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 with substantial revisions and reorganization efforts unfolding during the Summer 2023. I express my sincere gratitude toward all of my students in MA281 (Introduction to Linear Algebra), but I am especially indebted to those who assisted in the enhancement of these notes with comments and suggestions, including Sydney Delfelder, Owen Gonzales, Ben Stubblefield, and April Thomas.

Elsewhere, some of the material from Chapter 2 (e.g., regarding the Smith Normal Form, the Rational Canonical Form, and the Jordan Canonical Form) was written independently from August 2018 to January 2022 at the request of students from the Algebra Qualifying Exam Study Group at the University of Kansas. I would like to thank the graduate students there for their motivation and participation in these study sessions — especially Wayne Ng Kwing King, Enrique Salcido, Neethu Suma-Raveendran, and Christopher Wong. I am also grateful to Souvik Dey and Monalisa Dutta for their assistance with these meetings and their peripheral contributions to these notes.

Contents

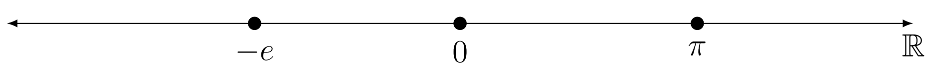
Chapter 1

Vectors and Matrices

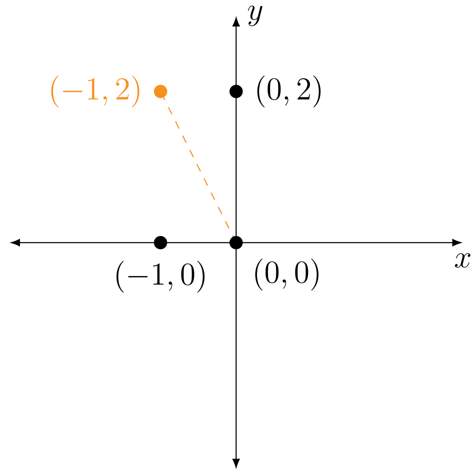
Often, in dealing with real-word problems, we are immediately met with large amounts of data and information. Even an activity as simple as baking a cake requires many ingredients and steps that must be completed in careful order, and the complexity of a task may grow exponentially as the number of inputs increases. One way to efficiently organize data is according to rows and columns in what we will refer to as vectors and matrices. We will demonstrate in this chapter that vectors and matrices admit an arithmetic that yields a highly sophisticated and widely applicable theory.

1.1 Real n-Space

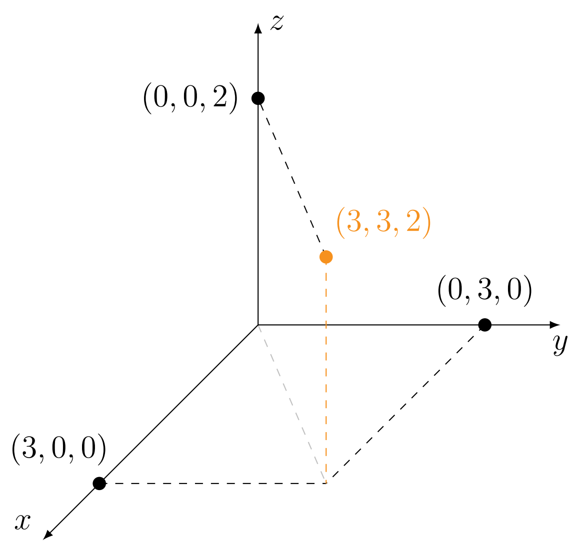
Consider the set $\mathbb{R}$ consisting of real numbers. Like usual, we may geometrically realize $\mathbb{R}$ as a line (the real number line) consisting of **points** x that lie a distance of $|x|$ from the **origin** 0 for each real number x. Explicitly, the point π lies π units to the right of the origin, whereas the point $-e$ lies e units to the left of the origin. Given any pair of real numbers $a \leq b$, the distance between the points a and b along the real number line is given by the length of the closed interval $[a, b]$; we learn in Calculus I that this distance is exactly the real number $b - a$. Consequently, the real numbers $\mathbb{R}$ admit a notion of geometry since we can conceive of things like lines and distances. Below is a visual representation of the real number line with the points $-e$, 0, and π plotted for reference.



