

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 — especially those who assisted in the enhancement of these notes with comments and suggestions, including Ben Stubblefield.

Contents

1	Matrices and Vector Spaces	7
1.1	Matrices and Matrix Addition	7
1.2	Rotation Matrices and Matrix Multiplication	10
1.3	Elementary Row and Column Operations	14
1.4	The Method of Gaussian Elimination in Linear Systems	17
1.5	Invertible Matrices	21
1.6	Vector Spaces	26
1.7	Span and Linear Independence	31
1.8	Vector Space Dimension	34
1.9	Matrix Rank	37
1.10	Linear Transformations	41
1.11	Kernels and Images of Linear Transformations	44
1.12	The Rank-Nullity Theorem	49
1.13	Composition and Inversion of Linear Transformations	51
1.14	Matrix Representations of Linear Transformations	55
1.15	Chapter 1 Overview	63
2	Canonical Forms for Matrices	65
2.1	Determinants of 3×3 Matrices	65
2.2	The Adjugate of a Matrix	70
	References	76

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 i th row and j th 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 $[x]$.

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 $[1 \ 2 \ 3 \ 4 \ 5]$ 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 i th row and j th 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 j th column consists of the integer j for each integer $1 \leq j \leq 5$. Under this identification, we will adopt the one-line notation $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ for the $m \times n$ matrix A with a_{ij} in its i th row and j th column.

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

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

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 $I_{m \times n} = [\delta_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ 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 $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, 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 $A^t = [a_{ji}]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$. Put another way, the (i, j) th entry of A^t is the (j, i) th entry of A , hence the i th row of A^t is precisely the i th 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 $I_{m \times n}^t = I_{n \times m}$, hence we have that $I_{n \times n}^t = I_{n \times n}$.

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 $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ and $B = [b_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, the **matrix sum** of A and B is the $m \times n$ matrix $A + B = [a_{ij} + b_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$. 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 $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, 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 $cA = [ca_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$. 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 $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, we will write $-A = [-a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$. 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 $(A + B)^t = A^t + B^t$. \square

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 $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ and any $n \times r$ matrix $B = [a_{ij}]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq r}}$, 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 $v = [v_{11} \ v_{12} \ \cdots \ v_{1n}]$ 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.,

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

Given any $m \times n$ matrix A and any $n \times r$ matrix B , the i th row of A may be viewed as the $1 \times n$ vector $A_i = [a_{i1} \ a_{i2} \ \cdots \ a_{in}]$ and the j th 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 $1 \leq \ell \leq r$, 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) th entry of $B + C$ 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 $(AB)^t = B^t A^t$. 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 entry of 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 $B^t A^t$. We conclude therefore that $(AB)^t = B^t A^t$. \square

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.

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

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 $a_1x_1 + \dots + a_nx_n$ for some (real) scalars a_1, \dots, a_n . Consequently, a (real) $1 \times n$ **linear equation** is any equation of the form $a_1x_1 + \dots + a_nx_n = b$ 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.

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

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 ξ_1, \dots, ξ_n constitute a **solution** to a (real) $m \times n$ system of linear equations if it holds that $a_{i1}\xi_1 + \dots + a_{in}\xi_n = b_i$ 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 $x = [x_1 \ x_2 \ \cdots \ x_n]$, a $1 \times m$ matrix $b = [b_1 \ b_2 \ \cdots \ b_m]$, and an $m \times n$ matrix A whose (i, j) th entry is the coefficient a_{ij} of the j th variable x_j of the i th 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 $A = [1 \ 2 \ 3]$ whose j th column is the point value of a j -pointer; the matrix $x = [x_1 \ x_2 \ x_3]$ whose j th column is the number of j -pointers made by Curry; and the matrix $b = [43]$ 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 $\xi_1 = [43 \ 0 \ 0]$ and $\xi_2 = [41 \ 1 \ 0]$ 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 $[1 \ 2]$, 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 $[1 \ 1]$, 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 i th equation by a nonzero (real) scalar c .
- (2.) We may add c times the i th equation to the j th equation for all integers $1 \leq i, j \leq m$.
- (3.) We may interchange the i th and j th 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 i th equation are encoded in the i th row of the matrix A and the i th row of the matrix b^t , we must henceforth work with the **augmented matrix** $[A \mid b^t]$. By definition, this is simply the matrix A with one additional column in the form of b^t . We use the bar $|$ 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 i th 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 $[A \mid b^t]$ 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.

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

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.

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

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 i th row of an $m \times n$ system of linear equations to the j th row corresponds to left multiplication of the $m \times (n+1)$ matrix $[A \mid b^t]$ 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 $[A \mid b^t]$ 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.

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

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 $[A \mid b^t]$ 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 $[A \mid b^t]$ 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.

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

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 i th row of I_m to the j th 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 $[A \mid b^t]$ 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 i th row of the matrix do not appear in the i th 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 i th 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. \square

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 j th 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 j th column of C contains a pivot in row k , then we may add $-c_{ij}$ times the k th row of C to the i th row of C to obtain 0 in the i th row and j th column of C . Continuing in this manner yields the result. \square

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 i th 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 j th row and first column of the matrix obtained in step (3.), then add $-r_j$ times the first row of this matrix to the j th 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 k th row of the matrix of step (6.) to the i th 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** $\text{RREF}(A)$.

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.

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

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

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

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.

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

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.

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

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.

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

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.

