $x^2 + x + 1$ is the following.

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

Crucially, the characteristic polynomial and minimal polynomial of the companion matrix of a monic polynomial $p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ are both simply $p(x)$.

Proposition 2.9.3. *Consider the monic polynomial* $p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ *of positive degree* n *and the companion matrix* $C_{p(x)}$ *of the polynomial* $p(x)$.

$$C_{p(x)} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & -\alpha_1 \\ 0 & 1 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\alpha_{n-1} \end{bmatrix}$$

Both the characteristic polynomial and the minimal polynomial of $C_{p(x)}$ *are equal to* $p(x)$.

Proof. We will prove that the characteristic polynomial of $C_{p(x)}$ is equal to $p(x)$. We proceed by induction on the degree n of $p(x)$. Certainly, if $n = 1$, then the companion matrix of $p(x) = x + \alpha_0$ is annihilated by $p(x)$ because it holds that $C_{p(x)} = \begin{bmatrix} -\alpha_0 \end{bmatrix}$ so that $p(C_{p(x)}) = C_{p(x)} + \alpha_0 I = O$. We conclude in this case that $p(x)$ is the minimal polynomial of $C_{p(x)}$ by Proposition 2.3.9, hence it is the characteristic polynomial by Proposition 2.3.14. We will assume by induction that the claim holds for all monic polynomials of degree $n-1$. Consider the characteristic matrix $xI - C_{p(x)}$.

$$xI - C_{p(x)} = \begin{bmatrix} x & 0 & \cdots & 0 & \alpha_0 \\ -1 & x & \cdots & 0 & \alpha_1 \\ 0 & -1 & \ddots & 0 & \alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & x + \alpha_{n-1} \end{bmatrix}$$

By definition, the characteristic polynomial of $C_{p(x)}$ is $\det(xI - C_{p(x)})$. Expanding the determinant along the first row yields $\det(xI - C_{p(x)}) = x \det(xI - C_{q(x)}) + (-1)^{n+1} \alpha_0 \det(A)$ for the matrices

$$xI - C_{q(x)} = \begin{bmatrix} x & 0 & \cdots & 0 & \alpha_1 \\ -1 & x & \cdots & 0 & \alpha_2 \\ 0 & -1 & \ddots & 0 & \alpha_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & x + \alpha_{n-1} \end{bmatrix} \text{ and } A = \begin{bmatrix} -1 & x & 0 & \cdots & 0 \\ 0 & -1 & x & \cdots & 0 \\ 0 & 0 & -1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

obtained as $(n-1) \times (n-1)$ submatrices of $xI - C_{p(x)}$ by deleting the first row and first column and the first row and nth column of $xI - C_{p(x)}$, respectively. Observe that $C_{q(x)}$ is the companion matrix of the monic polynomial $q(x) = x^{n-1} + \alpha_{n-1} x^{n-2} + \cdots + \alpha_3 x^2 + \alpha_2 x + \alpha_1$, hence by induction, the characteristic polynomial and the minimal polynomial of $C_{q(x)}$ are both $q(x)$. Particularly, it follows that $x \det(xI - C_{q(x)}) = xq(x) = x^n + \alpha_{n-1} x^{n-1} + \cdots + \alpha_3 x^3 + \alpha_2 x^2 + \alpha_1 x = p(x) - \alpha_0$. On the other hand, we note that A is an upper-triangular matrix with $n-1$ copies of -1 along the diagonal, hence we conclude by Proposition 2.7.1 that $\det(A) = (-1)^{n-1}$ and $(-1)^{n+1} \alpha_0 \det(A) = \alpha_0$. Combined, these two calculations reveal that $\det(xI - C_{p(x)}) = p(x) - \alpha_0 + \alpha_0 = p(x)$, as desired.

Even though it is a bit contrived, we will prove that $p(x)$ is the minimal polynomial of $C_{p(x)}$ by demonstrating that no monic polynomial of strictly lesser degree annihilates $C_{p(x)}$. Observe that for the $n \times 1$ standard basis vector E_1 consisting of one in the first row and zeros elsewhere, we have that $C_{p(x)} E_1 = E_2$ so that $C_{p(x)}^2 E_1 = C_{p(x)} E_2 = E_3$ and $C_{p(x)}^k E_1 = E_{k+1}$ for all integers $1 \leq k \leq n-1$. Consequently, for any monic polynomial $q(x) = x^{n-1} + \beta_{n-2} x^{n-2} + \cdots + \beta_2 x^2 + \beta_1 x + \beta_0$, we have that $q(C_{p(x)}) E_1 = E_n + \beta_{n-2} E_{n-1} + \cdots + \beta_2 E_3 + \beta_1 E_2 + \beta_0 E_1$. We conclude that $q(C_{p(x)})$ is nonzero, hence there cannot be a monic polynomial of degree less than n that annihilates $C_{p(x)}$. $\square$

Given any (real) matrices $A_1, \dots, A_k$ such that A_i is an $n_i \times n_i$ matrix for each integer $1 \leq i \leq k$, the **direct sum** of $A_1, \dots, A_k$ is the (real) $(n_1 + \cdots + n_k) \times (n_1 + \cdots + n_k)$ matrix

$$A_1 \oplus \cdots \oplus A_k = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix}$$

constructed by arranging the matrices $A_1, \dots, A_k$ along the main diagonal and completing the matrix with zeros elsewhere. We refer to a square matrix of this form as a **block diagonal** matrix.

Example 2.9.4. Every diagonal matrix can be realized as a block diagonal matrix whose components along the main diagonal are simply 1×1 matrices. Explicitly, we have the following.

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} \end{bmatrix} \oplus \begin{bmatrix} a_{22} \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} a_{nn} \end{bmatrix}$$

Example 2.9.5. By definition, the direct sum of a 1×1 and a 2×2 matrix matrix is a 3×3 block diagonal matrix. Explicitly, the direct sum is a matrix of the following form.

$$\begin{bmatrix} a_{11} \end{bmatrix} \oplus \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & b_{11} & b_{12} \\ 0 & b_{21} & b_{22} \end{bmatrix}$$

Block diagonal matrices behave in a civilized manner with respect to taking determinants and computing their characteristic matrices. Consequently, the determinant, characteristic polynomial, and minimal polynomial of a block diagonal matrix can be easily deduced as follows.

Proposition 2.9.6. *Given any square matrices $A_1, \dots, A_k$, we have that*

$$\det(A_1 \oplus \cdots \oplus A_k) = \det(A_1) \cdots \det(A_k).$$

Proof. By definition of the direct sum of matrices, we have the following.

$$A_1 \oplus \cdots \oplus A_k = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix}$$

By Corollary 2.1.13, there exist scalars $\alpha_1, \dots, \alpha_k$ such that $\det(A_i) = \alpha_i \det(\operatorname{RREF}(A_i))$ for each integer $1 \leq i \leq k$. Considering that the matrix $A_1 \oplus \cdots \oplus A_k$ is block diagonal, performing elementary row operations on any submatrix A_i does not affect any of the other submatrices, hence we may reduce each of the matrices $A_1, \dots, A_k$ to its reduced row echelon form at the cost of some scalar.

$$\det(A_1 \oplus \cdots \oplus A_k) = \alpha_1 \cdots \alpha_k \begin{vmatrix} \operatorname{RREF}(A_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \operatorname{RREF}(A_k) \end{vmatrix}$$

Either the reduced row echelon form of each of the matrices $A_1, \dots, A_k$ is the appropriately-sized identity matrix, or the reduced row echelon form of some matrix possesses a zero row. Certainly, in the first case, the determinant of the matrix in the above displayed equation is one, and we conclude that $\det(A_1 \oplus \cdots \oplus A_k) = \alpha_1 \cdots \alpha_k$. Even more, the determinant of each matrix A_i satisfies that $\det(A_i) = \alpha_i$, hence it holds that $\det(A_1 \oplus \cdots \oplus A_k) = \det(A_1) \cdots \det(A_k)$. Conversely, if the reduced row echelon form of some matrix possesses a zero row, then the determinant of the matrix in the above displayed equation is zero so that $\det(A_1 \oplus \cdots \oplus A_k) = 0 = \det(A_1) \cdots \det(A_k)$. $\square$

Corollary 2.9.7. *Given any square matrices $A_1, \dots, A_k$ with respective characteristic polynomials $\chi_1(x), \dots, \chi_k(x)$, the characteristic polynomial of $A_1 \oplus \cdots \oplus A_k$ is $\chi_1(x) \cdots \chi_k(x)$.*

Proof. Considering that $xI - (A_1 \oplus \cdots \oplus A_k) = (xI - A_1) \oplus \cdots \oplus (xI - A_k)$, the claim follows immediately from the definition of the characteristic polynomial and Proposition 2.9.6. $\square$

Proposition 2.9.8. *Given any square matrices $A_1, \dots, A_k$ with respective minimal polynomials $\mu_1(x), \dots, \mu_k(x)$, the minimal polynomial of $A_1 \oplus \cdots \oplus A_k$ is* $\operatorname{lcm}(\mu_1(x), \dots, \mu_k(x))$.

Proof. We claim that for any polynomial $p(x)$, we have that $p(A_1 \oplus \cdots \oplus A_k) = p(A_1) \oplus \cdots \oplus p(A_k)$. Considering that the identity $\alpha(A_1 \oplus \cdots \oplus A_k) = (\alpha A_1) \oplus \cdots \oplus (\alpha A_k)$ clearly holds, it suffices to prove that $(A_1 \oplus \cdots \oplus A_k)^n = (A_1^n) \oplus \cdots \oplus (A_k^n)$ for any positive integer n: indeed, we have that

$$(A_1 \oplus \cdots \oplus A_k)^2 = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix} \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix} = \begin{bmatrix} A_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k^2 \end{bmatrix} = (A_1^2) \oplus \cdots \oplus (A_k^2)$$

because the only nonzero entries of this matrix product come from the rows and columns corresponding to the matrix A_i for each integer $1 \leq i \leq k$. Certainly, it is possible to repeat this process for any positive integer n, hence the desired identity $p(A_1 \oplus \cdots \oplus A_k) = p(A_1) \oplus \cdots \oplus p(A_k)$ holds.

Consider the least common multiple $p(x) = \operatorname{lcm}(\mu_1(x), \dots, \mu_k(x))$ of the minimal polynomials of $A_1, \dots, A_k$. By definition, for each integer $1 \leq i \leq k$, there exists a polynomial $q_i(x)$ such that $p(x) = \mu_i(x) q_i(x)$, from which it follows that $p(x)$ annihilates the matrices $A_1, \dots, A_k$. Consequently, we find that $p(x)$ annihilates $A_1 \oplus \cdots \oplus A_k$, hence by Proposition 2.3.9, we conclude that $p(x)$ must be divisible by the minimal polynomial $\mu(x)$ of $A_1 \oplus \cdots \oplus A_k$. Conversely, if $\mu(x)$ annihilates the direct sum $A_1 \oplus \cdots \oplus A_k$, then it must annihilate each of the matrices $A_1, \dots, A_k$ because it holds by the previous paragraph that $\mu(A_1 \oplus \cdots \oplus A_k) = \mu(A_1) \oplus \cdots \oplus \mu(A_k)$, and the latter is equal to the zero matrix if and only if $\mu(A_i)$ is equal to the zero matrix for each integer $1 \leq i \leq k$. By Proposition 2.3.9, $\mu(x)$ is divisible by $\mu_1(x), \dots, \mu_k(x)$, hence it is divisible by $p(x) = \operatorname{lcm}(\mu_1(x), \dots, \mu_k(x))$. $\square$

We are at last ready to construct the Rational Canonical Form of a real $n \times n$ matrix.

Definition 2.9.9 (Rational Canonical Form)**.** Consider any (real) $n \times n$ matrix A with invariant factors $p_1(x), p_2(x), \dots, p_\ell(x)$ whose companion matrices are $C_{p_1(x)}, C_{p_2(x)}, \dots, C_{p_\ell(x)}$, respectively. We define the **Rational Canonical Form** of A as the (real) $n \times n$ matrix

$$\operatorname{RCF}(A) = C_{p_1(x)} \oplus C_{p_2(x)} \oplus \cdots \oplus C_{p_\ell(x)} = \begin{bmatrix} C_{p_1(x)} & 0 & 0 & 0 \\ 0 & C_{p_2(x)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & C_{p_\ell(x)} \end{bmatrix}$$

Example 2.9.10. Let us compute the Rational Canonical Form for the matrix of Example 2.8.2.

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

We proved in that example that the only invariant factor of A is $(x-1)(x+1) = x^2 - 1$. Consequently, the Rational Canonical Form for A is the companion matrix of this quadratic polynomial.

$$\operatorname{RCF}(A) = C_{x^2-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Example 2.9.11. Let us compute the Rational Canonical Form for the matrix of Example 2.8.3.

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

We proved in that example that the only invariant factor of A is x^2. Like the previous example, the Rational Canonical Form for A must be the companion matrix of x^2.

$$\operatorname{RCF}(A) = C_{x^2} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Example 2.9.12. Let us compute the Rational Canonical Form for the matrix of Example 2.8.4.

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

We proved in that example that the invariant factors of A are x and $x(x-6) = x^2 - 6x$. Consequently, the Rational Canonical Form for A is the direct sum of the companion matrices of x and $x^2 - 6x$.

$$\operatorname{RCF}(A) = C_x \oplus C_{x^2-6x} = \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 6 \end{bmatrix}$$

Example 2.9.13. Let us compute the Rational Canonical Form for the matrix of Example 2.8.5.

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

Considering that the invariant factors of A are $x - 1$ and $(x-1)^2 = x^2 - 2x + 1$ by the example, the Rational Canonical Form for A is the direct sum of the companion matrices of $x - 1$ and $x^2 - 2x + 1$.

$$\operatorname{RCF}(A) = C_{x-1} \oplus C_{x^2-2x+1} = \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} 0 & -1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

Example 2.9.14. Consider any matrix A whose invariant factors are $x - 1$ and $(x - 1)(x - 2)$. Observe that any such matrix must be a 3×3 matrix. By definition, the Rational Canonical Form for such a matrix is the direct sum of the companion matrices of $x - 1$ and $(x-1)(x-2) = x^2 - 3x + 2$.

$$\operatorname{RCF}(A) = C_{x-1} \oplus C_{x^2-3x+2} = \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} 0 & -2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -2 \\ 0 & 1 & 3 \end{bmatrix}$$

Example 2.9.15. Consider any matrix A whose invariant factors are x, x^2, and $x^3(x+1)^2$. Observe that any such matrix must be an 8×8 matrix. By definition, the Rational Canonical Form for such a matrix is the direct sum of the companion matrices of x, x^2, and $x^3(x+1)^2 = x^5 + 2x^4 + x^3$.

$$\operatorname{RCF}(A) = C_x \oplus C_{x^2} \oplus C_{x^5+2x^4+x^3} = \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix}$$

Example 2.9.16. Consider any matrix A whose invariant factors are x, x, $x^2(x^2+1) = x^4 + x^2$, and $x^3(x-1)(x^2+1) = x^3(x^3-x^2+x-1) = x^6 - x^5 + x^4 - x^3$. Observe that any such matrix must be a 12×12 matrix. By definition, the Rational Canonical Form for such a matrix is the direct sum of the companion matrices of x, x, $x^4 + x^2$, and $x^6 - x^5 + x^4 - x^3$.

$$\begin{aligned}\operatorname{RCF}(A) &= C_x \oplus C_x \oplus C_{x^4+x^2} \oplus C_{x^6-x^5+x^4-x^3} \\ &= [0] \oplus [0] \oplus \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \oplus [0] \oplus \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}\end{aligned}$$

Example 2.9.17. Consider any matrix A with two invariant factors of $x^2(x^2+x+1) = x^4+x^3+x^2$. Observe that any such matrix must be an 8×8 matrix, and the Rational Canonical Form for such a matrix must be the direct sum of the companion matrix of $x^4 + x^3 + x^2$ with itself.

$$\operatorname{RCF}(A) = C_{x^4+x^3+x^2} \oplus C_{x^4+x^3+x^2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

2.10 The Jordan Canonical Form

Like the Rational Canonical Form, the **Jordan Canonical Form** of an $n \times n$ matrix is a block diagonal matrix built as a direct sum of square matrices that are obtained from the Smith Normal Form of the characteristic matrix. Explicitly, suppose that A is a (real) $n \times n$ matrix with elementary divisors $(x - c_{i1})^{e_{i1}}, \dots, (x - c_{ik})^{e_{ik}}$. We refer to the following $e_{ij} \times e_{ij}$ upper-triangular matrix $J_{(x-c_{ij})^{e_{ij}}}$ as the **Jordan matrix** (or **Jordan block**) corresponding to the elementary divisor $(x - c_{ij})^{e_{ij}}$.

$$J_{(x-c_{ij})^{e_{ij}}} = \begin{bmatrix} c_{ij} & 1 & 0 & \cdots & 0 \\ 0 & c_{ij} & 1 & \cdots & 0 \\ 0 & 0 & c_{ij} & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & c_{ij} \end{bmatrix}$$

Put another way, the Jordan matrix corresponding to the elementary divisor $(x - c_{ij})^{e_{ij}}$ is the $e_{ij} \times e_{ij}$ upper-triangular matrix consisting of c_{ij} on the diagonal and ones along the **superdiagonal**.

Example 2.10.1. By definition, the Jordan matrix corresponding to any linear polynomial $x + c$ is the 1×1 matrix $J_{x+c} = \begin{bmatrix} -c \end{bmatrix}$. One might recognize this as the companion matrix of $x + c$.