Observe that forward and backward are the only two directions along the real number line, hence the geometry of $\mathbb{R}$ is in this sense quite simple. On the other hand, suppose that we want to keep track of both east-west movement and north-south movement. Given that an object lies x units from the origin in the east-west direction and y units in the north-south direction, we may canonically express this data as an **ordered pair** (x, y). Explicitly, if a particle lies 1 unit west and 2 units north of the origin $(0, 0)$, then it lies 1 unit to the left of the origin on the x-axis and 2 units north of the origin on the y-axis; the location of the particle in this case can be written as the ordered pair $(-1, 2)$. We refer to the collection of all ordered pairs of real numbers (x, y) as the **Cartesian product** $\mathbb{R} \times \mathbb{R}$ of the real numbers with itself, i.e., we have that $\mathbb{R} \times \mathbb{R} = \{(x, y) \mid x \text{ and } y \text{ are real numbers}\}$. Graphically, the totality of points in $\mathbb{R} \times \mathbb{R}$ form a plane, so $\mathbb{R} \times \mathbb{R}$ is often called the **Cartesian plane**. Conventionally, the Cartesian plane is denoted by $\mathbb{R}^2$ and referred to also as **real 2-space**.

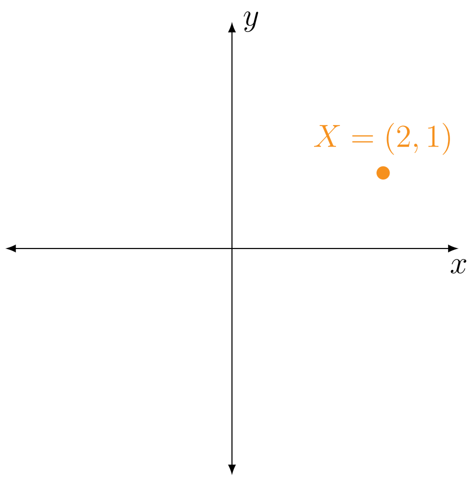


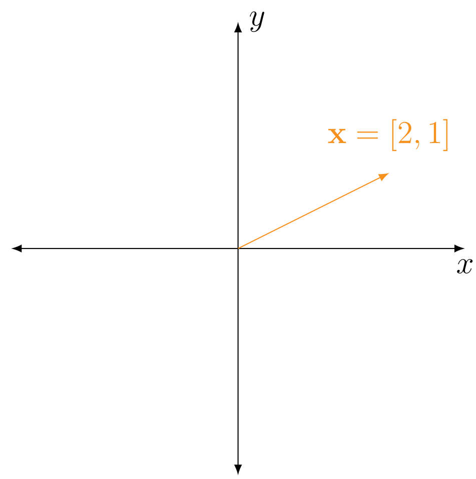
Going one step further, let us keep track of east-west, north-south, and up-down movements. Explicitly, if x measures the location of a particle in the x-axis; y measures the location of a particle in the y-axis; and z measures the location of particle in the z-axis, then the ordered triple (x, y, z) conveniently encodes this information. Like before, if the particle lies 3 units east of the origin, 3 units north of the origin, and 2 units above the origin, then the particle's location is determined by the ordered triple $(3, 3, 2)$. We denote by $\mathbb{R}^3$ the collection of all ordered triples of real numbers, i.e., we have that $\mathbb{R}^3 = \{(x, y, z) \mid x, y, \text{ and } z \text{ are real numbers}\}$; we refer to $\mathbb{R}^3$ as real 3-space.



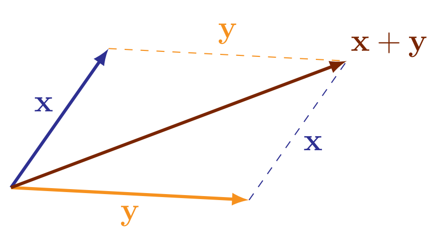
Once and for all, if n is a positive integer, then we will denote by $\mathbb{R}^n$ the collection of all n-**tuples** of real numbers, i.e., we have that $\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \mid x_1, x_2, \dots, x_n \text{ are real numbers}\}$. We will typically use a capital letter X to denote a real n-tuple $(x_1, x_2, \dots, x_n)$. We refer to the real number x_1 as the first **coordinate** of X; we refer to the real number x_2 as the second coordinate of X; we refer to the real number x_n as the nth coordinate of X; and in general, the real number x_i is called the ith coordinate of X for each integer $1 \leq i \leq n$. Every point in real n-space is uniquely determined by its coordinates: indeed, if we consider any pair of points $X = (x_1, x_2, \dots, x_n)$ and $Y = (y_1, y_2, \dots, y_n)$ such that $(x_1, x_2, \dots, x_n) = X = Y = (y_1, y_2, \dots, y_n)$, then each of the coordinates on the left-hand side must be equal to the corresponding coordinate on the right-hand side, i.e., we must have that $x_i = y_i$ for all integers $1 \leq i \leq n$. Even though it is not possible to geometrically visualize points in real n-space for any integer $n \geq 4$, it is still meaningful to discuss this notion. Explicitly, every set of data consisting of n distinct real parameters induces an element of real n-space $\mathbb{R}^n$.