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

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

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

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 $L' = L$. \square

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 $(A^{-1})^t A^t = (AA^{-1})^t = I_{n \times n}^t = I_{n \times n}$, and we conclude that $(A^t)^{-1} = (A^{-1})^t$ by the uniqueness of the matrix inverse guaranteed by Proposition 1.5.1. \square

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. \square

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 i th row of $I_{n \times n}$ to the j th 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 i th row of $I_{n \times n}$ by a nonzero scalar c , then E^{-1} is obtained from $I_{n \times n}$ by multiplying the i th 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 i th row of $I_{n \times n}$ to the j th row of $I_{n \times n}$, then E^{-1} is obtained from $I_{n \times n}$ by adding $-c$ times the i th row of $I_{n \times n}$ to the j th 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. \square

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: $I_{n \times n} = B(E_k \cdots E_1 A)^{-1} = B A^{-1} E_1^{-1} \cdots E_k^{-1}$. \square

By Corollary 1.4.5, every $n \times n$ matrix A is row equivalent to its reduced row echelon form $\text{RREF}(A)$. Consequently, by the previous corollary, it follows that A is invertible if and only if $\text{RREF}(A)$ is invertible. Particularly, if $\text{RREF}(A)$ admits any rows consisting entirely of zeros, then it is not invertible (because the last row of $\text{RREF}(A)B$ is zero for all $n \times r$ matrices B), hence A cannot be invertible. Conversely, we will demonstrate that if all rows $\text{RREF}(A)$ 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 i th 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 i th 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 n th 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 $I_{n \times n}$. \square

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 $\text{RREF}(A)$ is invertible if and only if $\text{RREF}(A) = I_{n \times n}$. 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.

$$A \xrightarrow{a^{-1}R_1 \rightarrow R_1} \begin{bmatrix} 1 & a^{-1}b \\ c & d \end{bmatrix}$$

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.

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

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.

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

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 $[A \mid I_{n \times n}]$ 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 $[I_{n \times n} \mid A^{-1}]$.

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

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

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

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

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

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

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 $|$ 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 $(\alpha + \beta)u = \alpha u + \beta u$ for all (real) scalars α and β and each element $u \in V$.

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 $f : \mathbb{R} \rightarrow \mathbb{R}$. We may define function addition so that if $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$ 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$, 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 $O : \mathbb{R} \rightarrow \mathbb{R}$ defined by $O(x) = 0$ for all real numbers x . Given any function $f : \mathbb{R} \rightarrow \mathbb{R}$, 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 $f : \mathbb{R} \rightarrow \mathbb{R}$, we may define the function $-f : \mathbb{R} \rightarrow \mathbb{R}$ 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ 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 $f : \mathbb{R} \rightarrow \mathbb{R}$, we have that $(1f)(x) = 1f(x) = f(x)$ for all real numbers x .
- (8.) Given any function $f : \mathbb{R} \rightarrow \mathbb{R}$, 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$, 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 $f : \mathbb{R} \rightarrow \mathbb{R}$, 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 $0v = 0v + O_V = 0v + [0v + (-0v)] = (0v + 0v) + (-0v) = 0v + (-0v) = O_V$. \square

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 $w + (-w) = O_V$. \square

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 $f : \mathbb{R} \rightarrow \mathbb{R}$ and its subset $\mathcal{C}^1(\mathbb{R})$ of functions $f : \mathbb{R} \rightarrow \mathbb{R}$ such that f' is continuous. Clearly, the zero function $O : \mathbb{R} \rightarrow \mathbb{R}$ 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ such that f' is continuous. Consider the set $W = \{f \in \mathcal{C}^1(\mathbb{R}) \mid f(0) = 0\}$. Clearly, the zero function $O : \mathbb{R} \rightarrow \mathbb{R}$ 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 : although the $n \times n$ zero matrix $O_{n \times n}$ 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. \square

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 $v + w = 1v + 1w \in W$ and $\alpha w = 0v + \alpha w \in W$; we are done by the Two-Step Subspace Test. \square

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. \square

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 $\alpha_1 v_1 + \dots + \alpha_n v_n$ 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 $\text{span}\{v_1, \dots, v_n\}$ to denote this set. Explicitly, an element of $\text{span}\{v_1, \dots, v_n\}$ is of the form $\alpha_1 v_1 + \dots + \alpha_n v_n$ 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 $V = \text{span}\{v_1, \dots, v_n\}$.

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 $x_1 E_1 + \dots + x_n E_n$ for some real numbers x_1, \dots, x_n , where E_i is the $1 \times n$ row vector consisting of 1 in the i th column and zeros elsewhere. Consequently, it follows that $\mathbb{R}^{1 \times n} = \text{span}\{E_1, \dots, E_n\}$.

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 $\mathbb{R}^{2 \times 2} = \text{span}\{E_{11}, E_{12}, E_{21}, E_{22}\}$.

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 $f(x) = a_n x^n + \dots + a_1 x + a_0$ for some real numbers a_0, a_1, \dots, a_n , hence we conclude that $P_n(x) = \text{span}\{1, x, \dots, x^n\}$.

We say that a collection of vectors v_1, \dots, v_n are **linearly independent** whenever it holds that $\alpha_1 v_1 + \dots + \alpha_n v_n = O_V$ implies that $\alpha_1 = \dots = \alpha_n = 0$, 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 $\alpha_1 v_1 + \dots + \alpha_n v_n = 0$, 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 $\alpha_i v_i = -\alpha_1 v_1 - \dots - \alpha_n v_n$ and $v_i = -\alpha_1 \alpha_i^{-1} v_1 - \dots - \alpha_n \alpha_i^{-1} v_n$, 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 i th 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 $a_nx^n + \dots + a_1x + a_0 = 0$, then all of the derivatives of this polynomial are zero. Particularly, the n th 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 $a_0 = a_1 = \dots = a_n = 0$. 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.

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

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 j th 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 j th column of the unique reduced row echelon form $\text{RREF}(A)$ of the $m \times n$ matrix A , or there is not. By definition of the reduced row echelon form, if the j th column of $\text{RREF}(A)$ 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 $\text{RREF}(A)$ is nonzero, there exists an integer $1 \leq k \leq j$ such that the (i, k) th component of $\text{RREF}(A)$ is a pivot of 1. Consequently, the j th column of $\text{RREF}(A)$ 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 j th 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.

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

- (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.) $V = \text{span}\{v_1, \dots, v_n\}$, i.e., V is spanned by v_1, \dots, v_n and
- (2.) v_1, \dots, v_n are linearly independent, i.e., $\alpha_1 v_1 + \dots + \alpha_n v_n = 0$ if and only if $\alpha_1 = \dots = \alpha_n = 0$.