Example 2.10.2. By definition, the Jordan matrix corresponding to the polynomial $(x-1)^2$ is the 2×2 upper-triangular matrix with ones along the diagonal and ones along the superdiagonal.

$$J_{(x-1)^2} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Example 2.10.3. By definition, the Jordan matrix corresponding to the polynomial $(x+3)^3$ is the 3×3 upper-triangular matrix with -3s along the diagonal and ones along the superdiagonal.

$$J_{(x+3)^3} = \begin{bmatrix} -3 & 1 & 0 \\ 0 & -3 & 1 \\ 0 & 0 & -3 \end{bmatrix}$$

Definition 2.10.4 (Jordan Canonical Form)**.** Consider any (real) $n \times n$ matrix A with elementary divisors $(x - c_{i1})^{e_{i1}}, (x - c_{i2})^{e_{i2}}, \dots, (x - c_{ik})^{e_{ik}}$ and their corresponding Jordan matrices $J_{(x-c_{i1})^{e_{i1}}}$, $J_{(x-c_{i2})^{e_{i2}}}, \dots, J_{(x-c_{ik})^{e_{ik}}}$. We define the **Jordan Canonical Form** of A as the $n \times n$ matrix

$$\mathrm{JCF}(A) = J_{(x-c_{i1})^{e_{i1}}} \oplus J_{(x-c_{i2})^{e_{i2}}} \oplus \cdots \oplus J_{(x-c_{ik})^{e_{ik}}} = \begin{bmatrix} J_{(x-c_{i1})^{e_{i1}}} & 0 & 0 & 0 \\ 0 & J_{(x-c_{i2})^{e_{i2}}} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & J_{(x-c_{ik})^{e_{ik}}} \end{bmatrix}$$

Example 2.10.5. Let us compute the Jordan Canonical Form for the matrix of Example 2.8.2.

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

We proved in that example that the elementary divisors of A are $x - 1$ and $x + 1$. Consequently, the Jordan Canonical Form for A is the direct sum of the 1×1 Jordan matrices J_{x-1} and J_{x+1}.

$$\mathrm{JCF}(A) = J_{x-1} \oplus J_{x+1} = \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Example 2.10.6. Let us compute the Jordan Canonical Form for the matrix of Example 2.8.3.

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

We proved in that example that the only elementary divisor of A is x^2. Like the previous example, the Jordan Canonical Form for A must be the 2×2 Jordan matrix J_{x^2}.

$$\mathrm{JCF}(A) = J_{x^2} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Example 2.10.7. Let us compute the Jordan Canonical Form for the matrix of Example 2.8.4.

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

We proved in that example that the elementary divisors of A are x, x, and $x - 6$. Consequently, the Jordan Canonical Form for A is the direct sum of the 1×1 Jordan matrices J_x, J_x, and J_{x-6}.

$$\mathrm{JCF}(A) = J_x \oplus J_x \oplus J_{x-6} = \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

Example 2.10.8. Let us compute the Jordan Canonical Form for the matrix of Example 2.8.5.

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

By the example, the elementary divisors of A are $x-1$ and $(x-1)^2$, hence the Jordan Canonical Form for A is the direct sum of the 1×1 Jordan matrix J_{x-1} and the 2×2 Jordan matrix $J_{(x-1)^2}$.

$$\text{JCF}(A) = J_{x-1} \oplus J_{(x-1)^2} = \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Example 2.10.9. Consider any matrix A whose elementary divisors are $x-1$, $x-1$, and $x-2$. Observe that any such matrix must be a 3×3 matrix. By definition, the Jordan Canonical Form for such a matrix is the direct sum of the Jordan matrices J_{x-1}, J_{x-1}, and J_{x-2}.

$$\text{JCF}(A) = J_{x-1} \oplus J_{x-1} \oplus J_{x-2} = \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Example 2.10.10. Consider any matrix A whose elementary divisors are x, x^2, x^3, and $(x+1)^2$. Observe that any such matrix must be an 8×8 matrix. By definition, the Jordan Canonical Form for such a matrix is the direct sum of the Jordan matrices corresponding to x, x^2, x^3, and $(x+1)^2$.

$$\text{JCF}(A) = J_x \oplus J_{x^2} \oplus J_{x^3} \oplus J_{(x+1)^2} = \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

Example 2.10.11. Consider any real matrix A whose invariant factors are x, x, $x^2(x^2+1)$, and $x^3(x-1)(x^2+1)$. Observe that both roots of the polynomial x^2+1 are complex numbers: indeed, the roots of x^2+1 are i and $-i$. Consequently, if we view A a real matrix, then A does not admit a Jordan Canonical Form. Explicitly, the Jordan Canonical Form is built from the Jordan matrices corresponding to powers of linear polynomials: if x^2+1 is an elementary divisor of A, then viewed as a real polynomial, this polynomial does not **split** as a product of linear polynomials; however, if we view A as a matrix whose entries are complex numbers, then we may view x^2+1 as a polynomial with complex coefficients, hence it is permissible to factor x^2+1 as $(x+i)(x-i)$. Under this lens, the elementary divisors of A are $x, x, x^2, x^3, x-1, x-i, x+i, x-i$, and $x+i$. Consequently, the Jordan Canonical Form for A is the following 12×12 complex upper-triangular matrix.

$$\begin{aligned} \text{JCF}(A) &= J_x \oplus J_x \oplus J_{x^2} \oplus J_{x^3} \oplus J_{x-1} \oplus J_{x-i} \oplus J_{x+i} \oplus J_{x-i} \oplus J_{x+i} \\ &= \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} i \end{bmatrix} \oplus \begin{bmatrix} -i \end{bmatrix} \oplus \begin{bmatrix} i \end{bmatrix} \oplus \begin{bmatrix} -i \end{bmatrix} \end{aligned}$$

Example 2.10.12. Consider any matrix A with elementary divisors of x^2, x^2, x^2+x+1, and x^2+x+1. Observe that the Jordan Canonical Form for such a matrix exists if and only if we view A as a matrix with complex entries: indeed, the polynomial $x^2 + x + 1$ has two complex roots $-\frac{1}{2} + \frac{\sqrt{3}}{2}i$ and $-\frac{1}{2} - \frac{\sqrt{3}}{2}i$. Consequently, the Jordan Canonical Form for A is the following.

$$\operatorname{JCF}(A) = J_{x^2} \oplus J_{x^2} \oplus J_{x+\frac{1}{2}-\frac{\sqrt{3}}{2}i} \oplus J_{x+\frac{1}{2}+\frac{\sqrt{3}}{2}i} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} -\frac{1}{2} + \frac{\sqrt{3}}{2}i \end{bmatrix} \oplus \begin{bmatrix} -\frac{1}{2} - \frac{\sqrt{3}}{2}i \end{bmatrix}$$

Remark 2.10.13. Examples 2.10.11 and 2.10.12 raise an important point regarding the Jordan Canonical Form of a square matrix A: it exists if and only if the elementary divisors of A are all power of linear polynomials. Consequently, if we want the Jordan Canonical Form to exist for any square matrix, we must assume that the entries of our matrix lie in an **algebraically closed** field, i.e., we must ensure that the characteristic polynomial of our matrix can be written as a product of (not necessarily distinct) linear polynomials. Often, the caveat with the Jordan Canonical Form is that it is an upper-triangular matrix with entries in the complex numbers. Conversely, the Rational Canonical Form of a matrix always exists; however, it is rarely an upper-triangular matrix. Even still, in most cases, the Jordan Canonical Form is preferable to the Rational Canonical Form because of its upper-triangular form. One can prove that the determinant of a matrix is the product of its eigenvalues, hence the product of the eigenvalues of a real matrix must be a real number. We could have predicted this based on the fact that complex roots come in conjugate pairs whose product is a real number. Even more, the **trace** of a matrix is the sum of the diagonal components of the matrix; this can be achieved as the sum of the eigenvalues. Once again, if the matrix is real, then the sum of its eigenvalues is a real number because each conjugate pair of complex eigenvalues sum to a real number. Consequently, the requirement to pass to the complex numbers is not detrimental.

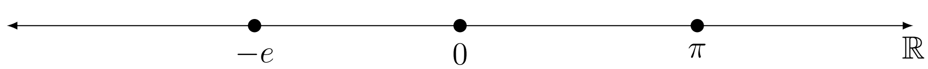
Chapter 3

Inner Product Spaces

Previously, we dedicated the second chapter of these lecture notes to the algebraic properties of matrices. Explicitly, we studied determinants, characteristic polynomials, minimal polynomials, eigenvalues, eigenvectors, eigenspaces, and canonical forms for matrices such as the Smith Normal Form, the Rational Canonical Form, and the Jordan Canonical Form. We noticed in our study of eigenvalues, eigenvectors, and eigenspaces that these algebraic objects possess some innate geometric properties. Explicitly, the Spectral Theorem tells us that every real symmetric matrix induces a basis of eigenvectors for the space of real column vectors for which every pair of eigenvectors corresponding to distinct eigenvalues is orthogonal. We are interested throughout this chapter in further unravelling the geometry of vector spaces that admit a notion of orthogonality.

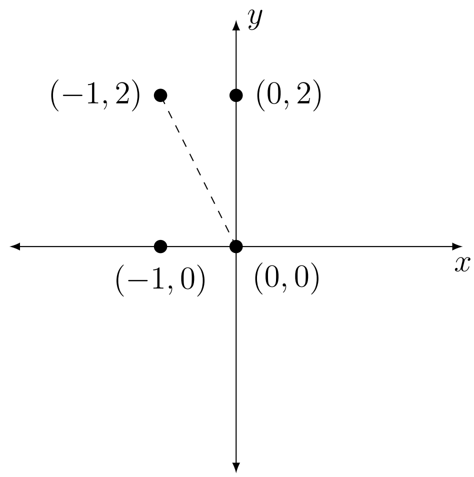
3.1 Real n-Space

Consider the set $\mathbb{R}$ consisting of real numbers. Like usual, we may geometrically realize $\mathbb{R}$ as a line (the real number line) consisting of **points** x that lie a distance of $|x|$ from the origin 0 for each real number x. Explicitly, the point π lies π units to the right of the origin, and the point $-e$ lies e units to the left of the origin. Given any pair of real numbers $a \leq b$, the distance between the points a and b in $\mathbb{R}$ is given by the length of the closed interval $[a, b]$; we learn back in calculus that this distance is exactly the real number $b - a$. Consequently, the real numbers $\mathbb{R}$ give rise to the geometric notions of a line and the notion of distance between two points on a line.

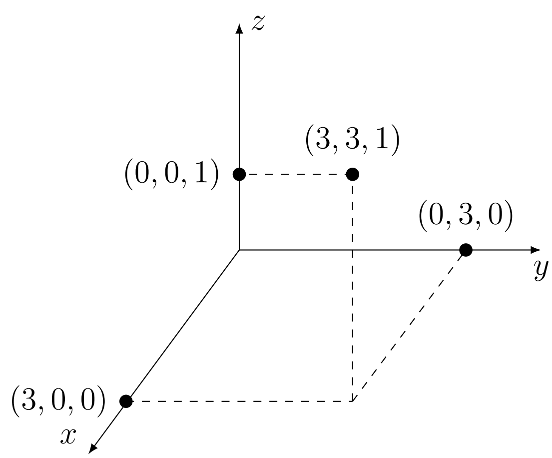


One can only move forward and backward on the real number line, hence the geometry of $\mathbb{R}$ is (in this sense) quite simple. On the other hand, suppose that we want to keep track of both east-west movement and north-south movement. Given that an object lies x units from the origin in the east-west direction and y units in the north-south direction, we may canonically express this data as the **ordered pair** (x, y). Explicitly, if a particle lies 1 unit west and 2 units north of the origin $(0, 0)$, then it lies 1 unit to the left of the origin on the x-axis and 2 units north of the origin on the y-axis; the location of the particle in this case can be written as the ordered pair $(-1, 2)$. We refer to the collection of all ordered pairs of real numbers (x, y) as the **Cartesian product** $\mathbb{R} \times \mathbb{R}$ of the real numbers with itself, i.e., we have that $\mathbb{R} \times \mathbb{R} = \{(x, y) \mid x \text{ and } y \text{ are real numbers}\}$.

Graphically, the points in $\mathbb{R} \times \mathbb{R}$ form a plane, so $\mathbb{R} \times \mathbb{R}$ is often called the **Cartesian plane**. Conventionally, the Cartesian plane is denoted by $\mathbb{R}^2$ and referred to also as **real** 2-**space**.



Going one step further, let us keep track of east-west, north-south, and up-down movements. Explicitly, if x measures the location of a particle in the x-axis; y measures the location of a particle in the y-axis; and z measures the location of particle in the z-axis, then the ordered triple (x, y, z) conveniently encapsulates this information. Like before, if the particle lies 3 units east of the origin; 3 units north of the origin; and 1 unit above the origin, then the particle's location is given by the ordered triple $(3, 3, 1)$. We denote by $\mathbb{R}^3$ the collection of all ordered triples of real numbers, i.e., we have that $\mathbb{R}^3 = \{(x, y, z) \mid x, y, \text{ and } z \text{ are real numbers}\}$; we refer to $\mathbb{R}^3$ as real 3-space.



Once and for all, if n is a positive integer, then we will denote by $\mathbb{R}^n$ the collection of all n-**tuples** of real numbers, i.e., we have that $\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \mid x_1, x_2, \dots, x_n \text{ are real numbers}\}$. Like before with real column vectors, we will use a capital letter X to denote a real n-tuple $(x_1, x_2, \dots, x_n)$; we refer to the real number x_1 as the first **coordinate** of X; we refer to the real number x_2 as the second coordinate of X; we refer to the real number x_n as the nth coordinate of X; and in general, we refer to x_i as the ith coordinate of X for each integer $1 \leq i \leq n$. Every point in real n-space is uniquely determined by its coordinates. Explicitly, if $(x_1, x_2, \dots, x_n) = (y_1, y_2, \dots, y_n)$, then each of the coordinates on the left-hand side must be equal to the corresponding coordinate on the right-hand side, i.e., $y_i = x_i$ for all integers $1 \leq i \leq n$. Even though it is not possible to envision points in real n-space for $n \geq 5$, it is still meaningful to discuss this notion. Explicitly, every set of data consisting of n distinct real parameters induces an element of real n-space $\mathbb{R}^n$.

Our next proposition illustrates that $\mathbb{R}^n$ forms a real n-dimensional vector space.

Proposition 3.1.1. *Real n-space $\mathbb{R}^n$ forms a real vector space of dimension n.*

Proof. We define addition of points in real n-space **componentwise** by

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n).$$

Considering that addition of real numbers constitutes an associative and commutative binary operation on the real numbers, conditions (1.), (2.), and (3.) of Definition 1.6.5 are satisfied. Even more, the zero vector in $\mathbb{R}^n$ is the n-tuple $O = (0, 0, \dots, 0)$, and for any real n-tuple $(x_1, x_2, \dots, x_n)$, we have that $-(x_1, x_2, \dots, x_n) = (-x_1, -x_2, \dots, -x_n)$. We conclude that conditions (4.) and (5.) of the definition hold, hence we may turn our attention to scalar multiplication in $\mathbb{R}^n$. We define $\alpha(x_1, x_2, \dots, x_n) = (\alpha x_1, \alpha x_2, \dots, \alpha x_n)$ for any real number α and any real n-tuple $(x_1, x_2, \dots, x_n)$. Considering that multiplication of real numbers constitutes an associative, commutative, and distributive binary operation on the real numbers, it follows that $\mathbb{R}^n$ is a real vector space.

Last, the dimension of $\mathbb{R}^n$ is n: the standard basis of $\mathbb{R}^n$ consists of the vectors E_i whose ith coordinate is 1 and whose other coordinates are 0, i.e., $E_1 = (1, 0, \dots, 0)$, $E_2 = (0, 1, \dots, 0)$, etc. $\square$

Example 3.1.2. Consider the points $X = (1, 1, -1)$, $Y = (1, 2, 3)$, and $Z = (0, -2, -2)$ in $\mathbb{R}^3$. Observe that $X + Y = (2, 3, 2)$, $-Z = (0, 2, 2)$, $Y - Z = (1, 4, 5)$, and $3X = (3, 3, -3)$.

Consequently, we will henceforth refer to points in real n-space as both points and vectors. By Theorem 1.13.12, if we wish to understand any real vector space of dimension n, it suffices to understand the real vector space $\mathbb{R}^n$; the advantage of dealing directly with real n-space itself is that we have access to **Euclidean geometry**. Our aim throughout this chapter is to develop this tool. We begin by defining a notion of **distance** in real n-space. Given any points $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$, we define the distance between X and Y as the following real number.

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

Consequently, the distance from the origin to the point X is given and denoted as follows.

$$\|X\| = d(X, O) = \sqrt{x_1^2 + \cdots + x_n^2}$$

Often, we will simply refer to the quantity $\|X\|$ as the **magnitude** of the vector X. We note that this definition of distance is merely a generalization of the length of the hypotenuse of the right triangle formed by the x-axis, the y-axis, and a point in the Cartesian plane: indeed, if we could visualize the right triangle formed by the origin of $\mathbb{R}^n$, the point $(x_1, x_2, \dots, x_{n-1}, 0)$, and the point $X = (x_1, x_2, \dots, x_{n-1}, x_n)$ in $\mathbb{R}^n$, then the length of its hypotenuse is precisely $\|X\|$.

Example 3.1.3. Consider the vectors X, Y, and Z from Example 3.1.2. Computing the magnitudes of each vector yields $\|X+Y\| = \sqrt{2^2 + 3^2 + 2^2} = \sqrt{17}$ and $\|-Z\| = \sqrt{0^2 + 2^2 + 2^2} = 2\sqrt{2} = \|Z\|$ and $\|3X\| = \sqrt{3^2 + 3^2 + (-3)^2} = 3\sqrt{3} = 3\|X\|$; these last two examples indicate a general phenomenon.

Proposition 3.1.4. *Consider any positive integer n and any vector $X = (x_1, \dots, x_n)$ of $\mathbb{R}^n$.*

1.) *We have that* $\|X\| = 0$ *if and only if* X *is the zero vector.*

2.) *We have that* $\|\alpha X\| = |\alpha|\|X\|$ *for all real numbers* α.

Proof. (1.) By definition, we have that $\|X\| = \sqrt{x_1^2 + \cdots + x_n^2} = 0$ if and only if $x_1^2 + \cdots + x_n^2 = 0$. Clearly, if X is the zero vector, then $x_1 = \cdots = x_n = 0$ so that $x_1^2 + \cdots + x_n^2 = 0^2 + \cdots + 0^2 = 0$. Conversely, if X is a nonzero vector, then its ith coordinate x_i must be nonzero for some integer $1 \leq i \leq n$. Considering that the square of a nonzero real number if a positive real number, we have that $x_i^2 > 0$. Even more, the square of any real number is non-negative, hence we have that $\|X\|^2 = x_1^2 + \cdots + x_n^2 \geq x_i^2 > 0$. We conclude that $\|X\|$ must be nonzero if X is nonzero.