Continuing from a geometric perspective, it is useful to distinguish between points and **vectors** in real n-space. Explicitly, we may view the vector $\mathbf{x} = [x_1, x_2, \dots, x_n]$ corresponding to the point $X = (x_1, x_2, \dots, x_n)$ in real n-space as a ray (or arrow) emanating from the origin and extending to the point $(x_1, x_2, \dots, x_n)$. Explicitly, the vector $\mathbf{x} = [1, 2, 3, 4]$ of $\mathbb{R}^4$ can be represented by the ray extending from the origin $(0, 0, 0, 0)$ to the point $(1, 2, 3, 4)$ in $\mathbb{R}^4$. We refer to the vector $\mathbf{x}$ in this case as lying in **standard position**. We will come to find that despite the mathematical equivalence of points X and vectors $\mathbf{x}$ in real n-space, the benefit of this distinction is that vectors in real n-space are translation-invariant and possess a notion of length. Often, we will restrict our attention to the Cartesian plane $\mathbb{R}^2$ or real 3-space $\mathbb{R}^3$, where we can visualize these vectors.

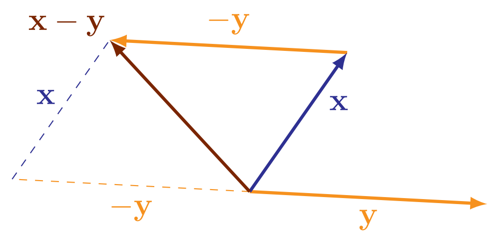




Geometrically, we may prescribe the arithmetic of **vector addition** as follows: to determine the vector sum $\mathbf{x} + \mathbf{y}$ pictorially, visualize $\mathbf{x}$ and $\mathbf{y}$ as rays emanating from the origin; translate $\mathbf{y}$ so that the "foot" of $\mathbf{y}$ lies at the "head" of $\mathbf{x}$; and draw the ray emanating from the "foot" of $\mathbf{x}$ to the "head" of $\mathbf{y}$. Equivalently, one could also determine $\mathbf{x} + \mathbf{y}$ by translating $\mathbf{x}$ so that the "foot" of $\mathbf{x}$ lies at the "head" of $\mathbf{y}$ and subsequently drawing the raw emanating from the "foot" of $\mathbf{y}$ to the "head" of $\mathbf{x}$. Either way, the resulting **vector sum** can be pictured as follows.



We refer to the process of computing the vector sum $\mathbf{x} + \mathbf{y}$ in this manner as the **Parallelogram Law** because the resulting diagram forms a parallelogram. We will in no time describe the algebraic operations of vector addition and scalar multiplication, but for now, we note that for any vector $\mathbf{x}$ emanating from the origin to a point X in real n-space, the point $-X$ is obtained from X by taking the coordinates of X with opposite sign. Explicitly, if we assume that $X = (x_1, x_2, \dots, x_n)$, then $-X = (-x_1, -x_2, \dots, -x_n)$. By identifying the vector $-\mathbf{x}$ in standard position with the ray emanating from the origin to the point $-X$, we find that $-\mathbf{x}$ is nothing more than $\mathbf{x}$ in the "opposite direction." Consequently, translating $-\mathbf{x}$ so that it overlaps $\mathbf{x}$, the "head" of $-\mathbf{x}$ lies at the "foot" of $\mathbf{x}$ and vice-versa. We may in this way describe **vector subtraction** pictorially as follows.



Even more, **scalar multiplication** of a vector $\mathbf{x}$ by a real number (or **scalar**) α can be visualized by taking the vector $\alpha\mathbf{x}$ as the ray emanating from the origin with length $|\alpha|$ times the length of $\mathbf{x}$ in the same direction of $\mathbf{x}$ if α is positive and in the opposite direction if α is negative. We will henceforth say that two vectors $\mathbf{x}$ and $\mathbf{y}$ in real n-space are **parallel** if there exists a nonzero real number α such that $\mathbf{y} = \alpha\mathbf{x}$. We will say that $\mathbf{x}$ and $\alpha\mathbf{x}$ have the **same direction** if $\alpha > 0$; they have the **opposite direction** if $\alpha < 0$; and the vector $\mathbf{0}$ corresponding to the origin has no direction. Certainly, a pair of vectors in real n-space need not be parallel, hence in general, it might not be possible to say that an arbitrary pair of vectors have the same or opposite direction.