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 i th 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 $v = \alpha_1 v_1 + \dots + \alpha_n v_n$. 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 $x_1 E_1 + \dots + x_n E_n$; 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 $\alpha_1 v_1 + \dots + \alpha_n v_n = \beta_1 v_1 + \dots + \beta_n v_n$, 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 $\alpha_1 v_1 + \dots + \alpha_n v_n = \beta_1 v_1 + \dots + \beta_n v_n$, then subtracting $\beta_1 v_1 + \dots + \beta_n v_n$ from both sides and combining like terms gives $(\alpha_1 - \beta_1)v_1 + \dots + (\alpha_n - \beta_n)v_n = O_V$. 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.

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

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.

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

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 β_1, \dots, β_m such that $\beta_1 w_1 + \dots + \beta_m w_m = O_V$. \square

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. \square

Example 1.8.3. By Example 1.7.11, the real $1 \times n$ matrices E_i consisting of 1 in the i th 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 $\alpha_1 v_1 + \dots + \alpha_m v_m + \alpha_k v_k = O_V$. Observe that if $\alpha_k = 0$, then $\alpha_1 = \dots = \alpha_m = 0$ 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 $v_k = -\alpha_1\alpha_k^{-1}v_1 - \dots - \alpha_m\alpha_k^{-1}v_m$. 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 $v = \alpha_1v_1 + \dots + \alpha_nv_n$. 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 . \square

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 $\alpha_1v_1 + \dots + \alpha_nv_n + \alpha v = O_V$ 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 $\alpha_1 = \dots = \alpha_n = 0$. Consequently, we must have that α is nonzero so that $v = \alpha_1\alpha^{-1}v_1 + \dots + \alpha_n\alpha^{-1}v_n$. We conclude that $V = \text{span}\{v_1, \dots, v_n\}$. \square

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. \square

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. \square

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 $a_n x^n + \dots + a_1 x + a_0$ 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 j th 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 $\text{span}\{A_1, \dots, A_n\} = \{c_1 A_1 + \dots + c_n A_n \mid c_1, \dots, c_n \text{ are scalars}\}$. We will refer to the vector space $\text{span}\{A_1, \dots, A_n\}$ as the **column space** of A ; the dimension of $\text{span}\{A_1, \dots, A_n\}$ 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 $Ac^t = c_1 A_1 + \dots + c_n A_n$.

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 $\text{span}\{E_1^t, E_2^t, E_3^t\} = \{\alpha_1 E_1^t + \alpha_2 E_2^t + \alpha_3 E_3^t \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}\} = \mathbb{R}^{3 \times 1}$ 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.

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

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

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

for A has column space $\text{span}\{(1, 0)^t\} = \{(\alpha, 0)^t \mid \alpha \in \mathbb{R}\}$, 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 $c_1 R_1 + \dots + c_n R_n = O$, then multiplying both sides of this vector equation on the left by E yields that $c_1 A_1 + \dots + c_n A_n = O$. Conversely, if there exist scalars d_1, \dots, d_n such that $d_1 A_1 + \dots + d_n A_n = O$, then multiplying both sides of this vector equation on the left by E^{-1} yields that $d_1 R_1 + \dots + d_n R_n = O$. 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 i th 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 $\text{span}\{a_1, \dots, a_m\} = \{c_1 a_1 + \dots + c_m a_m \mid c_1, \dots, c_m \text{ are scalars}\}$.

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 , E_2 , and E_3 such that $E_1 = (1, 0, 0)$, $E_2 = (0, 1, 0)$, 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.

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

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

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

Likewise, the row rank of $\text{RREF}(A)$ 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 if $c_1 = c_2 = c_3 = 0$. Last, the row space of $\text{RREF}(A)$ is

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

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

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

Consequently, A and $\text{RREF}(A)$ 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 $c_1 a_1 + \dots + c_m a_m$ in the span of a_1, \dots, a_m is now given by $c_1 a_1 + \dots + (c_1 c^{-1}) c a_i + \dots + c_m a_m$.

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 $c_1a_1 + \dots + c_ma_m$ in the span of a_1, \dots, a_m can be achieved as $c_1a_1 + \dots + (c_i - c_jc)a_i + \dots + c_j(ca_i + a_j) + \dots + c_ma_m$. 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 $\text{span}\{a_1, \dots, a_m\} = \text{span}\{r_1, \dots, r_m\}$, 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. \square

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. \square

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 $C_j - aC_i \mapsto C_j$ 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 . \square

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.

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

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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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. \square

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 $T : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{n \times m}$ 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 $T_A : \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{m \times r}$ 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 $T_A : \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{m \times r}$ 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 $T : \mathbb{R}^{1 \times 3} \rightarrow \mathbb{R}^{1 \times 2}$ 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 $D : P_1(x) \rightarrow P_1(x)$ defined by $D(mx + b) = m$. Explicitly, the function D maps a polynomial to its first derivative. Observe that

$$\alpha(m_1x + b_1) + (m_2x + b_2) = \alpha m_1x + \alpha b_1 + m_2x + b_2 = (\alpha m_1 + m_2)x + (\alpha b_1 + b_2),$$

and the derivative of this function is $\alpha m_1 + m_2 = \alpha D(m_1x + b_1) + D(m_2x + b_2)$. 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 $\det : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}$ 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 $f : \mathbb{R} \rightarrow \mathbb{R}$. 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 $A : \mathcal{C}^0(\mathbb{R}) \rightarrow \mathcal{C}^1(\mathbb{R})$ 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 $R_a : \mathcal{C}^0(\mathbb{R}) \rightarrow \mathcal{C}^1(\mathbb{R})$ 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

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

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

Proposition 1.10.8. *Let $T : V \rightarrow W$ 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 $T(O_V) = O_W$. \square

By Example 1.6.7, the collection of real functions $f : \mathbb{R} \rightarrow \mathbb{R}$ 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 $f : \mathbb{R} \rightarrow \mathbb{R}$ 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., $\mathcal{L}(V, W) = \{T : V \rightarrow W \mid T \text{ is a linear transformation}\}$. 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 $S : V \rightarrow W$ and $T : V \rightarrow W$ are linear transformations, then $S + T : V \rightarrow W$ 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 $S + T : V \rightarrow W$ is a linear transformation.