(2.) We define $\alpha X = \alpha(x_1, \dots, x_n) = (\alpha x_1, \dots, \alpha x_n)$. Consequently, the definition of magnitude yields $\|\alpha X\| = \sqrt{(\alpha x_1)^2 + \cdots + (\alpha x_n)^2} = \sqrt{\alpha^2(x_1^2 + \cdots + x_n)^2} = |\alpha|\sqrt{x_1^2 + \cdots + x_n^2} = |\alpha|\|X\|$. $\square$

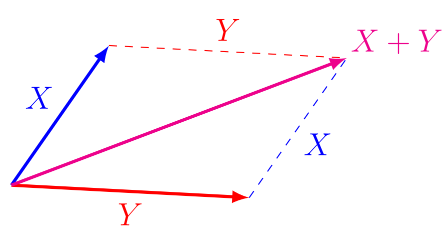
Conventionally, vectors of magnitude one are referred to as **unit vectors**. By Proposition 3.1.4, every nonzero vector X gives rise to a unique unit vector $\frac{1}{\|X\|}X$.

Corollary 3.1.5. *Every nonzero vector* X *of* $\mathbb{R}^n$ *induces a unit vector* $\frac{1}{\|X\|}X$ *of* $\mathbb{R}^n$.

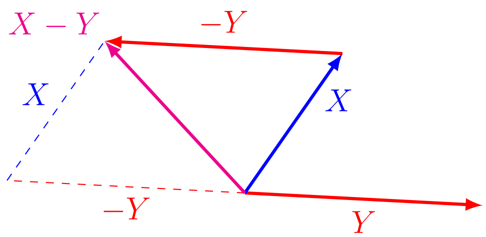
Proof. By Proposition 1.14.11, if X is any nonzero vector of $\mathbb{R}^n$, then $\|X\|$ is a positive real number. Consequently, we have that $\alpha = \frac{1}{\|X\|}$ is a positive real number such that $\|\alpha X\| = \alpha\|X\| = 1$. $\square$

Example 3.1.6. Consider the vectors X, Y, and Z from Example 3.1.2. We demonstrated that $\|X + Y\| = \sqrt{17}$ and $\|Z\| = 2\sqrt{2}$, hence $\frac{1}{\sqrt{17}}(X + Y)$ and $\frac{1}{2\sqrt{2}}Z$ are unit vectors of $\mathbb{R}^3$.

Even more, vectors in real n-space can be considered as rays (or arrows) emanating from the origin and extending to a point in real n-space. Explicitly, the vector $X = (1, 2, 3, 4)$ of $\mathbb{R}^4$ can be represented by the ray extending from the origin $(0, 0, 0, 0)$ to the point $(1, 2, 3, 4)$ in $\mathbb{R}^4$. We refer to the vector X in this case as lying in **standard position**. Often, we will restrict our attention to the Cartesian plane $\mathbb{R}^2$ or real 3-space $\mathbb{R}^3$, where we can visualize this notion. Under this identification, vector addition can be described geometrically as follows: if we determine the vector sum $X + Y$, then we may realize X and Y as rays emanating from the origin; translate Y so that the "foot" of Y lies as the "head" of X; and draw the ray emanating from the "foot" of X to the "head" of Y. Considering that vector addition is commutative, one could also determine $X + Y$ by translating X so that the "foot" of X lies at the "head" of Y and subsequently drawing the raw emanating from the "foot" of Y to the "head" of X. Either way, this situation can be visualized as follows.



We refer to the process of computing the vector sum $X + Y$ pictorially in this way as the **Parallelogram Law**. Observe that for any vector X, the vector $-X$ satisfies that $X + (-X) = O$. Consequently, if we place the "foot" of $-X$ at the "head" of X and draw the ray emanating from the "foot" of X to the "head" of $-X$, we obtain the zero vector. Put another way, the "head" of the translated $-X$ and the "foot" of X coincide, hence $-X$ is nothing more than X in the **opposite direction**. We are therefore able to describe vector subtraction pictorially as follows.

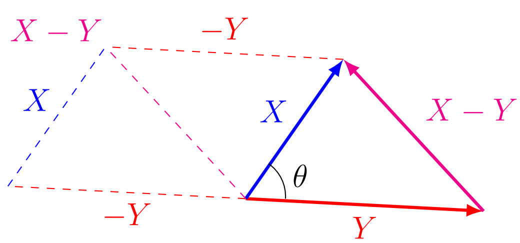


We will henceforth say to two vectors X and Y in $\mathbb{R}^n$ are **parallel** if there exists a nonzero real number α such that $Y = \alpha X$. By extension of the previous definition, we will say then that X and αX have the **same direction** if $\alpha > 0$; they have the **opposite direction** if $\alpha < 0$. Certainly, a pair of vectors in $\mathbb{R}^n$ need not be parallel, hence in general, it might not be possible to say that an arbitrary pair of vectors have the same direction or the opposite direction.

Example 3.1.7. Observe that the vectors $X = (1, 0, -1)$ and $Y = (-3, 0, 3)$ are parallel because we have that $Y = -3X$, hence X and Y have the opposite direction; however, the vector $Z = (-1, 1, 1)$ is not parallel to either X or Y. We will soon see that it is in fact **perpendicular** to both X and Y.

3.2 The Dot Product

Consider any pair of vectors X and Y lying in standard position in real n-space $\mathbb{R}^n$ for some positive integer n. Certainly, if $n = 2$ or $n = 3$, then we could visualize X and Y in the Cartesian plane $\mathbb{R}^2$ or in the real 3-space $\mathbb{R}^3$ that we occupy; more specifically, we could take a protractor and measure the angle θ formed by the intersection of X and Y at the origin. Pictorially, we have the following.



Consequently, by applying the Law of Cosines to this triangle, we obtain the following formula.

$$\|X - Y\|^2 = \|X\|^2 + \|Y\|^2 - 2\|X\|\|Y\|\cos(\theta)$$

Observe that if $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$, then by definition of the magnitude of a vector, it holds that $\|X\|^2 = x_1^2 + \cdots + x_n^2$ and $\|Y\|^2 = y_1^2 + \cdots + y_n^2$ so that

$$\|X - Y\|^2 = (x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 = x_1^2 + \cdots + x_n^2 + y_1^2 + \cdots + y_n^2 - 2(x_1 y_1 + \cdots + x_n y_n).$$

Combining this formula with the Law of Cosines formula from above yields that

$$\|X\|^2 + \|Y\|^2 - 2\|X\|\|Y\|\cos(\theta) = \|X - Y\|^2 = \|X\|^2 + \|Y\|^2 - 2(x_1 y_1 + \cdots + x_n y_n)$$

so that $\|X\|\|Y\|\cos(\theta) = x_1 y_1 + \cdots + x_n y_n$. We refer to the real number $x_1 y_1 + \cdots + x_n y_n$ as the **dot product** of the real vectors X and Y, and we write $X \cdot Y = x_1 y_1 + \cdots + x_n y_n$. We are already familiar with the vector dot product of the first two chapters; the dot product of vectors in $\mathbb{R}^n$ behaves in the same way as the vector dot product, but its output is a real number rather than a real 1×1 matrix. We demonstrate next that the dot product informs the geometry of $\mathbb{R}^n$.

Proposition 3.2.1. *Given any pair of nonzero vectors X and Y lying in standard position in $\mathbb{R}^n$, the angle θ of intersection between the vectors X and Y satisfies that*

$$\theta = \cos^{-1}\left(\frac{X \cdot Y}{\|X\|\|Y\|}\right)$$

Essentially, the formula is obtained from the previous paragraph by solving for θ in the identity $X \cdot Y = \|X\|\|Y\|\cos(\theta)$. Often, we will refer to the formula $X \cdot Y = \|X\|\|Y\|\cos(\theta)$ as the **geometric interpretation** of the dot product. Like before with the vector dot product, we will say that the real vectors X and Y in $\mathbb{R}^n$ are **orthogonal** if and only if it holds that $X \cdot Y = 0$.

Example 3.2.2. Consider the vectors $X = (1, 1, -1)$, $Y = (1, 2, 3)$, and $Z = (0, -2, -2)$ in $\mathbb{R}^3$. By definition of the dot product, we obtain the following identities.

$$\begin{aligned}
X \cdot X &= (1)(1) + (1)(1) + (-1)(-1) = 3 \\
X \cdot Y &= (1)(1) + (1)(2) + (-1)(3) = 0 \\
X \cdot Z &= (1)(0) + (1)(-2) + (-1)(-2) = 0 \\
Y \cdot Z &= (1)(0) + (2)(-2) + (3)(-2) = -10
\end{aligned}$$

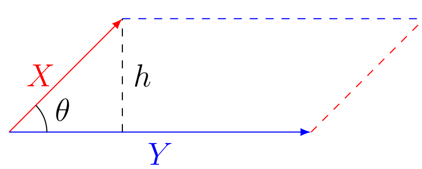
Consequently, we have that X is orthogonal to both Y and Z, but X is not orthogonal to itself, and Y is not orthogonal to Z. Even more, we have that $X \cdot X = \|X\|^2$.

Example 3.2.3. Consider the vectors $X = (1, 2, 0, 2)$ and $Y = (-3, 1, 1, 5)$ in $\mathbb{R}^4$. Even though we cannot visualize these vectors because they live in real 4-space, we can find the angle θ between them. By definition of the magnitude of a vector, we have that $\|X\| = \sqrt{1^2 + 2^2 + 0^2 + 2^2} = \sqrt{9} = 3$ and $\|Y\| = \sqrt{(-3)^2 + 1^2 + 1^2 + 5^2} = \sqrt{36} = 6$. By definition of the dot product, we have that $X \cdot Y = (1)(-3) + (2)(1) + (0)(1)(+(2)(5) = 9$. Consequently, we conclude that

$$\theta = \cos^{-1}\left(\frac{X \cdot Y}{\|X\|\|Y\|}\right) = \cos^{-1}\left(\frac{9}{(3)(6)}\right) = \cos^{-1}\left(\frac{1}{2}\right) = 60^\circ.$$

Proposition 3.2.4. *Given any nonzero, non-parallel vectors X and Y lying in standard position in $\mathbb{R}^n$, the area of the* **parallelogram** *spanned by X and Y is $\|X\|\|Y\|\sin(\theta)$.*

Proof. Pictorially, the parallelogram spanned by X and Y can be determined as follows.



Observe that the angle θ between X and Y satisfies that $h = \|X\|\sin(\theta)$. Because the area of a parallelogram is the base times the height of the parallelogram, it is $h\|Y\| = \|X\|\|Y\|\sin(\theta)$. □

We illustrate next that the dot product satisfies many nice arithmetic properties.

Proposition 3.2.5. *Consider any vectors X, Y, and Z of $\mathbb{R}^n$.*

1.) *We have that $X \cdot Y = Y \cdot X$, i.e., the dot product is commutative.*

2.) *We have that* $X \cdot (Y + Z) = X \cdot Y + X \cdot Z$, *i.e., the dot product is distributive.*

3.) *We have that* $(\alpha X) \cdot Y = \alpha(X \cdot Y) = X \cdot (\alpha Y)$ *for all real numbers* α.

4.) *We have that* $X \cdot X = \|X\|^2$. *Consequently,* $X \cdot X$ *is nonzero if and only if* X *is nonzero.*

Proof. (1.) Let $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ for some real numbers $x_1, \dots, x_n, y_1, \dots, y_n$. Considering that multiplication of real numbers is commutative, for each integer $1 \leq i \leq n$, we have that $x_i y_i = y_i x_i$, from which it follows that $X \cdot Y = x_1 y_1 + \cdots + x_n y_n = y_1 x_1 + \cdots + y_n x_n = Y \cdot X$.

(2.) Given any vector $Z = (z_1, \dots, z_n)$ of $\mathbb{R}^n$, we have that $Y + Z = (y_1 + z_1, \dots, y_n + z_n)$ by definition of addition in $\mathbb{R}^n$. Considering that multiplication of real numbers is distributive, we have that $X \cdot (Y + Z) = x_1(y_1 + z_1) + \cdots + x_n(y_n + z_n) = x_1 y_1 + x_1 z_1 + \cdots + x_n y_n + x_n z_n = X \cdot Y + X \cdot Z$.

(3.) We have that $\alpha X = (\alpha x_1, \dots, \alpha x_n)$ for any real number α by definition of scalar multiplication in $\mathbb{R}^n$. We conclude that $(\alpha X) \cdot Y = (\alpha x_1) y_1 + \cdots + (\alpha x_n) y_n = \alpha(x_1 y_1 + \cdots + x_n y_n)$. Likewise, we have that $\alpha Y = (\alpha y_1, \dots, \alpha y_n)$ so that $X \cdot (\alpha Y) = x_1(\alpha y_1) + \cdots + x_n(\alpha y_n) = \alpha(x_1 y_1 + \cdots + x_n y_n)$. Each of these values is equal to the other, and they are both equal to $\alpha(X \cdot Y)$.

(4.) Last, we have that $X \cdot X = x_1^2 + \cdots + x_n^2 = (\sqrt{x_1^2 + \cdots + x_n^2})^2 = \|X\|^2$. By Proposition 3.1.4, we have that $X \cdot X$ is zero if and only if $\|X\|$ is zero if and only if X is the zero vector. $\square$

By applying the aforementioned properties of the dot product to the situation of orthogonal vectors, we can prove the following important properties of orthogonal vectors.

Proposition 3.2.6. *Consider any vector* X *and any vectors* Y *and* Z *that are orthogonal to* X.

1.) *We have that* X *is orthogonal to* $Y + Z$.

2.) *We have that* X *is orthogonal to* αY *for all real numbers* α.

3.) (**Pythagorean Theorem**) *We have that* $\|X + Y\|^2 = \|X\|^2 + \|Y\|^2$.

Even more, if X *and* Y *lie in standard position in* $\mathbb{R}^n$, *then the angle* θ *of intersection between the vectors* X *and* Y *satisfies that* $\theta = 90°$, *i.e., the vectors* X *and* Y *are* **perpendicular**.

Proof. (1.) By definition, if X and Y are orthogonal and X and Z are orthogonal, then $X \cdot Y = 0$ and $X \cdot Z = 0$. By Proposition 3.2.5, it follows that $X \cdot (Y + Z) = X \cdot Y + X \cdot Z = 0$.

(2.) By Proposition 3.2.5, we have that $X \cdot (\alpha Y) = \alpha(X \cdot Y) = 0$ for all real numbers α.

(3.) By Proposition 3.2.5, the dot product is commutative and distributive so that

$$\|X + Y\| = (X + Y) \cdot (X + Y) = X \cdot X + X \cdot Y + Y \cdot X + Y \cdot Y = \|X\|^2 + 2(X \cdot Y) + \|Y\|^2.$$

Considering that X and Y are orthogonal, we conclude that $2(X \cdot Y) = 0$, as desired.

Last, if X and Y are orthogonal vectors lying in standard position, then by Proposition 3.2.1, the angle θ of intersection between the vectors X and Y is given by $\theta = \cos^{-1}(0) = 90°$. $\square$

Example 3.2.7. We determine in this example unit vector perpendicular to $X = (-1, 3, 4)$. By definition, we require a vector $U = (u, v, w)$ such that $U \cdot X = 0$ and $\|U\| = 1$. Computing the dot product of X and U, we find that $U \cdot X = -u + 3v + 4w = 0$. We have three variables and only one equation, hence there must be two free variables that we are allowed to set equal to anything that is convenient. We will choose $u = 0$ and $v = -4$; the resulting equation is $3(-4) + 4w = 0$ so that $4w = 3(4)$ and $w = 3$. Consequently, the vector $U = (0, -4, 3)$ is orthogonal to X; however, its magnitude is $\sqrt{0^2 + (-4)^2 + 3^2} = 5$, so it is not a unit vector. By Proposition 3.1.5, we find that $\frac{1}{5}U$ is a unit vector; it is orthogonal to X by Proposition 3.2.6 because U is orthogonal to X.

Example 3.2.8. We determine in this example a unit vector perpendicular to $X = (-1, 3, 4)$ and $Y = (2, 1, -1)$. Like before in Example 3.2.8, we must solve the following system of equations.

$$\begin{aligned} -u + 3v + 4w &= (u, v, w) \cdot X = 0 \\ 2u + v - w &= (u, v, w) \cdot Y = 0 \end{aligned}$$

By adding twice the first equation to the second equation, we find that $7v + 7w = 0$ or $v = -w$. We have two equations in three unknowns, so we have one free variable. By declaring that $u = 0$ and $v = 1$, we find that $w = -1$ and $U = (0, 1, -1)$ is orthogonal to X and Y. Considering that $\|U\| = \sqrt{0^2 + 1^2 + (-1)^2} = \sqrt{2}$, we conclude that $\frac{1}{\sqrt{2}}U$ is a unit vector orthogonal to X and Y.

Before we conclude this section, we state and prove two inequalities regarding vectors in $\mathbb{R}^n$.