Until now, we have considered vectors in real n-space from a primarily geometric standpoint by way of diagrams and visualizations; however, this might very well come across as unsatisfactory to some readers for several reasons — not least of all that it is difficult to draw vectors in three-space and impossible to picture vectors with more coordinates than that. Bearing this in mind, we turn our attention to an algebraic description of vectors in real n-space. We will to this end represent vectors $\mathbf{v}$ and $\mathbf{w}$ in real n-space according to their coordinates. Explicitly, we will write $\mathbf{v} = [v_1, v_2, \dots, v_n]$ for some positive integer n and real numbers $v_1, v_2, \dots, v_n$. Given any positive integer m and any real numbers $w_1, w_2, \dots, w_m$, the vectors $\mathbf{v}$ and $\mathbf{w}$ are equal (i.e., $\mathbf{v} = \mathbf{w}$) if and only if $m = n$ and $w_i = v_i$ for each integer $1 \leq i \leq n$. Concretely, a pair of vectors expressed in terms of their coordinates are equal if and only if (1.) the number of coordinates of the vectors is the same and (2.) the corresponding coordinates of the vectors are the same. We reserve the notation $\mathbf{0}$ for the **zero vector** whose coordinates are all zero, i.e., $\mathbf{0} = [0, 0, \dots, 0]$. Crucially, all though we will indiscriminately use the symbol $\mathbf{0}$ to denote the zero vector in all contexts, it is important to realize that the zero vector in real n-space differs as n ranges across all positive integers.

We define vector addition and scalar multiplication **coordinatewise**. Explicitly, for any vectors $\mathbf{v} = [v_1, v_2, \dots, v_n]$ and $\mathbf{w} = [w_1, w_2, \dots, w_n]$ in real n-space and any real number α, we declare that

$$\begin{aligned} \mathbf{v} + \mathbf{w} &= [v_1 + w_1, v_2 + w_2, \dots, v_n + w_n] \text{ and} \\ \alpha\mathbf{v} &= [\alpha v_1, \alpha v_2, \dots, \alpha v_n]. \end{aligned}$$

Consequently, it follows that vector subtraction is carried out componentwise, as well.

$$\mathbf{v} - \mathbf{w} = [v_1 - w_1, v_2 - w_2, \dots, v_n - w_n]$$

Example 1.1.1. Consider the vectors $\mathbf{u} = [1, 1, -1]$, $\mathbf{v} = [1, 2, 3]$, and $\mathbf{w} = [0, -2, -2]$ in real 3-space. Observe that $\mathbf{u} + \mathbf{v} = [2, 3, 2]$, $-\mathbf{w} = [0, 2, 2]$, $\mathbf{v} - \mathbf{w} = [1, 4, 5]$, and $3\mathbf{u} = [3, 3, -3]$.

Example 1.1.2. Observe that the vectors $\mathbf{u} = [1, 0, -1]$ and $\mathbf{v} = [-3, 0, 3]$ are parallel because we have that $\mathbf{v} = -3\mathbf{u}$, hence $\mathbf{u}$ and $\mathbf{v}$ have the opposite direction; however, the vector $\mathbf{w} = [-1, 1, 1]$ is not parallel to either $\mathbf{u}$ or $\mathbf{v}$. (We will soon see that it is in fact perpendicular to $\mathbf{u}$ and $\mathbf{v}$.)

Considering that vector addition and scalar multiplication in real n-space are determined by the coordinates of the underlying vectors, the following proposition should not come as a surprise.

Proposition 1.1.3 (Properties of Vector Arithmetic in Real n-Space)**.** *Consider any vectors* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *in real n-space and any real numbers α and β. We have that*

1.) *vector addition is* **associative**, *i.e.,* $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$*;*

2.) *vector addition is* **commutative**, *i.e.,* $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$*;*

3.) *the zero vector* $\mathbf{0}$ *is the* **additive identity**, *i.e.,* $\mathbf{v} + \mathbf{0} = \mathbf{v}$*;*

4.) *the* **additive inverse** *of* $\mathbf{v}$ *is* $-\mathbf{v}$, *i.e.,* $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$*;*

5.) *scalar multiplication is associative, i.e.,* $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$*;*

6.) *scalar multiplication is* **distributive** *across vector addition, i.e.,* $\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$*;*

7.) *scalar multiplication is distributive across scalar addition, i.e.,* $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$*; and*

8.) *the* **multiplicative identity** 1 *preserves scale, i.e.,* $1\mathbf{v} = \mathbf{v}$.