- (4.) Consider the function $O : V \rightarrow W$ 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 $T : V \rightarrow W$, consider the function $-T : V \rightarrow W$. 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 $T : V \rightarrow W$ is any linear transformation, then the function $\alpha T : V \rightarrow W$ 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. \square

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 $T : \mathbb{R} \rightarrow \mathbb{R}$. Consequently, the elements of $\mathcal{L}(\mathbb{R}, \mathbb{R})$ are functions $T : \mathbb{R} \rightarrow \mathbb{R}$ 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 $T : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{n \times m}$ 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 $T_A : \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{m \times r}$ 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 $\ker(T_A) = \text{span}\{E_{11} + E_{21}, E_{12} + E_{22}\}$.

Example 1.11.4. Consider the linear transformation $T : \mathbb{R}^{1 \times 3} \rightarrow \mathbb{R}^{1 \times 2}$ 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

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

Example 1.11.5. Consider the differentiation transformation $D : P_1(x) \rightarrow P_1(x)$ 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 $T : V \rightarrow W$ with $\ker(T) = \{O_V\}$. We will say that $T : V \rightarrow W$ is **injective** if and only if $T(v_1) = T(v_2)$ implies that $v_1 = v_2$.

Proposition 1.11.6. *If $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 $\alpha_1 T(v_1) + \dots + \alpha_n T(v_n) = O_W$. By the first part of Proposition 1.10.8, we have that

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

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

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

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

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 $T : V \rightarrow W$ is a linear transformation of vector spaces V and W , then the range $\text{range}(T) = \{T(v) \mid v \in V\}$ 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 $O_W \in \text{range}(T)$.
- (2.) Consider any vectors $w, x \in \text{range}(T)$. By definition of $\text{range}(T)$, 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 $w + x \in \text{range}(T)$.
- (3.) Last, for any vector $w \in \text{range}(T)$ 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 $\text{range}(T)$. \square

Example 1.11.11. Consider the linear transformation $T : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{n \times m}$ of Example 1.10.2 defined by $T(A) = A^t$. By definition, we have that $\text{range}(T) = \{A^t \mid A \in \mathbb{R}^{m \times n}\}$. 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 $\text{range}(T) = \mathbb{R}^{n \times m}$.

Example 1.11.12. Given any real $m \times n$ matrix A , as in Example 1.10.3, we may define a linear transformation $T_A : \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{m \times r}$. We have that $\text{range}(T_A) = \{AB \mid B \in \mathbb{R}^{n \times r}\}$. 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 $\text{range}(T_A) = \mathbb{R}^{n \times r}$.

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

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

We will say that a linear transformation $T : V \rightarrow W$ is **surjective** if it holds that $\text{range}(T) = W$. 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 $T : V \rightarrow W$ is a linear transformation of finite-dimensional vector spaces V and W , then the following statements are equivalent.*

- 1.) *If $V = \text{span}\{v_1, \dots, v_n\}$, then $W = \text{span}\{T(v_1), \dots, T(v_n)\}$.*
- 2.) *We have that $\text{range}(T) = W$, 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 $V = \text{span}\{v_1, \dots, v_n\}$. Consequently, if the first statement of the proposition holds, then $W = \text{span}\{T(v_1), \dots, T(v_n)\}$. By definition, for every vector $w \in W$, there exist scalars $\alpha_1, \dots, \alpha_n$ such that $w = \alpha_1 T(v_1) + \dots + \alpha_n T(v_n) = T(\alpha_1 v_1 + \dots + \alpha_n v_n)$. Considering that $\alpha_1 v_1 + \dots + \alpha_n v_n$ lies in V , we conclude that $\text{range}(T) = W$, 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 $V = \text{span}\{v_1, \dots, v_n\}$, there exist scalars $\alpha_1, \dots, \alpha_n$ such that $v = \alpha_1 v_1 + \dots + \alpha_n v_n$ and $w = T(v) = T(\alpha_1 v_1 + \dots + \alpha_n v_n) = \alpha_1 T(v_1) + \dots + \alpha_n T(v_n)$, i.e., $W = \text{span}\{T(v_1), \dots, T(v_n)\}$. \square

We demonstrate next that linear transformations preserve vector subspaces.

Proposition 1.11.16. *Let $T : V \rightarrow W$ 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 $T^{-1}(U)$. \square

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 $T : V \rightarrow W$ 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 $v = \alpha_1 v_1 + \dots + \alpha_n v_n$ 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 $S : V \rightarrow W$ and $T : V \rightarrow W$ 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 $v = \alpha_1 v_1 + \dots + \alpha_n v_n$ 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 $S(\alpha_1 v_1 + \dots + \alpha_n v_n) = \alpha_1 S(v_1) + \dots + \alpha_n S(v_n) = \alpha_1 T(v_1) + \dots + \alpha_n T(v_n) = T(\alpha_1 v_1 + \dots + \alpha_n v_n)$. \square

1.12 The Rank-Nullity Theorem

Given any linear transformation $T : V \rightarrow W$ of vector spaces V and W , we obtain two vector spaces $\ker(T) = \{v \in V \mid T(v) = O_W\} \subseteq V$ and $\text{range}(T) = \{T(v) \mid v \in V\} \subseteq W$ 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 $\text{range}(T)$ 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ is to find the dimensions of its kernel and range, i.e., the **nullity** $\text{nullity}(T) = \dim(\ker(T))$ and the **rank** $\text{rank}(T) = \dim(\text{range}(T))$ 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 $T : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{n \times m}$ of Examples 1.11.2 and 1.11.11 defined by $T(A) = A^t$. Previously, we demonstrated that $\text{range}(T) = \mathbb{R}^{n \times m}$, hence we have that $\text{rank}(T) = \dim(\mathbb{R}^{n \times m}) = mn = \dim(\mathbb{R}^{m \times n})$. On the other hand, we have that $\ker(T) = \{O_{m \times n}\}$ so that $\text{nullity}(T) = 0$ and $\dim(\mathbb{R}^{m \times n}) = \text{rank}(T) + \text{nullity}(T)$.