Theorem 3.2.9 (Cauchy-Schwarz Inequality)**.** *Given any vectors X and Y of $\mathbb{R}^n$, we have that*

$$|X \cdot Y| \leq \|X\|\|Y\|.$$

Proof. Clearly, if either X or Y is the zero vector, then $X \cdot Y = 0$ and $\|X\|\|Y\| = 0$. Consequently, we may assume that neither X nor Y is zero. Even more, we may assume that the vectors X and Y lie in standard position in $\mathbb{R}^n$. By Proposition 3.2.1, we have that

$$\cos(\theta) = \frac{X \cdot Y}{\|X\|\|Y\|}$$

for the angle θ between X and Y. Considering that $|\cos(\theta)| \leq 1$, the inequality follows. □

Theorem 3.2.10 (Triangle Inequality)**.** *Given any vectors X and Y of $\mathbb{R}^n$, we have that*

$$\|X + Y\| \leq \|X\| + \|Y\|.$$

Proof. By Proposition 3.1.4, it follows that $\|X + Y\|$, $\|X\|$, and $\|Y\|$ are each non-negative real numbers, hence the desired inequality holds if and only if the inequality $\|X + Y\|^2 \leq (\|X\| + \|Y\|)^2$ holds. By Proposition 3.2.5, the left-hand side of this inequality is given by $(X + Y) \cdot (X + Y)$. By the proof of Proposition 3.2.6, we note that $(X + Y) \cdot (X + Y) = \|X\|^2 + 2(X \cdot Y) + |Y\|^2$. By the Cauchy-Schwarz Inequality, it follows that $2(X \cdot Y) \leq 2\|X\|\|Y\|$, and the desired inequality holds.

$$\|X + Y\|^2 = \|X\|^2 + 2(X \cdot Y) + |Y\|^2 \leq \|X\|^2 + 2\|X\|\|Y\| + |Y\|^2 = (\|X\| + \|Y\|)^2$$ □

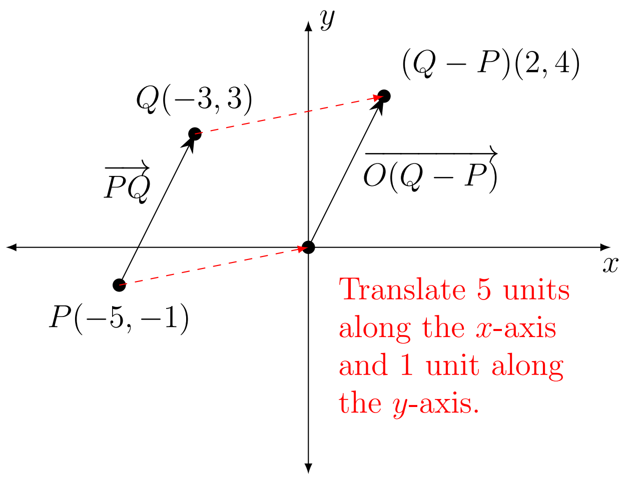
3.3 Lines and Planes

Given any pair of points $P = (x_1, \dots, x_n)$ and $Q = (y_1, \dots, y_n)$ of real n-space $\mathbb{R}^n$, we may construct the **located vector** $\overrightarrow{PQ}$ **beginning** at the point P in the **direction** of the point Q by declaring that $\overrightarrow{PQ} = \langle y_1 - x_1, \dots, y_n - x_n \rangle$. Pictorially, the located vector $\overrightarrow{PQ}$ is simply a ray emanating from the point P and extending to the point Q. Considering that $\mathbb{R}^n$ is a vector space, the located vector $\overrightarrow{PQ}$ can be uniquely identified with the located vector $\overrightarrow{O(Q-P)}$ beginning at the origin O in the direction of the vector $Q - P = (y_1 - x_1, \dots, y_n - x_n)$ that lies in standard position. Explicitly, if we view $\overrightarrow{PQ}$ lying in some hyperplane of $\mathbb{R}^n$, it is simply a **translation** of $\overrightarrow{O(Q-P)}$ or $Q - P$.

Example 3.3.1. Given the points $P = (-5, -1)$ and $Q = (-3, 3)$ in the Cartesian plane $\mathbb{R}^2$, the located vector $\overrightarrow{PQ}$ beginning at P in the direction of Q is $\overrightarrow{PQ} = \langle -3 + 5, 3 + 1 \rangle = \langle 2, 4 \rangle$. Essentially, the located vector $\overrightarrow{PQ}$ contains the same geometric information as the vector $Q - P$ lying in standard position; it is simply translated away from the origin. Explicitly, we have that

$$\|\overrightarrow{PQ}\| = \sqrt{2^2 + 4^2} = \sqrt{20} = 2\sqrt{5} = \|Q - P\|$$

for the vector $Q - P$ of $\mathbb{R}^2$ lying in standard position. Pictorially, the situation is as follows.



Observe that if we translate $\overrightarrow{PQ}$ a distance of 5 units along the x-axis and a distance of 1 unit along the y-axis, then the resulting vector is exactly the located vector $\overrightarrow{O(Q-P)}$ or the vector $Q - P$.

Example 3.3.2. Consider the points $P = (1, 2, 3)$ and $Q = (4, 5, 6)$ in $\mathbb{R}^2$. By definition, the located vector $\overrightarrow{PQ}$ beginning at P in the direction of Q is $\overrightarrow{PQ} = \langle 4 - 1, 5 - 2, 6 - 3 \rangle = \langle 3, 3, 3 \rangle$. We can identify $\overrightarrow{PQ}$ with the vector $Q - P = (3, 3, 3)$ lying in standard position in $\mathbb{R}^3$ by translating $\overrightarrow{PQ}$ a distance of -1 unit along the x-axis; -2 units along the y-axis; and -3 units along the z-axis.

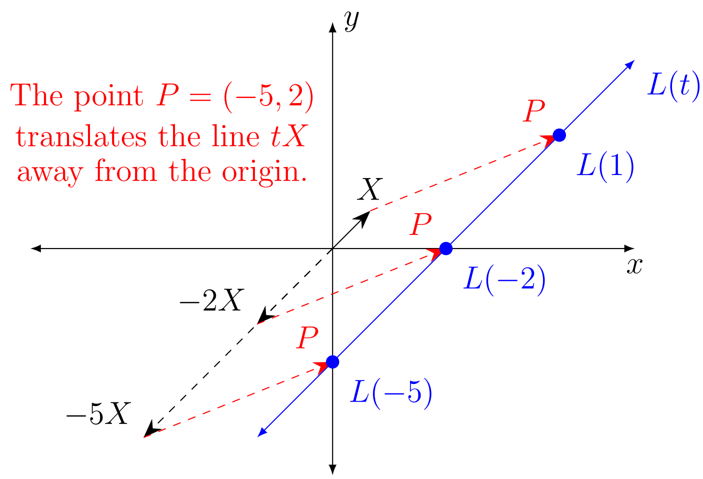
Observe that the located vector $\overrightarrow{PQ}$ beginning at the point P and in the direction of the point Q is always parallel to the vector $Q-P$ of $\mathbb{R}^n$ lying in standard position because $\overrightarrow{PQ}$ and $\overrightarrow{O(Q-P)}$ are translations of one another. Even more, the coordinates of $\overrightarrow{PQ}$ and $\overrightarrow{O(Q-P)}$ are the same, hence we have that $\|\overrightarrow{PQ}\| = \|\overrightarrow{O(Q-P)}\| = \|Q-P\|$, i.e., these vectors possess the same magnitude; however, the benefit of working with located vectors as opposed to vectors lying in standard position is that we are afforded the luxury of straying away from the origin if we work with located vectors. We will soon see that this allows us to define lines and planes in real n-space by generalizing these familiar notions from $\mathbb{R}^2$. We had tacitly assumed in the previous sections that we could freely translate vectors to and from the origin; the exposition so far in this section justifies this.

Given any point P in real n-space $\mathbb{R}^n$, any nonzero vector X lying in standard position in $\mathbb{R}^n$, and some real variable t, consider the vector equation $L(t) = tX + P$ obtained by viewing P as a vector of $\mathbb{R}^n$ lying in standard position. Graphically, for each real number t, the vector $tX + P$ can be constructed by placing the "foot" of the vector P at the "head" of the vector X and drawing the ray emanating from the origin to the "head" of P; then, by allowing t to run through all real numbers, we obtain a collection of points $L(t) = \{tX + P \mid t \text{ is a real number}\}$ that form a **line** in $\mathbb{R}^n$ in the **direction** of X **passing through** the point P. We may understand P as a translation of X by some distance in each of the coordinate axes of $\mathbb{R}^n$. Often, we will refer to the real variable t as the **parameter** (such as time), and we will refer to the vector equation $L(t) = tX + P$ as a **parametric equation** in one variable. Let us consider an example to see how this works.

Example 3.3.3. Given the point $P = (5, 2)$ and the vector $X = (1, 1)$ lying in standard position in $\mathbb{R}^2$, the **parametrization** of the line in the direction of X passing through the point P is

$$L(t) = tX + P = t(1, 1) + (5, 2) = (t, t) + (5, 2) = (t + 5, t + 2).$$

Clearly, the point $P = L(0)$ lies on the line $L(t)$. Other points on this line include the x-**intercept** $L(-2) = (3, 0)$ and the y-**intercept** $L(-5) = (0, -3)$. Pictorially, we have the following.



Crucially, observe that $L(t)$ is a line with slope 1 passing and y-intercept $(0, -3)$. Conventionally, we write $y = x - 3$; however, this example illustrates how to capture this information with vectors. One other way to see this is that $L(t) = (t + 5, t + 2)$ yields that $x(t) = t + 5$ and $y(t) = t + 2$. By solving for t in each of these identities, we find that $t = x - 5$ and $t = y - 2$. Comparing these two identities in t, we eliminate the variable t to find that $y - 2 = x - 5$ or $y = x - 3$.

Example 3.3.4. Conversely, let us illustrate how to express a line in $\mathbb{R}^2$ as a parametric equation $L(t)$. Consider the line $y = -2x + 3$ of slope -2 and y-intercept 3. By setting $x(t) = t$ and writing the pair (x, y) as a vector $L(t) = (t, -2t + 3)$, we find that $L(t) = (t, -2t) + (0, 3) = t(1, -2) + (0, 3)$, hence $L(t)$ is the line in the direction of the vector $X = (1, -2)$ passing through the point $P = (0, 3)$.

Caution: the parametric form of a line is not unique; indeed, we could have just as easily set $x(t) = -\frac{1}{2}t$ in the previous example to find that $L(t) = (-\frac{1}{2}t, t + 3) = t(-\frac{1}{2}, 1) + 3$.

Example 3.3.5. Consider the points $P = (1, 2, 3)$ and the vector $X = (-1, 0, 1)$ in $\mathbb{R}^3$. By definition, the parametric form of the line in the direction of X passing through the point P is given by

$$L(t) = tX + P = t(-1, 0, 1) + (1, 2, 3) = (-t, 0, t) + (1, 2, 3) = (-t + 1, 2, t + 3).$$

Consequently, the points $L(-1) = (2, 2, 2)$, $L(0) = (1, 2, 3) = P$, and $L(1) = (0, 2, 4)$ lie on $L(t)$. Observe that for each t-value, the y-coordinate of the line $L(t)$ is fixed at $y = 2$. Put another way,

$L(t)$ lies entirely in the plane $y = 2$. On the other hand, as a line in the xz-plane, the parametric equations $x(t) = -t + 1$ and $z(t) = t + 3$ yield that $t = 1 - x$ and $t = z - 3$ so that $z - 3 = 1 - x$ or $z = -x + 4$. By parametrizing the line $L(t)$ in terms of x, we find that $L(x) = (x, 2, -x + 4)$.

Caution: it is not possible in general for $n \geq 3$ to express a line in $\mathbb{R}^n$ in the form $y = mx + b$ or $z = mx + b$ or $z = my + b$. Explicitly, for $n \geq 3$, every line in $\mathbb{R}^n$ is given by a parametric equation of the form $L(t) = (at + b, ct + d, et + f)$ for some real numbers a, b, c, d, e, and f. By solving each of the equations $x(t) = at + b$, $y(t) = ct + d$, and $z(t) = et + f$ in order to eliminate t, we would obtain three separate linear equations in x and y, x and z, and y and z.

Example 3.3.6. We illustrate next how to determine the parametric form of a line passing through two points. We assume to this end that the two points in question are $P = (1, 1, 1)$ and $Q = (2, 2, 3)$. By definition of the line passing through P and Q, we must first determine a vector in the direction of both P and Q. Observe that the located vector $\overrightarrow{PQ} = \langle 1, 1, 2 \rangle$ does the job exactly. We must next find a point through which the line passes; by construction, two immediate options are P and Q. Choosing the point P gives the following parametric equation of the line in a real variable t.

$$L(t) = t\overrightarrow{PQ} + P = t(1, 1, 2) + (1, 1, 1) = (t, t, 2t) + (1, 1, 1) = (t + 1, t + 1, 2t + 1)$$

Choosing the point Q gives the following parametric equation of the line in a real variable s.

$$L(s) = s\overrightarrow{PQ} + Q = s(1, 1, 2) + (2, 2, 3) = (s, s, 2s) + (2, 2, 3) = (s + 2, s + 2, 2s + 3)$$

One can check that these two parametrizations constitute the same line by comparing coordinates.

$$\begin{aligned} t + 1 &= s + 2 \\ t + 1 &= s + 2 \\ 2t + 1 &= 2s + 3 \end{aligned}$$

Each of these equations yields that $t = s + 1$ or $s = t - 1$, hence the above system of equations is **consistent** (i.e., there exists a solution), and $L(t)$ and $L(s)$ represent the same line.

Perhaps the most general way to describe an object of **codimension** one (i.e., dimension $n - 1$) in real n-space $\mathbb{R}^n$ is by using the dot product. Considering that translation of objects in $\mathbb{R}^n$ does not change their inherent geometric properties, we will begin to view $\mathbb{R}^n$ as an **affine** vector space. Put simply, this means that we will not distinguish between a vector lying in standard position and a located vector lying in the plane. Generally, an affine vector space can be obtained from any vector space by "forgetting" the origin. We will henceforth refer to an $(n - 1)$-dimensional affine vector subspace H of $\mathbb{R}^n$ as a **hyperplane** of codimension one. We have already seen in $\mathbb{R}^2$ that the hyperplanes of codimension one are simply lines; likewise, in $\mathbb{R}^3$, the hyperplanes of codimension one are simply planes. Given any point P and any nonzero vector N (not necessarily lying in standard position) in $\mathbb{R}^n$, we define the hyperplane **passing through** the point P **perpendicular** to the vector N as the collection of points X in $\mathbb{R}^n$ such that $(X - P) \cdot N = 0$ or $X \cdot N = P \cdot N$. We refer to the vector N in this case as the **normal vector** to the hyperplane $X \cdot N = P \cdot N$.

Example 3.3.7. Consider the point $P = (5, 2)$ and the vector $N = (-1, 1)$ in $\mathbb{R}^2$. Every point in $\mathbb{R}^2$ is of the form $X = (x, y)$ for some real numbers x and y. By definition, the hyperplane passing through P perpendicular to N is given by the set of points $X = (x, y)$ such that $X \cdot N = P \cdot N$ or

$$-x + y = (x, y) \cdot (-1, 1) = X \cdot N = P \cdot N = (5, 2) \cdot (-1, 1) = 5(-1) + 2(1) = -3.$$

Consequently, this hyperplane is nothing more than the line $y = x - 3$ of Example 3.3.3.

Example 3.3.8. Consider the point $P = (1, 2, 3)$ and the vector $N = (0, 1, 0)$ in $\mathbb{R}^3$. Every point in $\mathbb{R}^3$ is of the form $X = (x, y, z)$ for some real numbers x, y, and z. By definition, the hyperplane passing through P perpendicular to N is given by the set of points $X = (x, y, z)$ such that

$$y = (x, y, z) \cdot (0, 1, 0) = X \cdot N = P \cdot N = (1, 2, 3) \cdot (0, 1, 0) = 2.$$

Consequently, this hyperplane is nothing more than the plane $y = 2$ of Example 3.3.5.

Example 3.3.9. Consider the point $P = (-1, 1, 5, 4)$ and the vector $N = (-1, 2, 4, 5)$ in $\mathbb{R}^4$. Every point in $\mathbb{R}^4$ is of the form $X = (w, x, y, z)$ for some real numbers w, x, y, and z. By definition, the hyperplane passing through P perpendicular to N consists of all points $X = (w, x, y, z)$ such that

$$-w + 2x + 4y + 5z = (w, x, y, z) \cdot (-1, 2, 4, 5) = X \cdot N = P \cdot N = (-1, 1, 5, 4) \cdot (-1, 2, 4, 5) = 43.$$

We can no longer visualize this hyperplane because it exists in four dimensions; however, if we set $w = 0$, then we obtain a plane $2x + 4y + 5z = 43$ called the **projection** onto the w-axis. We can likewise project onto the x-axis by declaring that $x = 0$ or onto the y-axis by declaring that $y = 0$.

Observe that for any real number t, we have that $(X - P) \cdot (tN) = t[(X - P) \cdot N]$ by Proposition 3.2.5, hence if $(X - P) \cdot N = 0$, then $(X - P) \cdot (tN) = 0$. We may therefore view the hyperplane $X \cdot N = P \cdot N$ passing through the point P perpendicular to the vector N as the hyperplane passing through the point P perpendicular to the line tN in the direction of N passing through the origin. One other thing to realize is that if we have an equation $a_1x_1 + \cdots + a_nx_n = b$ of a hyperplane in $\mathbb{R}^n$, then we may find a point $P = (b_1, \dots, b_n)$ in $\mathbb{R}^n$ for which $(b_1, \dots, b_n) \cdot (a_1, \dots, a_n) = b$. Consequently, we may view the nonzero vector $N = (a_1, \dots, a_n)$ as a normal vector to the hyperplane $a_1x_1 + \cdots + a_nx_n = b$. We demonstrate the usefulness of this observation next.

Example 3.3.10. Consider the line $y = -2x + 3$ of Example 3.3.4. Every point on this line is of the form $(x, -2x + 3)$, hence for $x = 0$, we obtain a point $P = (0, 3)$. By rearranging the equation $y = -2x + 3$, we find that $2x + y = 3$ so that $N = (2, 1)$ is a normal vector that defines this line. Explicitly, we have that $(x, y) \cdot (2, 1) = 2x + y = 3 = (0, 3) \cdot (2, 1)$, as desired.

Example 3.3.11. Consider the plane $x + y + z = -1$. Observe that the point $P = (0, 0, -1)$ lies on this plane. By reading off the coefficients of the left-hand side of the equation $x + y + z = -1$, we find that $N = (1, 1, 1)$ is a normal vector to this plane. Checking the dot product condition for the normal vector yields that $(x, y, z) \cdot (1, 1, 1) = x + y + z = -1 = (0, 0, -1) \cdot (1, 1, 1)$.