Proof. Each of the above properties can be verified directly by listing the coordinates of the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ and performing the vector addition and scalar multiplication coordinatewise. □

Example 1.1.4. Consider the vectors $\mathbf{u} = [1, 2, 5]$, $\mathbf{v} = [-1, 3, 4]$, and $\mathbf{w} = [3, 1, 6]$ in real 3-space. We can compute $3\mathbf{u} - 5(\mathbf{v} - \mathbf{w})$ according to the Properties of Vector Arithmetic in Real n-Space.

$$\begin{aligned} 3\mathbf{u} - 5(\mathbf{v} - \mathbf{w}) &= 3\mathbf{u} - 5\mathbf{v} + 5\mathbf{w} \\ &= 3[1, 2, 5] - 5[-1, 3, 4] + 5[3, 1, 6] \\ &= [3, 6, 15] + [5, -15, -20] + [15, 5, 30] \\ &= [23, -4, 25] \end{aligned}$$

We could alternatively computed the vector difference $\mathbf{v} - \mathbf{w} = [-4, 2, -2]$ and proceeded as follows.

$$3\mathbf{u} - 5(\mathbf{v} - \mathbf{w}) = 3[1, 2, 5] - 5[-4, 2, -2] = [3, 6, 15] + [20, -10, 10] = [23, -4, 25]$$

Either way, we obtain the same coordinates for the vector $3\mathbf{u} - 5(\mathbf{v} - \mathbf{w})$, as expected.

We refer to the vector $3\mathbf{u} - 5\mathbf{v} + 5\mathbf{w}$ in Example 1.1.4 as a **linear combination** of the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$. Generally, for any vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ in real n-space and any scalars $\alpha_1, \alpha_2, \dots, \alpha_k$, we refer to the vector $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k$ as the linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ with **scalar coefficients** $\alpha_1, \alpha_2, \dots, \alpha_k$. Later, these vectors will become a critical object of study; however, for now, it is important to note that every vector $\mathbf{v}$ in real n-space can be written uniquely as a linear combination of the **standard basis vectors** $\mathbf{e}_i$ whose ith coordinate is 1 and whose other coordinates are 0 for all integers $1 \leq i \leq n$. Concretely, there are two standard basis vectors in real 2-space: they are $\mathbf{e}_1 = [1, 0]$ and $\mathbf{e}_2 = [0, 1]$. Likewise, there are three standard basis vectors in real 3-space — namely, $\mathbf{e}_1 = [1, 0, 0]$, $\mathbf{e}_2 = [0, 1, 0]$, and $\mathbf{e}_3 = [0, 0, 1]$. Observe that

$$[v_1, v_2, v_3] = [v_1, 0, 0] + [0, v_2, 0] + [0, 0, v_3] = v_1[1, 0, 0] + v_2[0, 1, 0] + v_3[0, 0, 1] = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$$

yields an expression of the vector $[v_1, v_2, v_3]$ as a linear combination of the standard basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ with scalar coefficients corresponding to the coordinates of $\mathbf{v}$. By analogy, this process can be carried out for any vector in real n-space with respect to the standard basis $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$.

Example 1.1.5. Consider the vectors $\mathbf{u} = [1, 2, 5]$, $\mathbf{v} = [-1, 3, 4]$, and $\mathbf{w} = [3, 1, 6]$ in real 3-space. Let us verify that $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$. By definition, we must find real numbers α and β such that $\mathbf{w} = \alpha\mathbf{u} + \beta\mathbf{v}$. Expressing this relation in terms of coordinates yields that

$$[3, 1, 6] = \mathbf{w} = \alpha\mathbf{u} + \beta\mathbf{v} = \alpha[1, 2, 5] + \beta[-1, 3, 4] = [\alpha - \beta, 2\alpha + 3\beta, 5\alpha + 4\beta],$$

so it suffices to solve the induced **system of equations** with three equations and two unknowns.

$$\begin{cases} \alpha - \beta = 3 \\ 2\alpha + 3\beta = 1 \\ 5\alpha + 4\beta = 6 \end{cases}$$

By the first equation, it follows that $\alpha = \beta + 3$; substitute this into the second equation to find that