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 $T_A : \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{m \times r}$. Like before, if we assume that $m = n$ and A is an invertible real $n \times n$ matrix, then $\text{range}(T_A) = \mathbb{R}^{n \times r}$ so that $\text{rank}(T_A) = nr$ and $\text{nullity}(T_A) = 0$. Once again, in this case, we have that $\dim(\mathbb{R}^{n \times r}) = nr = \text{rank}(T_A) + \text{nullity}(T_A)$.

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

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

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

Proof. Given any vectors $v_1, \dots, v_n \in V$, if $\alpha_1, \dots, \alpha_n$ are scalars such that $\alpha_1 v_1 + \dots + \alpha_n v_n = O_V$, then $O_W = T(O_V) = T(\alpha_1 v_1 + \dots + \alpha_n v_n) = \alpha_1 T(v_1) + \dots + \alpha_n T(v_n)$. Consequently, if $T(v_1), \dots, T(v_n)$ are linearly independent in W , then we must have that $\alpha_1 = \dots = \alpha_n = 0$, 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 $\text{range}(T)$, then v_1, \dots, v_n are linearly independent in V , hence we conclude that $\text{range}(T) = n \leq \dim(V)$. \square

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

Proof. Observe that if $\ker(T) = \{O_V\}$, i.e., if T is injective, then by Proposition 1.11.9, we have that $\dim(V) \leq \text{rank}(T)$. Conversely, by Proposition 1.12.6, it always holds that $\dim(V) \geq \text{rank}(T)$, hence in this case, we conclude that $\dim(V) = \text{rank}(T) = \text{rank}(T) + 0 = \text{rank}(T) + \text{nullity}(T)$. 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 $\text{range}(T)$. Every vector v of V can be written as

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

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

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

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

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

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

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

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

Corollary 1.12.8. *If $T : V \rightarrow W$ 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 $\text{nullity}(T) = 0$.*
- 3.) *We have that $\text{rank}(T) = \dim(W)$.*

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

1.13 Composition and Inversion of Linear Transformations

Given any linear transformations $S : U \rightarrow V$ and $T : V \rightarrow W$ of vector spaces U , V , and W , we may define the **composite function** $T \circ S : U \rightarrow W$ 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 $S : U \rightarrow V$ and $T : V \rightarrow W$ are linear transformations of vector spaces U , V , and W , then the composite function $T \circ S : U \rightarrow W$ 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 $S : U \rightarrow V$ 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 $(T \circ S)(\alpha u + v) = T(\alpha S(u)) + T(S(v)) = \alpha T(S(u)) + T(S(v)) = \alpha(T \circ S)(u) + (T \circ S)(v)$. \square

Example 1.13.2. Consider the linear transformations $S : \mathbb{R}^{1 \times 3} \rightarrow \mathbb{R}^{1 \times 2}$ and $T : \mathbb{R}^{1 \times 2} \rightarrow \mathbb{R}^{1 \times 1}$ 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 $D : P_3(x) \rightarrow P_2(x)$ 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 $R_0 : P_2(x) \rightarrow P_3(x)$ such that $R_0(ax^2 + bx + c) = \int_0^x (at^2 + bt + c) dt$. Observe that the composite functions $R_0 \circ D : P_3(x) \rightarrow P_3(x)$ and $D \circ R_0 : P_2(x) \rightarrow P_2(x)$ 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 $\ker(R_0 \circ D) = \{d \mid d \in \mathbb{R}\} = \text{span}\{1\}$ and $\text{range}(R_0 \circ D) = \{ax^3 + bx^2 + cx \mid a, b, c \in \mathbb{R}\} = \text{span}\{x, x^2, x^3\}$ so that $4 = \dim(P_3(x)) = \text{rank}(R_0 \circ D) + \text{nullity}(R_0 \circ D)$. On the other hand, we have that $\ker(D \circ R_0) = \{0\}$ and $\text{range}(D \circ R_0) = \{ax^2 + bx + c \mid a, b, c \in \mathbb{R}\} = P_2(x)$.

Corollary 1.13.4. *Composition of linear transformations is not commutative in general. Explicitly, if $S : V \rightarrow W$ and $T : W \rightarrow V$ are linear transformations of vector spaces V and W , then it is not necessarily true that $T \circ S : V \rightarrow V$ and $S \circ T : W \rightarrow W$ 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 $T : V \rightarrow V$ 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 $T : V \rightarrow U$ because the codomain U is a subset of the domain V . Last, we will denote by $I : V \rightarrow V$ the **identity operator** defined by $I(v) = v$ for all vectors $v \in V$. If $T : V \rightarrow W$ is a linear transformation of vector spaces V and W , then we say that $S : W \rightarrow V$ is a **left inverse** of T (or T is a **right inverse** of S) if $S \circ T : V \rightarrow V$ satisfies that $S \circ T = I$.

Proposition 1.13.5. *Let $T : V \rightarrow W$ 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 $S : W \rightarrow V$ 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 \mathcal{B} for V . By Proposition 1.11.9, the images $T(v_i)$ of the basis vectors $v_i \in \mathcal{B}$ form a collection $T(\mathcal{B})$ of linearly independent vectors of W . By Theorem 1.8.10, we may extend $T(\mathcal{B})$ to a basis for W . We define the linear transformation $S : W \rightarrow V$ 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 $\alpha_1, \dots, \alpha_n$ and basis vectors v_1, \dots, 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 $S : W \rightarrow V$, 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 $S : W \rightarrow V$ 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 \mathcal{B} for V gives rise to a spanning set $T(\mathcal{B})$ for W that can be reduced to a basis for W . Consequently, define the linear transformation $S : W \rightarrow V$ 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

$$\begin{aligned} \alpha_1 T(v_1) + \cdots + \alpha_n T(v_n) &= \alpha_1 (T \circ S)(T(v_1)) + \cdots + \alpha_n (T \circ S)(T(v_n)) \\ &= (T \circ S)(\alpha_1 T(v_1) + \cdots + \alpha_n T(v_n)). \end{aligned}$$