Remark 3.3.12. Our previous two examples stand as a reminder that if we want to find a point in a hyperplane $a_1x_1 + \cdots + a_nx_n = b$, we may choose $n - 1$ values for $n - 1$ of the variables $x_1, \dots, x_n$; then, we may solve for the remaining variable in terms of the chosen values of the other $n - 1$.

Given any point P and any nonzero vector N in $\mathbb{R}^n$, we may construct the hyperplane passing through P perpendicular to N. Back in college algebra, we learn that two points in $\mathbb{R}^2$ uniquely determine a line; this fact is typically stated as the **point-slope form** of the line. Likewise, it is true that three **non-collinear** points in $\mathbb{R}^3$ uniquely determine a plane. Last, we discuss a method for computing the equation of the plane determined by three non-collinear points. Given any pair of vectors $X = (x_1, x_2, x_3)$ and $Y = (y_1, y_2, y_3)$ in $\mathbb{R}^3$, we define the **vector cross product**

$$X \times Y = \begin{vmatrix} E_1 & E_2 & E_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = (x_2y_3 - x_3y_2)E_1 - (x_1y_3 - x_3y_1)E_2 + (x_1y_2 - x_2y_1)E_3$$

of the vectors X and Y as the **symbolic determinant** of the standard basis vectors E_1, E_2, and E_3 with the vectors X and Y expressed as the second and third rows, respectively. Crucially, observe that $X \times Y$ is in fact a vector in $\mathbb{R}^3$ that satisfies the following properties.

Proposition 3.3.13. *Consider any vectors X, Y, and Z in $\mathbb{R}^3$.*

1.) *We have that $X \times Y = -(Y \times X)$.*

2.) *We have that $X \times (\alpha X) = 0$ for all real numbers α.*

3.) *We have that $X \times (\alpha Y) = \alpha(X \times Y)$ for all real numbers α.*

4.) *We have that $(X + Y) \times Z = (X \times Z) + (Y \times Z)$ and $X \times (Y + Z) = (X \times Z) + (Y \times Z)$.*

5.) *We have that $(X \times Y) \cdot Z = (Z \times X) \cdot Y = (Y \times Z) \cdot X$.*

6.) *We have that $(X \times Y) \cdot X = 0$ and $(X \times Y) \cdot Y = 0$.*

Proof. Each of the first three properties follows immediately from Corollary 2.1.13 because the cross product is defined by a determinant. Likewise, the fourth property follows from Proposition 2.1.7 because the cross product $(X + Y) \times Z$ is determined by the matrix whose second row is the sum of the second row of the matrices that determine $X \times Z$ and $Y \times Z$. Consequently, it suffices to prove the fifth, sixth, and seventh properties of the cross product. We will assume that $X = (x_1, x_2, x_3)$, $Y = (y_1, y_2, y_3)$, and $Z = (z_1, z_2, z_3)$. Computing the cross products yields the following.

$$X \times Y = \begin{vmatrix} E_1 & E_2 & E_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = (x_2y_3 - x_3y_2)E_1 - (x_1y_3 - x_3y_1)E_2 + (x_1y_2 - x_2y_1)E_3$$

$$Y \times Z = \begin{vmatrix} E_1 & E_2 & E_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} = (y_2z_3 - y_3z_2)E_1 - (y_1z_3 - y_3z_1)E_2 + (y_1z_2 - y_2z_1)E_3$$

By subsequently taking the dot products, we obtain the following identities.

$$\begin{aligned} (X \times Y) \cdot Z &= (x_2y_3 - x_3y_2)z_1 - (x_1y_3 - x_3y_1)z_2 + (x_1y_2 - x_2y_1)z_3 \\ &= x_1y_2z_3 - x_1y_3z_2 + x_2y_3z_1 - x_2y_1z_3 + x_3y_1z_2 - x_3y_2z_1 \\ &= (y_2z_3 - y_3z_2)x_1 - (y_1z_3 - y_3z_1)x_2 + (y_1z_2 - y_2z_1)x_3 = (Y \times Z) \cdot X \end{aligned}$$

We omit the proof that either of these is equal to $(Z \times X) \cdot Y$ because it is the same process. Consequently, the fifth property of the cross product is established. Combining the first and fifth properties above, we conclude that $(X \times Y) \cdot X = (Y \times X) \cdot X = -(X \times Y) \cdot X$, hence we must have that $(X \times Y) \cdot X = 0$. Likewise, it follows that $(X \times Y) \cdot Y = 0$, as desired. $\square$

Consequently, the vector cross product yields a tried-and-true method to construct vectors that are orthogonal to any pair of vectors X and Y. Even more, if X and Y are nonzero, non-parallel vectors, then $X \times Y$ is a normal vector to the plane spanned by X and Y: indeed, for any vector of the form $\alpha X + \beta Y$, we have that $(X \times Y) \cdot (\alpha X + \beta Y) = \alpha(X \times Y) \cdot X + \beta(X \times Y) \cdot Y = 0$. We are now in a position to determine the plane spanned by any three non-collinear points in $\mathbb{R}^3$.

Example 3.3.14. Consider the points $P = (1, 2, 3)$, $Q = (2, 5, 0)$, and $R = (-1, 0, 3)$ in $\mathbb{R}^3$. We obtain a pair of located vectors $\overrightarrow{PQ} = \langle 1, 3, -3 \rangle$ and $\overrightarrow{PR} = \langle -2, -2, 0 \rangle$ beginning at the point P in the directions of the points Q and R, respectively. We note that if we wish to determine the equation of the plane spanned by the non-collinear points P, Q, and R, then it is enough to determine a vector normal to the vectors X and Y. By Proposition 3.3.13, we achieve this as follows.

$$N = \overrightarrow{PQ} \times \overrightarrow{PR} = \begin{vmatrix} E_1 & E_2 & E_3 \\ 1 & 3 & -3 \\ -2 & -2 & 0 \end{vmatrix} = -6E_1 + 6E_2 + 4E_3 = (-6, 6, 4)$$

Choosing any one of the points P, Q, or R and applying the definition of the plane passing through the point perpendicular to the normal vector N, we obtain the equation of the plane.

$$-6x + 6y + 4z = (x, y, z) \cdot (-6, 6, 4) = X \cdot N = P \cdot N = (1, 2, 3) \cdot (-6, 6, 4) = -6 + 12 + 12 = 18.$$

One can simplify this expression to obtain the equation of the plane $-3x + 3y + 2z = 9$.

3.4 Inner Products

Generally, there exist vector spaces other than real n-space $\mathbb{R}^n$ that admit a notion of lengths and orthogonality of vectors. Given any vector space V, an **inner product** on V is an assignment of a scalar $\langle v, w \rangle$ to each pair of vectors v and w in a manner consistent with the following properties.

1.) We have that $\langle v, w \rangle = \langle w, v \rangle$ for all vectors v and w of V.

2.) We have that $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ for all vectors u, v, and w of V.

3.) We have that $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle = \langle v, \alpha w \rangle$ for all scalars α and all vectors v and w of V.

4.) We have that $\langle v, v \rangle \geq 0$ with equality if and only if v is the zero vector of V.

We refer to the scalar $\langle v, w \rangle$ as the inner product of the vectors v and w. Often, the fourth property above is referred to in the literature as the **positive-definite** property of the inner product. We are familiar already with examples of inner product spaces from the previous sections.

Example 3.4.1. Consider the real vector space $\mathbb{R}^n$ of points in real n-space. By Proposition 3.2.5, it follows that the scalar dot product $\langle X, Y \rangle = X \cdot Y$ constitutes an inner product on $\mathbb{R}^n$.

Example 3.4.2. Consider the real vector space $\mathbb{R}^{n\times 1}$ of real $n \times 1$ column vectors. We claim that the vector dot product $\langle X, Y\rangle = X^t Y$ constitutes an inner product on $\mathbb{R}^{n\times 1}$. Considering that $X^t Y$ is a real 1×1 matrix, it follows that $\langle X, Y\rangle = X^t Y = (X^t Y)^t = Y^t X = \langle Y, X\rangle$. By Proposition 1.2.6, we have that $\langle X, Y + Z\rangle = X^t(Y + Z) = X^t Y + X^t Z = \langle X, Y\rangle + \langle X, Z\rangle$. By Proposition 1.2.5, it follows that $\langle \alpha X, Y\rangle = (\alpha X)^t = (\alpha X^t)Y = \alpha(X^t Y) = X^t(\alpha Y) = \langle X, \alpha Y\rangle$, and both of these inner products are equal to $\alpha\langle X, Y\rangle$. Last, we have that $\langle X, X\rangle = X^t X = \left[\|X\|^2\right]$ so that $\langle X, X\rangle \geq 0$ with equality if and only if X is the zero vector by the fourth part of Proposition 3.2.5.

Example 3.4.3. Consider the real vector space $\mathcal{C}^0(\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$. Given any pair of continuous functions $f(x)$ and $g(x)$, we may define an inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$. We must recall from integral calculus that the three properties of an inner product hold: explicitly, it is plain to see that $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx = \int_0^1 g(x)f(x)\,dx = \langle g, f\rangle$, and for any real number C, it holds that $\langle Cf, g\rangle = \int_0^1 (Cf(x))g(x) = C\int_0^1 f(x)g(x)\,dx = C\langle f, g\rangle = \int_0^1 f(x)(Cg(x))\,dx = \langle f, Cg\rangle$. Likewise, for any three continuous functions $f(x)$, $g(x)$, and $h(x)$, we have that

$$\int_0^1 f(x)[g(x) + h(x)]\,dx = \int_0^1 [f(x)g(x) + f(x)h(x)]\,dx = \int_0^1 f(x)g(x)\,dx + \int_0^1 f(x)h(x)\,dx$$

so that $\langle f, g + h\rangle = \langle f, g\rangle + \langle f, h\rangle$. Even though it is not clear that this inner product is positive-definite (we would have to demonstrate that $\int_0^1 [f(x)]^2\,dx = 0$ if and only if $f(x)$ is the zero function), it turns out to be the case; however, we will not bother with the details here.

We refer to a vector space V as an **inner product space** if it admits an inner product $\langle v, w\rangle$ for every pair of vectors v and w of V. Consequently, each of the real vector spaces $\mathbb{R}^n$, $\mathbb{R}^{n\times 1}$, and $\mathcal{C}^0(\mathbb{R})$ is a real inner product space. Like in the previous sections, with an arbitrary inner product on an inner product space V, we will say that a pair of vectors v and w of V are **orthogonal** (or **perpendicular**) (with respect to the underlying inner product) if it holds that $\langle v, w\rangle = 0$. We will also refer to the scalar $\|v\| = \sqrt{\langle v, v\rangle}$ as the **magnitude** of the vector v so that $\|v\|^2 = \langle v, v\rangle$. Like before, the **unit vectors** of V are precisely those vectors v satisfying that $\|v\| = 1$.

Proposition 3.4.4. *Every nonzero vector v of an inner product space V induces a unit vector $\frac{1}{\|v\|}v$.*

Proof. Compare the above properties of the inner product with the proof of Corollary 3.1.5. □

Example 3.4.5. Consider the real inner product space $\mathcal{C}^0(\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$ with respect to the inner product $\langle f, g\rangle = \int_0^\pi f(x)g(x)\,dx$. We have that

$$\|\sin x\|^2 = \langle \sin x, \sin x\rangle = \int_0^\pi \sin^2 x\,dx = \frac{1}{2}\int_0^\pi (1 - \cos(2x))\,dx = \frac{1}{2}\left[x - \frac{1}{2}\sin(2x)\right]_0^\pi = \frac{\pi}{2}.$$

Consequently, the function $\sqrt{\frac{2}{\pi}}\sin x$ is a unit vector of $\mathcal{C}^0(\mathbb{R})$. Even more, we have that

$$\langle \sin x, \cos x\rangle = \int_0^\pi \sin x \cos x\,dx = \frac{1}{2}\int_0^\pi \sin(2x)\,dx = \left[-\frac{1}{4}\cos(2x)\right]_0^\pi = 0,$$

hence the functions $\sin x$ and $\cos x$ are orthogonal vectors of $\mathcal{C}^0(\mathbb{R})$ with respect to the inner product.

Example 3.4.6. Consider the real inner product space $\mathcal{C}^0(\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$ with respect to the inner product $\langle f, g \rangle = \int_{-1}^1 f(x)g(x)\,dx$. We have that

$$\|x\|^2 = \langle x, x \rangle = \int_{-1}^1 x^2\,dx = \left[\frac{x^3}{3}\right]_{-1}^1 = \frac{2}{3}.$$

Consequently, the function $f(x) = \sqrt{\frac{3}{2}}x$ is a unit vector of $\mathcal{C}^0(\mathbb{R})$. Even more, we have that

$$\langle x, x^2 \rangle = \int_{-1}^1 x^3\,dx = \left[\frac{x^4}{4}\right]_{-1}^1 = 0,$$

hence the functions x and x^2 are orthogonal vectors of $\mathcal{C}^0(\mathbb{R})$ with respect to the inner product. Generally, the same argument shows that x^{2k} and $x^{2\ell+1}$ are orthogonal for any non-negative integers k and ℓ: indeed, $x^{2k}x^{2\ell+1} = x^{2(k+\ell)+1}$ is an odd function, so its integral over the symmetric interval $[-1, 1]$ is zero. By the same rationale, if $f(x)g(x)$ is odd, then $f(x)$ and $g(x)$ are orthogonal.

Proposition 3.4.7. *Consider any vector space V with a positive-definite inner product $\langle -, - \rangle$.*

1.) *If u is orthogonal to both v and w, then u and $v + w$ are orthogonal.*

2.) *If u and v are orthogonal, then u and αv are orthogonal for all scalars α.*

3.) (**Pythagorean Theorem**) *If u and v are orthogonal, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.*

Proof. We note that the same arguments as in Proposition 3.2.6 hold here. □

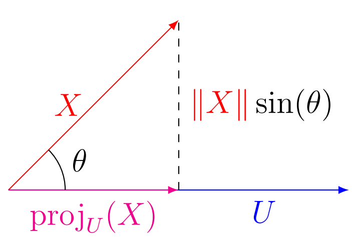
Given any pair of nonzero vectors v and w of an inner product space V, we refer to the scalar

$$\operatorname{comp}_w(v) = \frac{\langle v, w \rangle}{\langle w, w \rangle} = \frac{1}{\|w\|^2}\langle v, w \rangle$$

as the **component** of the vector v **along** the vector w; using the component of one vector along another, we may define the **projection** of the vector v **along** the vector w by setting

$$\operatorname{proj}_w(v) = \operatorname{comp}_w(v)w = \frac{\langle v, w \rangle}{\langle w, w \rangle}w.$$

Given any vectors lying in standard position in real n-space, the projection of any vector X along a unit vector U has a nice geometric interpretation. Explicitly, by Proposition 3.2.1, we have that $\langle X, U \rangle = X \cdot U = \|X\|\|U\|\cos(\theta) = \|X\|\cos(\theta)$ for the angle θ between X and U. Even more, we have that $\operatorname{proj}_U(X) = \operatorname{comp}_U(X)U = |X\|\cos(\theta)U$. Pictorially, this yields the following.



Put another way, the projection of X along U can be viewed as the "shadow" X casts along U.

Example 3.4.8. Consider the real vector space $\mathbb{R}^3$ with the inner product $\langle X, Y \rangle = X \cdot Y$. Observe that $(1, 0, 1) \cdot (1, 2, 3) = (1)(1) + (0)(2) + (1)(3) = 4$ and $(1, 2, 3) \cdot (1, 2, 3) = 1^2 + 2^2 + 3^2 = 14$. Consequently, we may find the component and the projection of $(1, 0, 1)$ along $(1, 2, 3)$ as follows.

$$\operatorname{comp}_{(1,2,3)}(1, 0, 1) = \frac{(1, 0, 1) \cdot (1, 2, 3)}{(1, 2, 3) \cdot (1, 2, 3)} = \frac{4}{14} = \frac{2}{7} \text{ and } \operatorname{proj}_{(1,2,3)}(1, 0, 1) = \frac{2}{7}(1, 2, 3)$$

Example 3.4.9. Consider the real vector space $\mathcal{C}^0(\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$ with respect to the inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx$. By definition, we have the following.

$$\operatorname{comp}_g(f) = \frac{\langle f, g \rangle}{\langle g, g \rangle} = \frac{\int_0^1 f(x)g(x)\,dx}{\int_0^1 [g(x)]^2\,dx}$$

Consider the vectors $f(x) = e^x$ and $g(x) = x$. Using integration by parts, we find that

$$\langle e^x, x \rangle = \int_0^1 xe^x\,dx = [xe^x]_0^1 - \int_0^1 e^x\,dx = e - [e^x]_0^1 = e - (e - 1) = 1.$$

Easier yet is the fact that $\langle x, x \rangle = \int_0^1 x^2\,dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$. Combined, these observations yield that

$$\operatorname{comp}_x(e^x) = \frac{\langle e^x, x \rangle}{\langle x, x \rangle} = \frac{1}{\frac{1}{3}} = 3 \text{ and } \operatorname{proj}_x(e^x) = \operatorname{comp}_x(e^x)x = 3x.$$

Our next proposition guarantees the existence of vectors orthogonal to any nonzero vector.

Proposition 3.4.10. *Consider any vector space V with a positive-definite inner product $\langle -, - \rangle$. Let v and w be any pair of nonzero vectors of V. We have that w and $v - \operatorname{proj}_w(v)$ are orthogonal.*

Proof. By definition, the projection of the vector v along the vector w is given by

$$\operatorname{proj}_w(v) = \operatorname{comp}_w(v)w = \frac{\langle v, w \rangle}{\langle w, w \rangle} w.$$

Computing the inner product of w and $v - \operatorname{proj}_w(v)$ yields the following.

$$\langle v - \operatorname{proj}_v(w), w \rangle = \left\langle v - \frac{\langle v, w \rangle}{\langle w, w \rangle} w, w \right\rangle = \langle v, w \rangle - \left\langle \frac{\langle v, w \rangle}{\langle w, w \rangle} w, w \right\rangle = \langle v, w \rangle - \frac{\langle v, w \rangle}{\langle w, w \rangle} \langle w, w \rangle = 0$$

We conclude that w and $v - \operatorname{proj}_w(v)$ are orthogonal, as desired. □

Even though we will not make explicit use of the following inequalities, their ubiquity in the fields of complex analysis, functional analysis, mathematical physics, partial differential equations, and many more areas of applied mathematics necessitate their inclusion in these lecture notes. Often in the literature, the following is abbreviated as the Schwarz Inequality; however, the inequality was discovered independently by each of the three eponymous mathematicians in chronological order according to the appearance of their names in the title, hence we provide the full name.