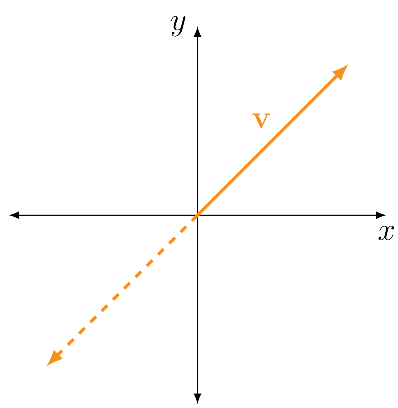
$$1 = 2\alpha + 3\beta = 2(\beta + 3) + 3\beta = 5\beta + 6,$$

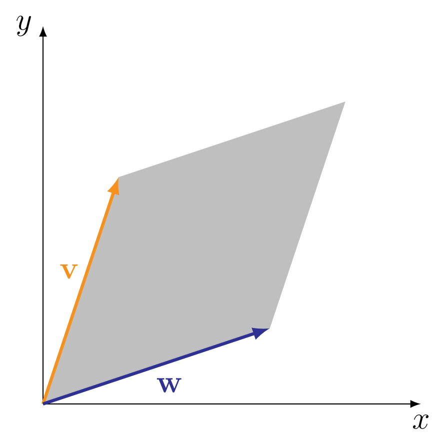
hence we conclude that $\beta = -1$, from which it follows that $\alpha = 2$. We conclude at last that

$$[3, 1, 2] = \mathbf{w} = 2\mathbf{u} - \mathbf{v} = 2[1, 2, 5] - [-1, 3, 4].$$

We remark that the third equation $5\alpha + 4\beta = 6$ was not required to solve this system.

Geometrically, linear combinations of vectors give rise to lines, planes, and hyperplanes. Explicitly, given any nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ in real n-space, the collection $\{\alpha\mathbf{v} \mid \alpha \in \mathbb{R}\}$ of all possible linear combinations of $\mathbf{v}$ is called the **line along v**, and the collection $\{\alpha\mathbf{v} + \beta\mathbf{w} \mid \alpha, \beta \in \mathbb{R}\}$ of all possible linear combinations of $\mathbf{v}$ and $\mathbf{w}$ is called the **plane spanned by v** and **w**.





Example 1.1.6. Consider the vectors $\mathbf{v} = [1, 3]$ and $\mathbf{w} = [3, 1]$ in real 2-space. By definition, the line along $\mathbf{v}$ is given by the collection of points $(\alpha, 3\alpha)$ such that α is a real number. Consequently, the points $(0, 0)$, $(3, 9)$, and $(-2, -6)$ lie on the line along $\mathbf{v}$; however, the point $(2, 2)$ does not lie on this line: indeed, the point $(2, 2)$ lies on the line along $\mathbf{v}$ if and only if there exists a real number α such that $(2, 2) = (\alpha, 3\alpha)$ if and only if $\alpha = 2$ and $3\alpha = 2$. Because this is impossible, the point $(2, 2)$ does not lie on the lie along $\mathbf{v}$. We say in this case that the system of equations

$$\begin{cases} \alpha = 2 \\ 3\alpha = 2 \end{cases}$$

is **inconsistent** because there is no real number α for which both equations hold.

Likewise, if we wish to determine if the point $(12, 12)$ lies in the planned spanned by $\mathbf{v}$ and $\mathbf{w}$, we seek real numbers α and β such that $[12, 12] = \alpha\mathbf{v} + \beta\mathbf{w} = \alpha[1, 3] + \beta[3, 1] = [\alpha + 3\beta, 3\alpha + \beta]$, hence we must solve the induced system of equations with two equations and two unknowns.

$$\begin{cases} \alpha + 3\beta = 12 \\ 3\alpha + \beta = 12 \end{cases}$$

Like before, if we substitute $\beta = -3\alpha + 12$ from the second equation into the first equation, then

$$12 = \alpha + 3\beta = \alpha + 3(-3\alpha + 12) = -8\alpha + 36$$

implies that $-8\alpha = -24$ so that $\alpha = 3$, from which it follows that $\beta = 3$.

By altering the presentation of our vectors from rows to columns, the relationship between linear combinations of vectors and systems of linear equations becomes all the more evident: by expressing the vectors $\mathbf{v} = [1, 3]$ and $\mathbf{w} = [3, 1]$ of Example 1.1.6 as column vectors, the containment of a point (x, y) within the plane spanned by $\mathbf{v}$ and $\mathbf{w}$ can be determined by solving the vector equation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \alpha\mathbf{v} + \beta\mathbf{w} = \alpha \begin{bmatrix} 1 \\ 3 \end{bmatrix} + \beta \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha + 3\beta \\ 3\alpha + \beta \end{bmatrix}.$$

Comparing the rows of the vectors on the left- and right-hand sides of this equation with the real numbers $x = 12$ and $y = 12$ yields the system of equations from Example 1.1.6.