Last, if we assume that T admits a right inverse $S : W \rightarrow V$, 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 $\text{range}(T) = W$, hence T is surjective. \square

We say that a linear transformation $T : V \rightarrow W$ admits a (two-sided) **inverse** transformation $S : W \rightarrow V$ 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 $T : V \rightarrow W$ 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 $L : W \rightarrow V$ and $R : W \rightarrow V$ 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 $T : V \rightarrow W$ 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 $T^{-1} : W \rightarrow V$. 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 $T : V \rightarrow W$ 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 $T^{-1} : W \rightarrow V$ 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 $f : \mathbb{R} \rightarrow \mathbb{R}$. Given any real number c , we may define a linear transformation $T_c : F(\mathbb{R}, \mathbb{R}) \rightarrow F(\mathbb{R}, \mathbb{R})$ by declaring that $T_c(f) = cf$. Observe that if $c = 0$, then $T_c(f) = 0$ for all functions $f : \mathbb{R} \rightarrow \mathbb{R}$; however, if c is nonzero, then T is invertible. Explicitly, the linear transformation $T_{c^{-1}} : F(\mathbb{R}, \mathbb{R}) \rightarrow F(\mathbb{R}, \mathbb{R})$ 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 $T_A : \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{n \times r}$ defined by $T_A(B) = AB$ is invertible. Explicitly, the linear transformation $T_{A^{-1}} : \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{n \times r}$ 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 $T_x : \mathbb{R}[x] \rightarrow \mathbb{R}[x]$ 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 $T_x : \mathbb{R}[x] \rightarrow \mathbb{R}[x]$ 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, $T_x : \mathbb{R}[x] \rightarrow W$ 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 $S_x : W \rightarrow \mathbb{R}[x]$. 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ determines its properties, i.e., none of the information in the definition of T is extraneous. Particularly, it is possible that $T : V \rightarrow W$ fails to be surjective; however, it is always true that $T : V \rightarrow \text{range}(T)$ 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 vector spaces V and W are **isomorphic** and we write $V \cong W$ if there exists a bijective linear transformation $T : V \rightarrow W$. Consequently, by Corollary [1.13.7](#), the isomorphisms between the vector spaces V and W are precisely the invertible linear transformations $T : V \rightarrow W$. Even more, if $T : V \rightarrow W$ is an isomorphism, then the inverse transformation $T^{-1} : W \rightarrow V$ 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 $T : V \rightarrow W$ 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 i th 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 $T : V \rightarrow \mathbb{R}^{1 \times n}$ by declaring that $T(v_i) = E_i$. Given any real $1 \times n$ matrix $[a_1 \ \cdots \ a_n]$ in $\mathbb{R}^{1 \times n}$, we have that

$$[a_1 \ \cdots \ a_n] = a_1 E_1 + \cdots + a_n E_n = a_1 T(v_1) + \cdots + a_n T(v_n) = T(a_1 v_1 + \cdots + a_n v_n).$$

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

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 $T : V \rightarrow W$ from a vector space V of dimension n to a vector space W of dimension m . Conversely, and more importantly, every linear transformation $T : V \rightarrow W$ 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

$$(a_1, a_2, a_3) = (a_1, 0, 0) + (0, a_2, 0) + (0, 0, a_3) = a_1 E_1 + a_2 E_2 + a_3 E_3,$$

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 i th 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

$$(a_1, a_2, \dots, a_n) = (a_1, 0, \dots, 0) + (0, a_2, \dots, 0) + \cdots + (0, 0, \dots, a_n) = a_1 E_1 + a_2 E_2 + \cdots + a_n E_n,$$

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 $T : V \rightarrow W$ 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 $T_A : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1}$ 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 $T_A : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1}$ 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 $T_A : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1}$. 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 $T_A : \mathbb{R}^{2 \times 1} \rightarrow \mathbb{R}^{2 \times 1}$ 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 $\frac{d}{dx} : P_1(x) \rightarrow P_1(x)$ 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 $\frac{d}{dx} : P_1(x) \rightarrow P_1(x)$ 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 $T : V \rightarrow W$ 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 $T(v_j) = a_{1j}w_1 + a_{2j}w_2 + \dots + a_{mj}w_m$ for some basis w_1, \dots, w_m of W as follows.

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

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

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

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 $v = \alpha_1v_1 + \dots + \alpha_nv_n$. Particularly, the **coordinate vector** of v_j with 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 $v = \alpha_1v_1 + \dots + \alpha_nv_n$ 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_m of W .

Algorithm 1.14.8 (Matrix Representation Algorithm). Given a linear transformation $T : V \rightarrow W$ between a vector space V of dimension n and a vector space W of dimension m and ordered bases $\mathcal{B}_V = \{v_1, \dots, v_n\}$ and $\mathcal{B}_W = \{w_1, \dots, w_m\}$ of V and W , respectively, use the following algorithm to find the matrix representation of T with respect to the ordered bases \mathcal{B}_V and \mathcal{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 $T(v_1) = a_{11}w_1 + a_{21}w_2 + \dots + a_{m1}w_m$. 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 $T(v_2) = a_{12}w_1 + a_{22}w_2 + \dots + a_{m2}w_m$. 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 j th 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 $T : V \rightarrow W$ with respect to the ordered bases \mathcal{B}_V and \mathcal{B}_W .