Theorem 3.4.11 (Cauchy-Bunyakovsky-Schwarz Inequality). *Consider any vector space V with a positive-definite inner product $\langle -, - \rangle$. We have that $\langle v, w \rangle^2 \leq \langle v, v \rangle \langle w, w \rangle$ for any vectors $v, w \in V$.*

Proof. Certainly, if either v or w is the zero vector, then we have that $\langle v, w \rangle = 0$ so that both the left- and right-hand sides of the desired inequality are zero. We may assume therefore that neither v nor w is the zero vector. Consider the projection $\operatorname{proj}_w(v) =$ of the vector v along the vector w. We may write $v = (v - \operatorname{proj}_w(v)) + \operatorname{proj}_w(v)$. By Proposition 3.4.10, it follows that $v - \operatorname{proj}_w(v)$ is orthogonal to w, hence $v - \operatorname{proj}_w(v)$ is orthogonal to $\operatorname{proj}_w(v)$ by Proposition 3.4.7; thus, by the Pythagorean Theorem for inner product spaces (cf. the aforementioned proposition), it follows that

$$\|v\|^2 = \|(v - \operatorname{proj}_w(v)) + \operatorname{proj}_w(v)\|^2 = \|v - \operatorname{proj}_w(v)\|^2 + \|\operatorname{proj}_w(v)\|^2.$$

By hypothesis that $\langle -, - \rangle$ is a positive-definite inner product, it follows that

$$\|v\|^2 = \|v - \operatorname{proj}_w(v)\|^2 + \|\operatorname{proj}_w(v)\|^2 \geq \|\operatorname{proj}_w(v)\|^2.$$

Explicit computation of the scalar $\|\operatorname{proj}_w(v)\|^2$ yields the following.

$$\|\operatorname{proj}_w(v)\|^2 = \langle \operatorname{proj}_w(v), \operatorname{proj}_w(v) \rangle = \left\langle \frac{\langle v, w \rangle}{\langle w, w \rangle} w, \frac{\langle v, w \rangle}{\langle w, w \rangle} w \right\rangle = \frac{\langle v, w \rangle^2}{\langle w, w \rangle^2} \langle w, w \rangle = \frac{\langle v, w \rangle^2}{\langle w, w \rangle}$$

Consequently, it follows that $\langle v, v \rangle = \|v\|^2 \geq \|\operatorname{proj}_w(v)\|^2 = \frac{\langle v, w \rangle^2}{\langle w, w \rangle}$ or $\langle v, w \rangle^2 \leq \langle v, v \rangle \langle w, w \rangle$. ☐

Theorem 3.4.12 (Triangle Inequality)**.** *Consider any vector space V with a positive-definite inner product $\langle -, - \rangle$. We have that $\|v + w\| \leq \|v\| + \|w\|$ for any vectors v and w of V.*

Proof. Expanding the inner product $\|v + w\|^2 = \langle v + w, v + w \rangle$, we find that

$$\|v + w\|^2 = \langle v + w, v + w \rangle = \langle v, v \rangle + 2\langle v, w \rangle + \langle w, w \rangle \leq \langle v, v \rangle + 2|\langle v, w \rangle| + \langle w, w \rangle.$$

By the Cauchy-Bunyakovsky-Schwarz Inequality, we have that $|\langle v, w \rangle| \leq \sqrt{\langle v, v \rangle}\sqrt{\langle w, w \rangle}$ so that $2|\langle v, w \rangle| \leq 2\sqrt{\langle v, v \rangle}\sqrt{\langle w, w \rangle}$ and $\langle v, v \rangle + 2|\langle v, w \rangle| + \langle w, w \rangle \leq \langle v, v \rangle + \sqrt{\langle v, v \rangle}\sqrt{\langle w, w \rangle} + \langle w, w \rangle$. We conclude that $\|v + w\|^2 \leq \langle v, v \rangle + 2|\langle v, w \rangle| + \langle w, w \rangle = (\sqrt{\langle v, v \rangle} + \sqrt{\langle w, w \rangle})^2 = (\|v\| + \|w\|)^2$. ☐

3.5 Orthogonal Bases and the Gram-Schmidt Process

We will continue to assume that V is a vector space that admits a positive-definite inner product $\langle -, - \rangle$. Explicitly, for any pair of vectors v and w in V, we have that $\langle v, w \rangle$ is a scalar called the inner product of v and w; a vector space that admits an inner product is called an inner product space. Common examples of real inner product spaces include real n-space $\mathbb{R}^n$ equipped with the scalar dot product $\langle X, Y \rangle = X \cdot Y$; the real vector space $\mathbb{R}^{n \times 1}$ of real $n \times 1$ column vectors with the vector dot product $\langle X, Y \rangle = X^t Y$; and the real vector space $\mathcal{C}^0(\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$ with the inner product $\langle f, g \rangle = \int_a^b f(x)g(x)\,dx$ for any pair of real numbers $a < b$. We know from the previous section that an inner product must be commutative and distributive with respect to vector addition and scalar multiplication, i.e., we must have that $\langle v, w \rangle = \langle w, v \rangle$ and $\langle \alpha u + v, w \rangle = \alpha \langle u, w \rangle + \langle v, w \rangle$ for all scalars α and all vectors u, v, and w in V. We say that the vectors v and w are orthogonal if and only if their inner product is zero if and only if $\langle v, w \rangle = 0$ (cf. Examples 3.2.2, 2.6.4, 3.4.5, and 3.4.6 for some instances of orthogonal vectors).

By the Spectral Theorem, every real symmetric $n \times n$ matrix induces a basis of $\mathbb{R}^{n\times 1}$ consisting of eigenvectors with the additional property that every pair of distinct eigenvectors are orthogonal. Explicitly, every real symmetric matrix is orthogonally diagonalizable; however, it is not the case that the eigenvectors of a real symmetric matrix must be orthogonal: the real symmetric matrix

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

induces a basis of eigenvectors $X_1 = (1, 0, -1)$, $X_2 = (0, 1, -1)$, and $X_3 = (1, 1, 1)$, but the eigenvectors X_1 and X_2 are not orthogonal because the dot product $X_1 \cdot X_2 = (1)(0)+(0)(1)+(-1)(-1) = 1$ is nonzero. We must therefore find some method by which to convert these eigenvectors into some eigenvectors that are orthogonal to one another; this is called the **Gram-Schmidt Process**.

Example 3.5.1. Consider the vectors $X_1 = (1, 0, -1)$, $X_2 = (0, 1, -1)$, and $X_3 = (1, 1, 1)$ of $\mathbb{R}^3$. We will perform the Gram-Schmidt Process to product three unit vectors that are orthogonal to one another. We begin with the vector $X_1 = (1, 0, -1)$ of magnitude $\|X\| = \sqrt{1^2 + 0^2 + (-1)^2} = \sqrt{2}$. By Corollary 3.1.5, it follows that $U_1 = \frac{1}{\sqrt{2}} X_1$ is a unit vector; even more importantly, the vector U_1 remains an eigenvector of A corresponding to the eigenvalue 0 because $AU_1 = O = 0 \cdot U_1$. We must next produce a unit vector U_2 that is orthogonal to U_1. Crucially, Proposition 3.4.10 guarantees that the vector $X_2 - \operatorname{proj}_{X_1}(X_2)$ is orthogonal to X_1; moreover, it satisfies that

$$A(X_2 - \operatorname{proj}_{X_1}(X_2)) = AX_2 - A \operatorname{proj}_{X_1}(X_2) = O - A\left(\frac{X_1 \cdot X_2}{X_1 \cdot X_1} X_2\right) = \frac{X_1 \cdot X_2}{X_1 \cdot X_1} AX_2 = O$$

so that $X_2 - \operatorname{proj}_{X_1}(X_2)$ is an eigenvector of A corresponding to the eigenvalue 0. Consequently,

$$U_2 = \frac{1}{\|X_2 - \operatorname{proj}_{X_1}(X_2)\|}(X_2 - \operatorname{proj}_{X_1}(X_2))$$

is a unit vector orthogonal to U_1 because X_1 and $X_2 - \operatorname{proj}_{X_1}(X_2)$ are orthogonal. Even though the closed form expression of this vector is less than ideal (because the closed form expression of $X_2 - \operatorname{proj}_{X_1}(X_2)$ and its magnitude are quite awful numerically), we provide the details as follows.

$$X_2 - \operatorname{proj}_{X_1}(X_2) = X_2 - \frac{X_1 \cdot X_2}{X_1 \cdot X_1} X_1 = (0, 1, -1) - \frac{1}{2}(1, 0, -1) = \left(-\frac{1}{2}, 1, -\frac{1}{2}\right)$$

$$\|X_2 - \operatorname{proj}_{X_1}(X_2)\| = \sqrt{\left(-\frac{1}{2}\right)^2 + 1^2 + \left(-\frac{1}{2}\right)^2} = \sqrt{\frac{3}{2}}$$

Last, we construct a unit vector U_3 that is orthogonal to both of the vectors U_1 and U_2. Considering that $X_1 \cdot X_3 = 0$ and $X_2 \cdot X_3 = 0$, it suffices to take $U_3 = \frac{1}{\|X_3\|} X_3 = \frac{1}{\sqrt{3}}(1, 1, 1)$.

Generally, we will refer to a basis $v_1, \dots, v_n$ of a vector space V as **orthogonal** if it holds that $\langle v_i, v_j \rangle = 0$ for all integers $1 \le i < j \le n$. Even more, if the vectors $v_1, \dots, v_n$ are all unit vectors (i.e., we have that $\|v_i\| = 1$ for each integer $1 \le i \le n$), then we say that the basis vectors $v_1, \dots, v_n$ form an **orthonormal** basis. We are already familiar with examples of orthonormal bases.

Example 3.5.2. Consider the standard basis of $\mathbb{R}^{n \times 1}$ given by the $n \times 1$ column vectors $E_1, \dots, E_n$ such that E_i consists of one in the ith row and zeros elsewhere. By definition, we have that

$$\|E_i\| = \sqrt{0^2 + \cdots + 0^2 + 1^2 + 0^2 + \cdots + 0^2} = 1,$$

hence E_i is a unit vector. Even more, for any integers $1 \leq i < j \leq n$, we have that

$$E_i \cdot E_j = (0)(0) + \cdots + (1)(0) + (0)(0) + \cdots + (0)(1) + (0)(0) + \cdots + (0)(0) = 0,$$

hence E_i and E_j are orthogonal. We conclude that $E_1, \dots, E_n$ form an orthonormal basis for $\mathbb{R}^{n \times 1}$.

Example 3.5.3. Consider the real vector space $\mathcal{C}^0(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ with the positive-definite inner product $\langle f, g \rangle = \int_0^\pi f(x)g(x)\,dx$. Consider the vector subspace W of $\mathcal{C}^0(\mathbb{R})$ spanned by the linearly independent vectors $\sin x$ and $\cos x$ of $\mathcal{C}^0(\mathbb{R})$. By Example 3.4.5, we have that $\sin x$ and $\cos x$ are orthogonal, hence in order to find an orthonormal basis for the vector space spanned by $\sin x$ and $\cos x$, it suffices to compute $\|\sin x\|$ and $\|\cos x\|$. By the same example, we know that $\sqrt{\frac{2}{\pi}} \sin x$ is a unit vector. Likewise, we may compute the magnitude of $\cos x$

$$\|\cos x\|^2 = \langle \cos x, \cos x \rangle = \int_0^\pi \cos^2 x\,dx = \frac{1}{2}\int_0^\pi (1 + \cos(2x)) = \frac{1}{2}\left[x + \frac{1}{2}\sin(2x)\right]_0^\pi = \frac{\pi}{2}$$

so that $\sqrt{\frac{2}{\pi}} \cos x$ is a unit vector and $\left\{\sqrt{\frac{2}{\pi}} \sin x, \sqrt{\frac{2}{\pi}} \cos x\right\}$ is an orthonormal basis for W.

Every nonzero finite-dimensional inner product space admits an orthonormal basis as follows.

Theorem 3.5.4 (Gram-Schmidt Process)**.** *Every nonzero finite-dimensional vector space with a positive-definite inner product $\langle -, - \rangle$ admits an orthonormal basis with respect to $\langle -, - \rangle$.*

Proof. By Theorem 1.8.10, there exist nonzero vectors $v_1, \dots, v_n$ that constitute a basis for V. We may successively replace each basis vector v_i with a nonzero vector w_i that is orthogonal to the vectors $w_1, \dots, w_{i-1}$ for each integer $2 \leq i \leq n$ to obtain an orthogonal basis as follows.

$$w_i = v_i - \sum_{j=1}^{i-1} c_{ij} w_j = v_i - \sum_{j=1}^{i-1} \frac{\langle v_i, w_j \rangle}{\langle w_j, w_j \rangle} w_j = v_i - \frac{\langle v_i, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \cdots - \frac{\langle v_i, w_{i-1} \rangle}{\langle w_{i-1}, w_{i-1} \rangle} w_{i-1}.$$

Explicitly, we have that $w_1 = v_1$ because the summation is empty for $i = 1$, and we have that

$$w_2 = v_2 - \frac{\langle v_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 \text{ and } w_3 = v_3 - \frac{\langle v_3, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle v_3, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2.$$

Each of the nonzero vectors w_i is orthogonal to the vectors $w_i, \dots, w_{i-1}$ for each integer $2 \leq i \leq n$ by construction. Crucially, we note that for each integer $1 \leq j \leq i - 1$, we have that

$$\langle v_i, w_j \rangle - c_{ij} \langle w_j, w_j \rangle = \langle v_i, w_j \rangle - \frac{\langle v_i, w_j \rangle}{\langle w_j, w_j \rangle} \langle w_j, w_j \rangle = \langle v_i, w_j \rangle - \langle v_i, w_j \rangle = 0.$$

Consequently, it suffices to prove that w_1 and w_2 are orthogonal: indeed, we have that

$$\langle w_2, w_1 \rangle = \langle v_2 - c_{21} w_1, w_1 \rangle = \langle v_2, w_1 \rangle - c_{21} \langle w_1, w_1 \rangle = 0.$$

Even more, it follows that w_3 is orthogonal to w_1 and w_2 by construction of w_3: we have that

$$\langle w_3, w_1\rangle = \langle v_3 - c_{31}w_1 - c_{32}w_2, w_1\rangle = \langle v_3, w_1\rangle - c_{31}\langle w_1, w_1\rangle - c_{32}\langle w_2, w_1\rangle = 0 \text{ and}$$
$$\langle w_3, w_2\rangle = \langle v_3 - c_{31}w_1 - c_{32}w_2, w_2\rangle = \langle v_3, w_2\rangle - c_{31}\langle w_1, w_2\rangle - c_{32}\langle w_2, w_2\rangle = 0.$$

Continuing in this manner reveals that w_i is orthogonal to $w_1, \dots, w_{i-1}$ for each integer $2 \le i \le n$, as desired. Each of the basis vectors v_i can be written as $v_i = c_{i1}w_1 + \cdots + c_{i,i-1}w_{i-1}$, hence we find that $V = \operatorname{span}\{v_1, \dots, v_n\} = \operatorname{span}\{w_1, \dots, w_n\}$; thus, the orthogonal vectors $w_1, \dots, w_n$ must form an orthogonal basis for V by the third part of Theorem 1.8.10. Last, we obtain an orthonormal basis for V by replacing each of the vectors w_i with the unit vector $u_i = \frac{1}{\|w_i\|}w_i$ by Corollary 3.1.5. □

Example 3.5.5. Let us carry out the Gram-Schmidt Process to find an orthonormal basis for the vector subspace W of $\mathbb{R}^3$ spanned by the linearly independent vectors $X = (1, 2, -1)$, $Y = (-1, 3, 2)$, and $Z = (2, 4, 3)$. We may choose any of the three vectors as our initial vector to construct the basis; we pick $X_1 = X$ arbitrarily; then, we convert Y into a vector orthogonal to X_1 as follows.

$$X_2 = Y - \frac{\langle Y, X_1\rangle}{\langle X_1, X_1\rangle}X_1 = (-1, 3, 2) - \frac{(-1, 3, 2)\cdot(1, 2, -1)}{(1, 2, -1)\cdot(1, 2, -1)}(1, 2, -1) = (-1, 3, 2) - \frac{3}{6}(1, 2, -1)$$

Carrying out the simplification and the subtraction yields that $X_2 = \left(-\frac{3}{2}, 2, \frac{5}{2}\right)$. Likewise, we may convert Z into a vector orthogonal to both X_1 and X_2 as follows.

$$\begin{aligned}
X_3 &= Z - \frac{\langle Z, X_1\rangle}{\langle X_1, X_1\rangle}X_1 - \frac{\langle Z, X_2\rangle}{\langle X_2, X_2\rangle}X_2 \\
&= (2, 4, 3) - \frac{(2, 4, 3)\cdot(1, 2, -1)}{6}(1, 2, -1) - \frac{(2, 4, 3)\cdot\left(-\frac{3}{2}, 2, \frac{5}{2}\right)}{\left(-\frac{3}{2}, 2, \frac{5}{2}\right)\cdot\left(-\frac{3}{2}, 2, \frac{5}{2}\right)}\left(-\frac{3}{2}, 2, \frac{5}{2}\right) \\
&= (2, 4, 3) - \frac{7}{6}(1, 2, -1) - \frac{\frac{25}{2}}{\frac{25}{2}}\left(-\frac{3}{2}, 2, \frac{5}{2}\right) \\
&= (2, 4, 3) - \left(\frac{7}{6}, \frac{7}{3}, -\frac{7}{6}\right) - \left(-\frac{3}{2}, 2, \frac{5}{2}\right) = \left(\frac{7}{3}, -\frac{1}{3}, \frac{5}{3}\right)
\end{aligned}$$

We have therefore obtained an orthogonal basis. Last, we obtain an orthonormal basis by dividing each of these basis vectors by its magnitude; the orthonormal basis vectors are

$$U_1 = \frac{1}{\sqrt{6}}(1, 2, -1) \text{ and } U_2 = \frac{\sqrt{2}}{5}\left(-\frac{3}{2}, 2, \frac{5}{2}\right) \text{ and } U_3 = \frac{\sqrt{3}}{5}\left(\frac{7}{3}, -\frac{1}{3}, \frac{5}{3}\right).$$

Example 3.5.6. Let us carry out the Gram-Schmidt Process to find an orthonormal basis for the vector subspace W of $\mathcal{C}^0(\mathbb{R})$ spanned by the linearly independent vectors t and t^2 with respect to the inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$. We may choose t or t^2 as our initial vector to construct the basis; we pick $f_1(t) = t$ arbitrarily; then, we convert t^2 into a vector orthogonal to $f_1(t)$ as follows.

$$f_2(t) = t^2 - \frac{\langle t^2, f_1(t)\rangle}{\langle f_1(t), f_1(t)\rangle}f_1(t) = t^2 - \frac{\langle t^2, t\rangle}{\langle t, t\rangle}t = t^2 - \frac{\int_0^1 t^3\,dt}{\int_0^1 t^2\,dt}t = \frac{\frac{1}{4}}{\frac{1}{3}}t = t^2 - \frac{3}{4}t$$

Consequently, the vectors t and $t^2 - \frac{3}{4}t$ form an orthogonal basis for W. Last, we obtain an orthonormal basis as $\frac{1}{\|t\|}t$ and $\frac{1}{\|t^2-\frac{3}{4}t\|}(t^2 - \frac{3}{4}t)$; thus, the orthonormal basis vectors are

$$u_1(t) = \frac{1}{\int_0^1 t^2\,dt}t = 3t \text{ and } u_2(t) = \frac{1}{\int_0^1 \left(t^2 - \frac{3}{4}t\right)^2 dt}\left(t^2 - \frac{3}{4}t\right) = 80\left(t^2 - \frac{3}{4}t\right).$$

We turn our attention next to the general theory of orthogonal vectors of an arbitrary inner product space. One of the premier reasons to work with orthogonal vectors is the following.