By analogy to lines and planes spanned by vectors in real n-space, given any nonzero vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ in real n-space, the collection of all possible linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ forms a hyperplane called the **span** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ and denoted by

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} = \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k \mid \alpha_1, \alpha_2, \ldots, \alpha_k \text{ are real numbers}\}.$$

We will return to discuss the notion of span in greater detail in Section 1.6.

1.2 Vector Magnitude and the Dot Product

Our aim throughout this section is to systematically develop the theory of **Euclidean geometry** in real n-space that was suggested peripherally (and perhaps unsatisfactorily) in the previous section.

We begin with a notion of distance. Given any points $X = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ in real n-space, we define the **distance** between X and Y as the following real number.

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

Consequently, the distance from the origin $O = (0, 0, \dots, 0)$ to the point X is denoted as follows.

$$d(X, O) = \sqrt{x_1^2 + \cdots + x_n^2}$$

We note that this definition of distance is merely a generalization of the length of the hypotenuse of the right triangle formed by the x-axis, the y-axis, and a point in the Cartesian plane: indeed, if we could visualize the right triangle formed by the origin of $\mathbb{R}^n$, the point $(x_1, \dots, x_{n-1}, 0)$, and the point $X = (x_1, \dots, x_{n-1}, x_n)$ in $\mathbb{R}^n$, then the length of its hypotenuse is precisely $d(X, O)$.

Consider the vector $\mathbf{x} = [x_1, \dots, x_n]$ lying in standard position in real n-space. Geometrically, $\mathbf{x}$ can be viewed as the ray emitting from the origin to the point $X = (x_1, \dots, x_n)$, hence the **length** of the vector $\mathbf{x}$ is precisely the distance from the origin O to the point X, i.e., the length of $\mathbf{x}$ is

$$\|\mathbf{x}\| = d(X, O) = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Often, we will rather refer to the quantity $\|\mathbf{x}\|$ as the **magnitude** or **norm** of the vector $\mathbf{x}$.

Example 1.2.1. Consider the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ from Example 1.1.1. Computing the magnitudes of each vector yields $\|\mathbf{u}+\mathbf{v}\| = \sqrt{2^2 + 3^2 + 2^2} = \sqrt{17}$ and $\|-\mathbf{w}\| = \sqrt{0^2 + 2^2 + 2^2} = 2\sqrt{2} = \|\mathbf{w}\|$ and $\|3\mathbf{x}\| = \sqrt{3^2 + 3^2 + (-3)^2} = 3\sqrt{3} = 3\|\mathbf{x}\|$; these last two examples indicate a general phenomenon.

Proposition 1.2.2. *Consider any positive integer n and any vector $\mathbf{x} = [x_1, \dots, x_n]$ in real n-space.*

1.) *We have that $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x}$ is the zero vector.*

2.) *We have that $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ for all real numbers α.*

Proof. (1.) By definition, we have that $\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2} = 0$ if and only if $x_1^2 + \cdots + x_n^2 = 0$. Clearly, if $\mathbf{x}$ is the zero vector, then $x_1 = \cdots = x_n = 0$ so that $x_1^2 + \cdots + x_n^2 = 0^2 + \cdots + 0^2 = 0$. Conversely, if $\mathbf{x}$ is a nonzero vector, then its ith coordinate x_i must be nonzero for some integer $1 \leq i \leq n$. Considering that the square of a nonzero real number if a positive real number, we have that $x_i^2 > 0$. Even more, the square of any real number is non-negative, hence we have that $\|\mathbf{x}\|^2 = x_1^2 + \cdots + x_n^2 \geq x_i^2 > 0$. We conclude that $\|\mathbf{x}\|$ must be nonzero if $\mathbf{x}$ is nonzero.