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

$$T(E_1) = T(1, 0, 0) = (1, 2, 3) = 1 \cdot E_1 + 2 \cdot E_2 + 3 \cdot E_3$$

$$T(E_2) = T(0, 1, 0) = (2, 3, 4) = 2 \cdot E_1 + 3 \cdot E_2 + 4 \cdot E_3$$

$$T(E_3) = T(0, 0, 1) = (3, 4, 5) = 3 \cdot E_1 + 4 \cdot E_2 + 5 \cdot E_3$$

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 $T : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}^{2 \times 2}$ 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 $T : V \rightarrow W$ 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 \mathcal{B}_V of V and \mathcal{B}_W of W , respectively.*

- 1.) *If $S : W \rightarrow V$ is a linear transformation and B is the $n \times m$ matrix representation of S with respect to the ordered bases \mathcal{B}_W of W and \mathcal{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 $T^{-1} : W \rightarrow V$ of T is represented by the matrix inverse A^{-1} of A with respect to the specified ordered bases \mathcal{B}_V of V and \mathcal{B}_W of W , respectively.*

Proof. (1.) We will assume that $\mathcal{B}_V = \{v_1, \dots, v_n\}$ and $\mathcal{B}_W = \{w_1, \dots, w_m\}$. By definition of the matrix representation of T , the i th row of A consists of the scalars a_{i1}, \dots, 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 j th column of B consists of the scalars b_{1j}, \dots, b_{mj} 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

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

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 $a_{i1}b_{1j} + \dots + a_{in}b_{mj} = \sum_{k=1}^n a_{ik}b_{kj}$, as desired.

(2.) We note that T is invertible if and only if there exists a unique linear transformation $T^{-1} : W \rightarrow V$ 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 \mathcal{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 \mathcal{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 $S : W \rightarrow V$ defined by $S(w_j) = b_{1j}v_1 + \dots + b_{mj}v_m$ for each basis vector w_1, \dots, w_m of W . We have that

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

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

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

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} .

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

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

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

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

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

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 $T : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}^{2 \times 2}$ 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 $\mathbb{R}^{2 \times 2}$.

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

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

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

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

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.

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

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

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

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

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

(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 $S : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}^{2 \times 2}$ 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 of a vector space; vector spaces are ubiquitous throughout mathematics, so it is critical to understand their properties. We defined functions between vector spaces called linear transformations, 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 3×3 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 $\det : \mathbb{R}^{2 \times 2} \rightarrow \mathbb{R}$ 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.

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

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 j th 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 i th row and j th 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 i th row of A by α . Consider the $(n-1) \times (n-1)$ matrix A_{ij} obtained from A by deleting the i th row and j th 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 $\det(B) = \sum_{j=1}^n (-1)^{i+j} b_{ij} \det(B_{ij}) = \alpha \left(\sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}) \right) = \alpha \det(A)$. \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 i th row of A is zero. Considering that A is obtained from some $n \times n$ matrix B by multiplying the i th 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 i th row of A is distinct from the i th 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 i th row and the k th 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 i th row of A and the j th 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 k th row and ℓ th 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 i th row of $A_{k\ell}$ and the j th 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 $\det(A) = \sum_{\ell=1}^n (-1)^{k+\ell} a_{k\ell} \det(A_{k\ell}) = 0$. \square

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 j th row of A with the sum of α times the i th row and the j th 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 j th row of A with α times the i th 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 j th rows of A and C as the j th 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 j th row of A with the i th row of A . Explicitly, we note that C is obtained from D by multiplying the j th row of D by α . By Proposition 2.1.4, we have that $\det(C) = \alpha \det(D)$. Considering that the i th and j th rows of D are equal, it follows from Proposition 2.1.8 that $\det(D) = 0$ so that $\det(B) = \det(A) + \det(C) = \det(A) + \alpha \det(D) = \det(A)$. \square

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 i th 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 $\alpha_1, \dots, \alpha_k$. 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_j th row of A from the i th row of A for each integer $1 \leq j \leq k$ to reduce the i th 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 $\det(A) = 0$. \square

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 i th row of A with the sum of the i th and j th 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 i th row of C from the j th row of C so that the j th row of D is the i th 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 j th row of D by -1 ; then, Proposition 2.1.4 yields that $\det(B) = -\det(D) = -\det(A)$. \square

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 that all of these properties can be stated analogously for the columns.

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. \square

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 or column of A .*
- 2.) *By multiplying any column of A by α , we multiply $\det(A)$ by α .*
- 3.) *By adding a scalar multiple of one column of A to another column, we do not change $\det(A)$.*
- 4.) *By swapping two columns of A , we change the sign of $\det(A)$.*
- 5.) *We have that $\det(A) = 0$ if any column of A is zero.*
- 6.) *We have that $\det(A) = 0$ if any pair of columns of A are equal.*
- 7.) *We have that $\det(A) = 0$ if any column of A is a linear combination of other columns of A .*

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 i th row and column j th 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** $\text{adj}(A)$. 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.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ and } \text{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

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.

$$\text{adj}(A)A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = \begin{bmatrix} \det(A) & 0 \\ 0 & \det(A) \end{bmatrix} = \det(A)I_{2 \times 2}$$

Consequently, if $\det(A)$ is nonzero, then A is an invertible 2×2 matrix with $A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$.

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 i th row and j th 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 i th row and j th 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 i th row and j th column. We refer to the matrix $\Gamma = [\gamma_{ij}]_{1 \leq i \leq n, 1 \leq j \leq n}$ 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 $\text{adj}(A) = \Gamma^t$ 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 $\text{adj}(A)A = -2I_{3 \times 3} = \det(A)I_{3 \times 3}$. By Definition 2.2.4, we note that $\text{adj}(A)$ 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 i th row and j th column. Consequently, we find the following minors.

$$\begin{aligned}\mu_{11} &= \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1 & \mu_{21} &= \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 & \mu_{31} &= \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \\ \mu_{12} &= \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 & \mu_{22} &= \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 & \mu_{32} &= \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 \\ \mu_{13} &= \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 & \mu_{23} &= \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 & \mu_{33} &= \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1\end{aligned}$$

Continuing from this point, we find the $9 = 3 \cdot 3$ cofactors $\gamma_{ij} = (-1)^{i+j}\mu_{ij}$.

$$\begin{aligned}\gamma_{11} &= (-1)^{1+1}\mu_{11} = -1 & \gamma_{21} &= (-1)^{2+1}\mu_{21} = -1 & \gamma_{31} &= (-1)^{3+1}\mu_{31} = 1 \\ \gamma_{12} &= (-1)^{1+2}\mu_{12} = -1 & \gamma_{22} &= (-1)^{2+2}\mu_{22} = 1 & \gamma_{32} &= (-1)^{3+2}\mu_{32} = -1 \\ \gamma_{13} &= (-1)^{1+3}\mu_{13} = 1 & \gamma_{23} &= (-1)^{2+3}\mu_{23} = -1 & \gamma_{33} &= (-1)^{3+3}\mu_{33} = -1\end{aligned}$$

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 $\text{adj}(A) = \Gamma^t = \Gamma$. Even more, the following holds.