Proposition 3.5.7. *Consider a vector space V with a positive-definite inner product $\langle -, - \rangle$. Given any nonzero vectors $v_1, \dots, v_n$ of V such that v_i and v_j are orthogonal for all integers $1 \leq i < j \leq n$, if there exist scalars $\alpha_1, \dots, \alpha_n$ such that $\alpha_1 v_1 + \cdots + \alpha_n v_n = O_V$, we must have $\alpha_1 = \cdots = \alpha_n = 0$. Consequently, nonzero orthogonal vectors of an inner product space are linearly independent.*

Proof. Consider any expression $\alpha_1 v_1 + \cdots + \alpha_n v_n = O$ of linear dependence among a collection $v_1, \dots, v_n$ of nonzero mutually orthogonal vectors. Expanding the inner product, we have that

$$0 = \langle v_i, O \rangle = \langle v_i, \alpha_1 v_1 + \cdots + \alpha_n v_n \rangle = \alpha_1 \langle v_i, v_1 \rangle + \cdots + \alpha_n \langle v_i, v_n \rangle.$$

Considering that the inner products $\langle v_i, v_j \rangle = 0$ for each integer $j \neq i$ by assumption, we conclude that $\alpha_i \langle v_i, v_i \rangle = 0$. By hypothesis that v_i is nonzero, we can divide by the nonzero scalar $\langle v_i, v_i \rangle$ to find that $\alpha_i = 0$. Continuing this for each integer $1 \leq i \leq n$ yields that $\alpha_1 = \cdots = \alpha_n = 0$. ☐

Even more, if we restrict our attention to orthonormal vectors, we have the following.

Corollary 3.5.8. *Consider any vector space V with a positive-definite inner product $\langle -, - \rangle$. Given any unit vectors $u_1, \dots, u_n$ of V such that u_i and u_j are orthogonal for all integers $1 \leq i < j \leq n$, the coefficients of any vector $v = \alpha_1 u_1 + \cdots + \alpha_n u_n$ are unique. Particularly, we have $\alpha_i = \langle v, u_i \rangle$.*

Proof. Consider any vector of V of the form $v = \alpha_1 u_1 + \cdots + \alpha_n u_n$. By the proof of Proposition 3.5.7, we have that $\langle v, u_i \rangle = \alpha_i \langle u_i, u_i \rangle = \alpha_i$ by assumption that u_i is a unit vector. ☐

Combined, the previous proposition and corollary assert that the matrix representation of a linear transformation $T : V \to V$ with respect to an orthonormal basis $u_1, \dots, u_n$ of an inner product space is simply the matrix whose (i, j)th entry is the inner product $\langle T(u_i), u_j \rangle$. We will return to this notion in the next section. We conclude our present discussion with an important decomposition theorem regarding finite-dimensional inner product spaces. We will say that a vector subspace W of an inner product space V is **orthogonal** to a vector subspace U of V if it is the case that for every vector w of W and every vector u of U, we have that $\langle u, w \rangle = 0$. Before we state the next theorem, we must recall also that the sum of the vector subspaces U and W is the vector subspace $U + W = \{u + w \mid u \in U \text{ and } w \in W\}$ of V; if $U \cap W = \{O\}$, we write $U + W = U \oplus W$.

Theorem 3.5.9. *Consider any finite-dimensional vector space V with a positive-definite inner product $\langle -, - \rangle$. Given any vector subspace W of V, there exists a unique vector subspace $W^\perp$ of V that is orthogonal to W and satisfies that $V = W \oplus W^\perp$. We refer to the vector space $W^\perp$ as the* **orthogonal complement** *of W (and vice-versa). Put another way, every finite-dimensional inner product space can be decomposed as the direct sum of any subspace and its orthogonal complement.*

Proof. Certainly, the orthogonal complement of the zero subspace is V itself, hence we may assume that W is a nonzero subspace of V. By Theorem 1.8.10, for any basis $w_1, \dots, w_k$ of W, there exist nonzero vectors $v_{k+1}, \dots, v_n$ such that $w_1, \dots, w_k, v_{k+1}, \dots, v_n$ constitute a basis for V. By the Gram-Schmidt Process, we may obtain an orthogonal basis $u_1, \dots, u_k, u_{k+1}, \dots, u_n$ for V satisfying that $W = \operatorname{span}\{w_1, \dots, w_k\} = \operatorname{span}\{u_1, \dots, u_k\}$ and $W^\perp = \operatorname{span}\{u_{k+1}, \dots, u_n\}$. Even more, every vector v of V can be written as $v = \alpha_1 u_1 + \cdots + \alpha_k v_k + \alpha_{k+1} u_{k+1} + \cdots + \alpha_n u_n$ so that $V = W + W^\perp$ because the vector $\alpha_1 u_1 + \cdots + \alpha_k u_k$ lies in W, and the vector $\alpha_{k+1} u_{k+1} + \cdots + \alpha_n u_n$ lies in $W^\perp$. Given any vector $v \in W \cap W^\perp$, there exist scalars $\alpha_1, \dots, \alpha_k$ and $\alpha_{k+1}, \dots, \alpha_n$ such that $v = \alpha_1 u_1 + \cdots + \alpha_k v_k$ and $v = \alpha_{k+1} u_{k+1} + \cdots + \alpha_n u_n$. By taking the inner product of v with itself, we obtain

$$\langle v, v \rangle = \langle \alpha_1 u_1 + \cdots + \alpha_k u_k, \alpha_{k+1} u_{k+1}, \dots, \alpha_n u_n \rangle = \alpha_1 \alpha_{k+1} \langle u_1, u_{k+1} \rangle + \cdots + \alpha_k \alpha_n \langle u_k, u_n \rangle.$$

Each of the inner products $\langle u_i, u_j \rangle$ is zero for all integers $1 \leq i \leq k$ and $k+1 \leq j \leq n$ by construction, hence we conclude that $\langle v, v \rangle = 0$ so that $v = O_V$ by definition of a positive-definite inner product. We conclude therefore that $V = W \oplus W^\perp$. Last, we assert the uniqueness of $W^\perp$. We will demonstrate that every vector of V that is orthogonal to every vector of W lies in $W^\perp$; therefore, any vector subspace of V that is orthogonal to W must be contained in $W^\perp$. Every vector of V can be written as $v = w_1 + w_2$ for some vectors $w_1 \in W$ and $w_2 \in W^\perp$. Given that v is orthogonal to every vector of W, we must have that $0 = \langle v, w_1 \rangle = \langle w_1, w_1 \rangle + \langle w_1, w_2 \rangle = \langle w_1, w_1 \rangle$ so that w_1 is the zero vector, and the vector $v = w_2$ lies in $W^\perp$, as desired. □

Corollary 3.5.10. *Consider any finite-dimensional vector space V with a positive-definite inner product $\langle -, - \rangle$. Given any vector subspace W of V, we have that* $\dim(V) = \dim(W) + \dim(W^\perp)$.

Example 3.5.11. Let us determine the orthogonal complement of $W = \operatorname{span}\{(1, 0, 1), (1, 2, -2)\}$ in the vector space $\mathbb{R}^3$. By the proof of Theorem 3.5.9, we must first extended the basis of W to a basis of V. Every vector of W is of the form $(a + b, 2b, a - 2b) = a(1, 0, 1) + b(1, 2, -2)$ for some real numbers a and b, hence it suffices to choose a vector that is not of this form. Consequently, if we want the second coordinate of our vector to be 0, then in order for this vector to lie in W, it must be of the form $(a, 0, a)$. We conclude therefore that $(1, 0, -1)$ does not lie in W so that $(1, 0, 1), (1, 2, -2)$, and $(1, 0, -1)$ form a basis for $\mathbb{R}^3$. By the Gram-Schmidt Process, we find that

$$(1, 2, -2) - \frac{(1, 2, -2) \cdot (1, 0, 1)}{(1, 0, 1) \cdot (1, 0, 1)}(1, 0, 1) = (1, 2, -2) + \frac{1}{2}(1, 0, 1) = \left(\frac{3}{2}, 2, -\frac{3}{2}\right)$$

is orthogonal to $(1, 0, 1)$; it is also orthogonal to $(1, 0, -1)$ by inspection, so we have produced an orthogonal basis for V. By Corollary 3.5.10, we have that $\dim W^\perp = \dim(\mathbb{R}^3) - \dim(W) = 3 - 2 = 1$, hence we conclude that $W^\perp = \operatorname{span}\{(1, 0, -1)\}$ is the orthogonal complement of W.

Example 3.5.12. We conclude this section with an example to determine the orthogonal complement of $W = \operatorname{span}\{(1, 1, 1)\}$ in the vector space $\mathbb{R}^3$ in a different manner than the previous example. By definition, we seek all vectors (x, y, z) of $\mathbb{R}^3$ satisfying that $x + y + z = (x, y, z) \cdot (1, 1, 1) = 0$. Clearly, we have that $(1, -1, 0)$ and $(1, 0, -1)$ satisfy the aforementioned equation, hence they are orthogonal to $(1, 1, 1)$. Even more, they are linearly independent, hence they span a vector space of dimension two. By Corollary 3.5.10, we need only demonstrate that the vectors $(a, a, a) = a(1, 1, 1)$ and $(b + c, -b, -c) = b(1, -1, 0) + c(1, 0, -1)$ are orthogonal. But this is clear by taking the dot product: indeed, we have that $(a, a, a) \cdot (b + c, -b, -c) = a(b + c) - ab - ac = ab + ac - ab - ac = 0$.

3.6 Linear Functionals

Quite sneakily, we have yet to address the meaning of the scalars with which we have worked since the very first chapter of these lecture notes. We must at last deal with this situation. We say that a nonempty set k is a **field** if every pair of elements α and β of k can be added, subtracted, multiplied, and (if either α or β is nonzero) divided. Explicitly, if α and β are nonzero elements of k, then there must exist elements $-\beta$ and β^{-1} of k such that $\beta + (-\beta) = 0$, $\beta\beta^{-1} = 1$, and $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$, and $\alpha\beta^{-1}$ are all elements of k. We have throughout these notes dealt exclusively with the field $\mathbb{R}$ of real numbers, but there are other fields such as the rational numbers $\mathbb{Q}$ or the complex numbers $\mathbb{C}$ that are of interest in linear algebra. We refer to a vector space V with scalars in the field k as a k-**vector space**. We have simply said "real vector space" to mean an $\mathbb{R}$-vector space. We say that a linear transformation $f : V \to k$ is a **linear functional**. Explicitly, a linear functional is a function $f : V \to k$ from a vector space V to its field of scalars k satisfying that

1.) $f(v + w) = f(v) + f(w)$ for all vectors v and w of V and

2.) $f(\alpha v) = \alpha f(v)$ for all scalars α of k and all vectors v of V.

Example 3.6.1. Every line through the origin in the Cartesian plane $\mathbb{R}^2$ defines a linear functional on $\mathbb{R}$. Explicitly, for any real number m, the linear function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = mx$ is a linear functional because it satisfies that $f(x + y) = m(x + y) = mx + my = f(x) + f(y)$ and $f(\alpha x) = m(\alpha x) = \alpha(mx) = \alpha f(x)$ for all real numbers $\alpha, x,$ and y. Conversely, if $f : \mathbb{R} \to \mathbb{R}$ is a linear functional, then $f(x) = f(1)x$ for all real numbers x by the second property of linear functionals above, hence $f(x)$ is a line of slope $f(1)$ and y-intercept 0. Consequently, the linear functionals on the real vector space $\mathbb{R}$ are precisely the linear functions passing through the origin.

Example 3.6.2. We note that the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(x, y, z) = x + y + z$ is a linear functional: indeed, we have that $f(\alpha x, \alpha y, \alpha z) = \alpha x + \alpha y + \alpha z = \alpha(x + y + z) = \alpha f(x, y, z)$ and $f(x_1 + x_2, y_1 + y_2, z_1 + z_2) = (x_1 + x_2) + (y_1 + y_2) + (z_1 + z_2) = (x_1 + y_1 + z_1) + (x_2 + y_2 + z_2)$ or $f(x_1 + x_2, y_1 + y_2, z_1 + z_2) = f(x_1, y_1, z_1) + f(x_2, y_2, z_2)$ for all real numbers $\alpha, x_1, y_1, z_1, x_2, y_2, z_2$.

Example 3.6.3. Consider the function $E : F(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$ defined by $E(f(x)) = f(0)$ on the real vector space $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$. We can easily verify that E is a linear functional on $F(\mathbb{R}, \mathbb{R})$ because it is clear that $E(f(x) + g(x)) = E((f + g)(x)) = (f + g)(0) = f(0) + g(0)$ and $E(\alpha f(x)) = \alpha f(0) = \alpha E(f(x))$ for all real numbers α and all real functions $f(x)$ and $g(x)$.

Example 3.6.4. We define the **trace** an $n \times n$ matrix as the sum of its diagonal components.

$$\operatorname{trace}\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{nn} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}\right) = a_{11} + a_{22} + \cdots + a_{nn}$$

We claim that the function $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ defined by $f(A) = \operatorname{trace}(A)$ is a linear functional. Given any pair of real $n \times n$ matrices A and B, the diagonal components of $A + B$ are by definition the sums of the diagonal components of A and B. Explicitly, we have that $(A + B)_{ij} = A_{ij} + B_{ij}$ for all integers $1 \leq i, j \leq n$. Consequently, we have that $\operatorname{trace}(A + B) = (A + B)_{11} + \cdots + (A + B)_{nn}$ and

$$\operatorname{trace}(A) + \operatorname{trace}(B) = (A_{11} + \cdots + A_{nn}) + (B_{11} + \cdots + B_{nn}) = (A_{11} + B_{11}) + \cdots + (A_{nn} + B_{nn}).$$

We conclude that $f(A+B) = \operatorname{trace}(A+B) = \operatorname{trace}(A)+\operatorname{trace}(B) = f(A)+f(B)$. Even more, if α is any scalar, then the diagonal components of αA are by definition α times the diagonal components of A. Explicitly, we have that $(\alpha A)_{ij} = \alpha(A_{ij})$ for all integers $1 \leq i, j \leq n$ so that

$$\operatorname{trace}(\alpha A) = (\alpha A)_{11} + \cdots (\alpha A)_{nn} = \alpha(A_{11}) + \cdots + \alpha(A_{nn}) = \alpha(A_{11} + \cdots + A_{nn}) = \alpha \operatorname{trace}(A).$$

By the above displayed equation, it follows that $f(\alpha A) = \operatorname{trace}(\alpha A) = \alpha \operatorname{trace}(A) = \alpha f(A)$.

Example 3.6.5. Consider the linear functional $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(1,0,0) = 1$, $f(0,1,0) = -1$, and $f(0,0,1) = 2$. By Proposition 1.11.17, every linear transformation is uniquely determined by how it acts on a basis, hence the information provided is enough to uniquely determined $f(x,y,z)$ for all real numbers x, y, and z. Explicitly, we must have that $f(x,y,z) = x - y + 2z$ because

$$f(x,y,z) = f(1,0,0)x + f(0,1,0)y + f(0,0,1)z = x - y + 2z.$$

Each of the real numbers x, y, and z can be factored out from f by the second property of a linear functional above; the first property above yields that $f(x,y,z) = f(x,0,0) + f(0,y,0) + f(0,0,z)$.

We refer to the collection $V^* = \{f : V \to k \mid f \text{ is linear}\}$ of all linear functionals from a k-vector space V to its field k of scalars as the **dual** of V (or the **dual space** of V). Crucially, we note that V^* is a k-vector space with respect to function addition and pointwise scalar multiplication.