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

Observe that if we divide both sides by $\det(A) = -2$, then we find the following.

$$A^{-1} = \frac{1}{\det(A)} \text{adj}(A) = -\frac{1}{2} \begin{bmatrix} -1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

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 $\text{adj}(A)$ 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 i th row and j th column. Consequently, we find the following minors.

$$\mu_{11} = \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} = -3 \quad \mu_{21} = \begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} = -6 \quad \mu_{31} = \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} = -3$$

$$\mu_{12} = \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = -6 \quad \mu_{22} = \begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} = -12 \quad \mu_{32} = \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} = -6$$

$$\mu_{13} = \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = -3 \quad \mu_{23} = \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} = -6 \quad \mu_{33} = \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} = -3$$

Continuing from this point, we find the $9 = 3 \cdot 3$ cofactors $\gamma_{ij} = (-1)^{i+j} \mu_{ij}$.

$$\gamma_{11} = (-1)^{1+1} \mu_{11} = -3 \quad \gamma_{21} = (-1)^{2+1} \mu_{21} = 6 \quad \gamma_{31} = (-1)^{3+1} \mu_{31} = -3$$

$$\gamma_{12} = (-1)^{1+2} \mu_{12} = 6 \quad \gamma_{22} = (-1)^{2+2} \mu_{22} = -12 \quad \gamma_{32} = (-1)^{3+2} \mu_{32} = 6$$

$$\gamma_{13} = (-1)^{1+3} \mu_{13} = -3 \quad \gamma_{23} = (-1)^{2+3} \mu_{23} = 6 \quad \gamma_{33} = (-1)^{3+3} \mu_{33} = -3$$

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 $\text{adj}(A) = \Gamma^t = \Gamma$. Even more, the following holds.

$$\text{adj}(A)A = \begin{bmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = O_{3 \times 3} = 0I_{3 \times 3} = \det(A)I_{3 \times 3}$$

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 $\text{adj}(A)A = \det(A)I_{n \times n}$.*

Proof. By Definition 2.2.4, we have that $\text{adj}(A) = \Gamma^t$, 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 i th row and j th column of A . Consequently, the (i, j) th component of $\text{adj}(A)$ 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 $\text{adj}(A)A$ is the sum of the products of the (i, k) th component of $\text{adj}(A)$ and the (k, j) th component of A for each integer $1 \leq k \leq n$, i.e., the (i, j) th component of $\text{adj}(A)A$ 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 $\text{adj}(A)A$ 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 i th column of A with the j th column of A . Observe that for each integer $1 \leq k \leq n$, we have that $b_{ki} = a_{kj}$ because the i th column of B is equal to the j th column of A . Even more, we have that $B_{ki} = A_{ki}$ for all integers $1 \leq k \leq n$ because A and B only differ in the i th 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 $\text{adj}(A)A$ are zero, as desired. \square

Proposition 2.2.8. *Given any $n \times n$ matrix A , we have that $\text{adj}(A^t) = \text{adj}(A)^t$. Put another way, the adjugate of the transpose is the transpose of the adjugate.*

Proof. Crucially, observe that deleting the i th row and j th column of A^t is the same as deleting the i th column and j th row of A and taking its transpose because the i th row of A^t is the i th column of A and the j th column of A^t is the j th 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 $\text{adj}(A^t)$ is $(-1)^{i+j} \det((A^t)_{ij})$, hence by our opening remarks, the (i, j) th component of $\text{adj}(A^t)$ is exactly $(-1)^{i+j} \det((A_{ji})^t)$. By Proposition 2.1.12, it follows that the (i, j) th component of $\text{adj}(A^t)$ is $(-1)^{i+j} \det(A_{ji})$. Considering that this is the (j, i) th component of $\text{adj}(A)$ by definition, we conclude that the (i, j) th component of $\text{adj}(A^t)$ is the (i, j) th component of $\text{adj}(A)^t$, hence the two matrices in consideration are equal. \square

Corollary 2.2.9. *Given any $n \times n$ matrix A , we have that $A \text{adj}(A) = \det(A) I_{n \times n}$.*

Proof. By Proposition 2.2.7, we have that $\text{adj}(A^t)A^t = \det(A^t) I_{n \times n}$. By Proposition 2.1.12, we have that $\det(A^t) = \det(A)$ so that $\text{adj}(A^t)A^t = \det(A) I_{n \times n}$. By Proposition 2.2.8, we have that $\text{adj}(A^t) = \text{adj}(A)^t$ so that $\text{adj}(A)^t A^t = \det(A) I_{n \times n}$. Last, by Proposition 1.2.8, we conclude that

$$\det(A) I_{n \times n} = \det(A) I_{n \times n}^t = (\det(A) I_{n \times n})^t = (\text{adj}(A)^t A^t)^t = (A^t)^t (\text{adj}(A)^t)^t = A \text{adj}(A). \quad \square$$

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

$$\left(\frac{1}{\det(A)} \text{adj}(A) \right) A = I_{n \times n} = A \left(\frac{1}{\det(A)} \text{adj}(A) \right)$$

and $A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$. Conversely, if $\det(A) = 0$, then $\text{adj}(A)A = \det(A) I_{n \times n} = O_{n \times n}$. Consequently, there is no $n \times n$ matrix B such that $AB = I_{n \times n} = BA$, i.e., A is not invertible. \square

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.

References

- [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.
- [Moo68] J.T. Moore. *Elements of Linear Algebra and Matrix Theory*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Company, 1968.
- [Str06] G. Strang. *Linear Algebra and Its Applications*. 4th ed. Cengage Learning, 2006.