Proposition 3.6.6. *Given any k-vector space V, the dual space V^* is a k-vector space with respect to function addition $(f+g)(v) = f(v) + g(v)$ and pointwise scalar multiplication $(\alpha f)(v) = \alpha f(v)$.*

Proof. Considering that V and k are both k-vector spaces, this holds by Proposition 1.10.9 □

Clearly, it is advantageous to view the dual space V^* of V as a vector space over the same field of scalars because this vantage point affords us all of the tools that we developed in Chapter 1 to study V^*. We put these techniques to immediate use in the following fundamental proposition.

Proposition 3.6.7. *Consider any finite-dimensional k-vector space V. Given any basis $v_1, \dots, v_n$ of V, there exists a unique basis $f_1, \dots, f_n$ of V^* called the* **dual basis** *such that $f_i(v_j) = \delta_{ij}$ is the Kronecker delta. Even more, any linear functional f on V satisfies that $f = f(v_1)f_1 + \cdots + f(v_n)f_n$.*

Proof. Every vector of V can be written as $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ for some unique scalars $\alpha_1, \dots, \alpha_n$ by assumption that $v_1, \dots, v_n$ constitute a basis for V. Considering that every linear functional $f : V \to k$ must satisfy that $f(v) = f(\alpha_1 v_1) + \cdots + f(\alpha_n v_n) = \alpha_1 f(v_1) + \cdots + \alpha_n f(v_n)$, it follows that every element f of V^* is uniquely determined by $f(v_1), \dots, f(v_n)$ (cf. Proposition 1.11.17). Consequently, we may define unique linear functionals $f_1, \dots, f_n$ of V^* by declaring that $f_i(v_j) = \delta_{ij}$ and **extending linearly** to determine the image of f_i on any vector of V. Explicitly, we must have that $f_i(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 f_i(v_1) + \cdots + \alpha_n f_i(v_n) = \alpha_i$ because $f_i(v_j) = 0$ for all indices $i \neq j$ and $f_i(v_i) = 1$. We must next demonstrate that $f_1, \dots, f_n$ span V^* and that they are linearly independent. We will assume first that there exist scalars $\alpha_1, \dots, \alpha_n$ such that $\alpha_1 f_1 + \cdots + \alpha_n f_n$ is the zero functional. Consequently, for each integer $1 \leq i \leq n$, we have that

$$0 = (\alpha_1 f_1 + \cdots + \alpha_n f_n)(v_i) = \alpha_1 f_1(v_i) + \cdots + \alpha_n f_n(v_i) = \alpha_i$$

so that $f_1, \dots, f_n$ are linearly independent. Given any linear functional $f : V \to k$, we have already seen that $f(v) = \alpha_1 f(v_1) + \cdots + \alpha_n f(v_n)$ and $\alpha_i = f_i(\alpha_1 v_1 + \cdots + \alpha_n v_n) = f_i(v)$. Combined, these two formative observations yield the following identity for all vectors v of V.

$$f(v) = \alpha_1 f(v_1) + \cdots + \alpha_n f(v_n) = f(v_1) f_1(v) + \cdots + f(v_n) f_n(v) = (f(v_1) f_1 + \cdots + f(v_n) f_n)(v).$$

We conclude that $f = f(v_1) f_1 + \cdots + f(v_n) f_n$ so that the linear functionals $f_1, \dots, f_n$ span V^*. $\square$

Corollary 3.6.8. *Given any finite-dimensional k-vector space V, we have that* $\dim(V^*) = \dim(V)$.

Corollary 3.6.9. *Given any finite-dimensional k-vector space V, every vector of V can be written as $v = f_1(v) v_1 + \cdots + f_n(v) v_n$ for any basis $v_1, \dots, v_n$ of V and the dual basis $f_1, \dots, f_n$ of V^*.*

Proposition 3.6.10. *Every finite-dimensional k-vector space is isomorphic to its dual.*

Proof. We must provide a vector space isomorphism $T : V \to V^*$. Explicitly, we must construct a linear transformation $T : V \to V^*$ that is both injective and surjective. Given any basis $v_1, \dots, v_n$ of V, consider the linear transformation $T : V \to V^*$ uniquely determined by $T(v_i) = f_i$ for the dual basis $f_1, \dots, f_n$ of V^*. By Proposition 1.11.6, in order to demonstrate that T is injective, it suffices to prove that $\ker(T) = \{O_V\}$. Consider any vector v of V such that $T(v)$ is the zero functional. We may write $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ for some unique scalars $\alpha_1, \dots, \alpha_n$. By definition of the linear transformation T, we have that $T(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) = \alpha_1 f_1 + \cdots + \alpha_n f_n$. Considering that $f_1, \dots, f_n$ are linearly independent and $T(v)$ is the zero functional, we must have that $\alpha_1 = \cdots = \alpha_n = 0$ so that $v = O_V$. Every linear functional $f : V \to k$ can be written uniquely as $f = f(v_1) f_1 + \cdots + f(v_n) f_n$, hence the vector $v = f(v_1) v_1 + \cdots + f(v_n) v_n$ of V satisfies that

$$f = f(v_1) f_1 + \cdots + f(v_n) f_n = f(v_1) T(v_1) + \cdots + f(v_n) T(v_n) = T(f(v_1) v_1 + \cdots + f(v_n) v_n) = T(v).$$

We conclude that $T : V \to V^*$ is injective and surjective, so it is a vector space isomorphism. $\square$

Example 3.6.11. Consider the real vector space $\mathbb{R}^2$ consisting of points in the Cartesian plane. We note that the points $(1, 1)$ and $(1, -2)$ form a basis for $\mathbb{R}^2$ because they are linearly dependent: indeed, if there exist real numbers a and b for which $(a + b, a - 2b) = a(1, 1) + b(1, -2) = (0, 0)$, then $a + b = 0$ and $a - 2b = 0$ together yield that $-b = a = 2b$ or $b = 0$ so that $a = 0$. We will find a basis of $(\mathbb{R}^2)^*$ that is dual to the basis $(1, 1)$ and $(1, -2)$ of $\mathbb{R}^2$. By Proposition 3.6.7, we must furnish linear functionals $f_i : \mathbb{R}^2 \to \mathbb{R}$ defined by $f_i(v_j) = \delta_{ij}$ for each pair of integers $1 \leq i, j \leq 2$.

$$\begin{aligned} f_1(1, 1) &= 1 & \qquad f_1(1, -2) &= 0 \\ f_2(1, 1) &= 0 & \qquad f_2(1, -2) &= 1 \end{aligned}$$

Eventually, we wish to determine $f_1(x, y)$ and $f_2(x, y)$. Considering that each of these functions is linear, we have that $f_1(x, y) = f_1(x, 0) + f_1(0, y) = f_1(1, 0)x + f_1(0, 1)y$, so it suffices to find $f_1(1, 0)$ and $f_1(0, 1)$; the same rationale shows that $f_2(x, y) = f_2(x, 0) + f_2(0, y) = f_2(1, 0)x + f_2(0, 1)y$. We must write the vectors $(1, 0)$ and $(0, 1)$ in terms of the basis vectors $(1, 1)$ and $(1, -2)$. Explicitly, we have that $(1, 0) = a(1, 1) + b(1, -2) = (a + b, a - 2b)$ and $(0, 1) = c(1, 1) + d(1, -2) = (c + d, c - 2d)$.

By solving the linear equations $a + b = 1$ and $a - 2b = 0$, we find that $a = \frac{2}{3}$ and $b = \frac{1}{3}$. Likewise, by solving the linear equations $c + d = 0$ and $c - 2d = 1$, we find that $c = \frac{1}{3}$ and $d = -\frac{1}{3}$.

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

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

Considering that $f_1(1, 1) = 1$ and $f_1(1, -2) = 0$, it follows that $f_1(1, 0) = \frac{2}{3}$ and $f_1(0, 1) = \frac{1}{3}$. Even more, we have that $f_2(1, 1) = \frac{1}{3}$ and $f_2(0, 1) = -\frac{1}{3}$ so that $f_1(x, y) = \frac{2}{3}x + \frac{1}{3}y$ and $f_2(x, y) = \frac{1}{3}x - \frac{1}{3}y$.

Example 3.6.12. Consider the real vector space $\mathbb{R}^3$ consisting of points in real 3-space. By Algorithm 1.7.9, the points $(1, 0, -1)$, $(1, 1, 1)$, and $(2, 2, 0)$ of $\mathbb{R}^3$ are linearly independent because the real 3×3 matrix whose columns are these three vectors has three pivots as follows.

$$\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ -1 & 1 & 0 \end{bmatrix} \overset{R_3 + R_1 \mapsto R_3}{\sim} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 2 & 2 \end{bmatrix} \overset{R_3 - 2R_2 \mapsto R_3}{\sim} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & -2 \end{bmatrix}$$

By Proposition 3.6.7, dual basis of $(\mathbb{R}^3)^*$ corresponding to $(1, 0, -1)$, $(1, 1, 1)$, $(2, 2, 0)$ consists of the linear functionals $f_i : \mathbb{R}^3 \to \mathbb{R}$ defined by $f_i(v_j) = \delta_{ij}$ for each pair of integers $1 \leq i, j \leq 3$.

$$\begin{array}{lll} f_1(1, 0, -1) = 1 & f_1(1, 1, 1) = 0 & f_1(2, 2, 0) = 0 \\ f_2(1, 0, -1) = 0 & f_2(1, 1, 1) = 1 & f_2(2, 2, 0) = 0 \\ f_3(1, 0, -1) = 0 & f_3(1, 1, 1) = 0 & f_3(2, 2, 0) = 1 \end{array}$$

By taking inspiration from the exposition of Example 3.6.11, we may find real numbers such that

$$(1, 0, 0) = a(1, 0, -1) + b(1, 1, 1) + c(2, 2, 0) = (a + b + 2c, b + 2c, -a + b),$$

$$(0, 1, 0) = d(1, 0, -1) + e(1, 1, 1) + f(2, 2, 0) = (d + e + 2f, e + 2f, -d + e), \text{ and}$$

$$(0, 0, 1) = g(1, 0, -1) + h(1, 1, 1) + i(2, 2, 0) = (g + h + 2i, h + 2i, -g + h).$$

We leave it as an exercise for the reader to verify that $a = b = f = h = 1$, $c = i = -\frac{1}{2}$, $d = e = -1$, and $g = 0$. Consequently, we may write $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ as follows.

$$\begin{aligned} (1, 0, 0) &= (1, 0, -1) + (1, 1, 1) - \frac{1}{2}(2, 2, 0) \\ (0, 1, 0) &= -(1, 0, -1) - (1, 1, 1) + (2, 2, 0) \\ (0, 0, 1) &= 0(1, 0, -1) + (1, 1, 1) - \frac{1}{2}(2, 2, 0) \end{aligned}$$

Example 3.6.11 implies that the coefficients of $x, y,$ and z in $f_1(x, y, z)$ can be read from the first column above; the coefficients of $x, y,$ and z in $f_2(x, y, z)$ can be read from the second column above; and the coefficients of x, y, and z in $f_3(x, y, z)$ can be read from the third column above. We conclude that $f_1(x, y, z) = x - y$, $f_2(x, y, z) = x - y + z$, and $f_3(x, y, z) = -\frac{1}{2}x + y - \frac{1}{2}z$.

Even more, the linear functional $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(x, y, z) = 2x - y + z$ satisfies that $f(1, 0, -1) = 1$, $f(1, 1, 1) = 2$, and $f(2, 2, 0) = 2$ so that $f = f_1 + 2f_2 + 2f_3$ by Proposition 3.6.7.

Last, we will express the vectors $(2, -2, 1)$ and $(0, -2, -3)$ of $\mathbb{R}^3$ in terms of the basis vectors $(1, 0, -1)$, $(1, 1, 1)$, and $(2, 2, 0)$ using only the dual basis f_1, f_2, f_3 of $(\mathbb{R}^3)^*$. We note the following.

$$f_1(2, -2, 1) = 2 - (-2) = 4 \qquad f_1(0, -2, -3) = (0) - (-2) = 2$$

$$f_2(2, -2, 1) = 2 - (-2) + 1 = 5 \qquad f_2(0, -2, -3) = (0) - (-2) + (-3) = -1$$

$$f_3(2, -2, 1) = -\frac{1}{2}(2) + (-2) - \frac{1}{2}(1) = -\frac{7}{2} \qquad f_3(0, -2, -3) = -\frac{1}{2}(0) + (-2) - \frac{1}{2}(-3) = -\frac{1}{2}$$

$$(2, -2, 1) = 4(1, 0, -1) + 5(1, 1, 1) - \frac{7}{2}(2, 2, 0) \qquad (0, -2, -3) = 2(1, 0, -1) - (1, 1, 1) - \frac{1}{2}(2, 2, 0)$$

Consequently, linear functionals are quite useful — especially the dual basis for a vector space. Entire fields of mathematics are devoted to the study of linear functionals that satisfy additional analytic or geometric properties. Explicitly, in mathematical analysis, much of the work in the areas of functional analysis and numerical analysis takes place in a suitable vector space for such computations (such as a Hilbert space or a Banach space). Even though we will not discuss these concepts, we conclude this section with an astonishing result that states that over a positive-definite inner product space, every linear functional is simply the inner product with a fixed vector.

Theorem 3.6.13. *Consider any finite-dimensional k-vector space V with a positive-definite inner product $\langle -, - \rangle$. Given any linear functional $f : V \to k$, there exists a unique vector $w \in V$ such that $f(v) = \langle v, w \rangle$ for all vectors $v \in V$. Put another way, every linear functional on V is determined uniquely by the inner products of any basis vectors of V with some fixed vector of V.*

Proof. Choose an orthonormal basis $u_1, \dots, u_n$ for V by the Gram-Schmidt Process. Consider the vector $w = f(u_1)u_1 + \cdots + f(u_n)u_n$. We claim that $f_w : V \to k$ defined by $f_w(v) = \langle v, w \rangle$ is a linear functional. Explicitly, by the first and second properties of an inner product, for any vectors u and v of V, we have that $f_w(u+v) = \langle u+v, w \rangle = \langle u, w \rangle + \langle v, w \rangle = f_w(u) + f_w(v)$. Even more, by the third property of an inner product, we have that $f_w(\alpha v) = \langle \alpha v, w \rangle = \alpha \langle v, w = \alpha f_w(v)$ for any scalar α and any vector v. We conclude that f_w is a linear functional, hence it suffices to prove that $f(u_i) = f_w(u_i)$ for all basis vectors u_i by Proposition 1.11.17. By definition of f_w, we have that

$$f_w(u_i) = \langle u_i, w \rangle = \langle u_i, f(u_1)u_1 + \cdots + f(u_n)u_n \rangle = f(u_1)\langle u_i, u_1 \rangle + \cdots + f(u_n)\langle u_i, u_n \rangle = f(u_i)$$

by assumption that $u_1, \dots, u_n$ are orthonormal. Explicitly, we have that $\langle u_i, u_j \rangle = 0$ for all indices $i \neq j$ and $\langle u_i, u_i \rangle = 1$. Last, we prove the uniqueness of w: if $\langle v, w \rangle = \langle v, u \rangle$ for all vectors of V for some vector u, then $\langle w - u, w \rangle = \langle w - u, u \rangle$ yields that $\langle w - u, w - u \rangle = 0$ so that $w - u = O$. $\square$

Example 3.6.14. Consider the orthonormal basis $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ of $\mathbb{R}^3$. Observe that for the linear functional $f : \mathbb{R}^3 \to \mathbb{R}$ of Example 3.6.5, we have that $f(1, 0, 0) = 1$, $f(0, 1, 0) = -1$, and

$f(0,0,1) = 2$, hence $w = (1,0,0) - (0,1,0) + 2(0,0,1) = (1,-1,2)$ determines the linear functional f as an inner product. We note that $f(x,y,z) = x - y + 2z = (x,y,z) \cdot (1,-1,2)$, as desired.

Example 3.6.15. Consider the orthonormal basis consisting of $3t$ and $80(t^2 - \frac{3}{4}t)$ for the subspace $\operatorname{span}\{t, t^2\}$ of $\mathcal{C}^0(\mathbb{R})$ with respect to the inner product $\langle f, g \rangle = \int_0^1 f(t)g(t)\,dt$ (cf. Example 3.5.6). Observe that the linear functional $E : \mathcal{C}^0(\mathbb{R}) \to \mathbb{R}$ defined by $E(f(t)) = f(1)$ satisfies that $E(3t) = 3$ and $E(80(t^2 - \frac{3}{4}t)) = 20$, hence $f(t) = 3(3t) + 20(80(t^2 - \frac{3}{4}t)) = 1600t^2 - 1191t$ uniquely determines E as an inner product. Explicitly, one can verify that $E(t) = 3 = \int_0^1 (1600t^3 - 1191t^2)\,dt$, as desired.

References

[Con22] K. Conrad. *The Minimal Polynomial and Some Applications*. 2022. URL: https://kconrad.math.uconn.edu/blurbs/linmultialg/minpolyandappns.pdf.

[DF04] D.S. Dummit and R.M. Foote. *Abstract Algebra*. 3rd ed. John Wiley & Sons, Inc., 2004.

[FB95] J.B. Fraleigh and R.A. Beauregard. *Linear Algebra*. 3rd ed. Addison-Wesley Publishing Company, 1995.

[HK71] K. Hoffman and R. Kunze. *Linear Algebra*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1971.

[Lan86] S. Lang. *Introduction to Linear Algebra*. 2nd ed. Undergraduate Texts in Mathematics. Springer-Verlag New York, Inc., 1986.

[Lan87] S. Lang. *Introduction to Linear Algebra*. 3rd ed. Undergraduate Texts in Mathematics. Springer-Verlag New York, Inc., 1987.

[McK22] J. McKinno. *The Principal Axis Theorem*. 2022. URL: https://www.math.uwaterloo.ca/~jmckinno/Math225/Week7/Lecture2m.pdf.

[Moo68] J.T. Moore. *Elements of Linear Algebra and Matrix Theory*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Company, 1968.

[Smi17] K.E. Smith. *The Spectral Theorem*. 2017. URL: http://www.math.lsa.umich.edu/~kesmith/SpectralTheoremW2017.pdf.

[Str06] G. Strang. *Linear Algebra and Its Applications*. 4th ed. Cengage Learning, 2006.