(2.) We define $\alpha\mathbf{x} = \alpha[x_1, \dots, x_n] = [\alpha x_1, \dots, \alpha x_n]$. Consequently, the definition of magnitude yields $\|\alpha\mathbf{x}\| = \sqrt{(\alpha x_1)^2 + \cdots + (\alpha x_n)^2} = \sqrt{\alpha^2(x_1^2 + \cdots + x_n)^2} = |\alpha|\sqrt{x_1^2 + \cdots + x_n^2} = |\alpha|\|\mathbf{x}\|$. □

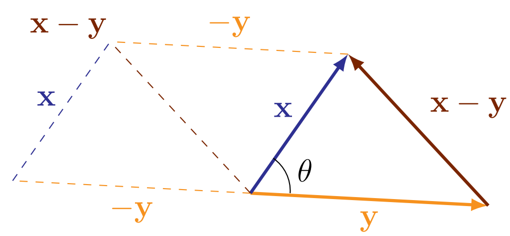
Conventionally, vectors of magnitude 1 are referred to as **unit vectors**. By Proposition 1.2.2, it can be shown that every nonzero vector $\mathbf{x}$ gives rise to a unique unit vector $\frac{1}{\|\mathbf{x}\|}\mathbf{x}$.

Corollary 1.2.3. *Every nonzero vector $\mathbf{x}$ of $\mathbb{R}^n$ induces a unit vector $\frac{1}{\|\mathbf{x}\|}\mathbf{x}$ of $\mathbb{R}^n$.*

Proof. By Proposition 1.2.2, if $\mathbf{x}$ is any nonzero vector of $\mathbb{R}^n$, then $\|\mathbf{x}\|$ is a positive real number. Consequently, we have that $\alpha = \frac{1}{\|\mathbf{x}\|}$ is a positive real number such that $\|\alpha\mathbf{x}\| = \alpha\|\mathbf{x}\| = 1$. □

Example 1.2.4. Consider the vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ from Example 1.2.1. We demonstrated previously that $\|\mathbf{x}+\mathbf{y}\| = \sqrt{17}$ and $\|\mathbf{z}\| = 2\sqrt{2}$, hence $\frac{1}{\sqrt{17}}(\mathbf{x}+\mathbf{y})$ and $\frac{1}{2\sqrt{2}}\mathbf{z}$ are unit vectors of $\mathbb{R}^3$.

Consider any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ lying in standard position in real n-space for some positive integer n. Certainly, if $n = 2$ or $n = 3$, then we could visualize $\mathbf{x}$ and $\mathbf{y}$ in the Cartesian plane $\mathbb{R}^2$ or in the real 3-space $\mathbb{R}^3$ that we occupy, take a protractor, and measure the angle θ formed by the intersection of $\mathbf{x}$ and $\mathbf{y}$ at the origin. Pictorially, we would obtain the following diagram.



By the Law of Cosines, the triangle spanned by the vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{x}-\mathbf{y}$ gives the following.

$$\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta)$$

Observe that if $\mathbf{x} = [x_1, \dots, x_n]$ and $\mathbf{y} = [y_1, \dots, y_n]$, then by definition of the magnitude of a vector, it follows that $\|\mathbf{x}\|^2 = x_1^2 + \cdots + x_n^2$ and $\|\mathbf{y}\|^2 = y_1^2 + \cdots + y_n^2$ so that

$$\|\mathbf{x}-\mathbf{y}\|^2 = (x_1-y_1)^2 + \cdots + (x_n-y_n)^2 = x_1^2 + \cdots + x_n^2 + y_1^2 + \cdots + y_n^2 - 2(x_1y_1 + \cdots + x_ny_n).$$

Combining this formula with the above equation obtained from the Law of Cosines yields that

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta) = \|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2(x_1y_1 + \cdots + x_ny_n)$$

so that $\|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta) = x_1y_1 + \cdots + x_ny_n$. We refer to the real number $x_1y_1 + \cdots + x_ny_n$ as the **dot product** $\mathbf{x}\cdot\mathbf{y}$ of the real vectors $\mathbf{x}$ and $\mathbf{y}$. Explicitly, if $\mathbf{x} = [x_1, \dots, x_n]$ and $\mathbf{y} = [y_1, \dots, y_n]$, then

$$\mathbf{x}\cdot\mathbf{y} = x_1y_1 + \cdots + x_ny_n.$$

Even more, it is clear from this exposition that the dot product informs the geometry of real n-space.

Proposition 1.2.5. *Given any pair of nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ lying in standard position in real n-space, the angle θ of intersection between the vectors $\mathbf{x}$ and $\mathbf{y}$ at the origin satisfies that*

$$\theta = \cos^{-1}\left(\frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\right)$$

Essentially, the formula is obtained from the previous paragraph by solving for θ in the identity $\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta)$. Consequently, we will typically refer to the identity $\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta)$ as the **geometric representation** of the dot product. Extending this notion of geometry of vectors in real n-space, we will say that $\mathbf{x}$ and $\mathbf{y}$ are **orthogonal** (or **perpendicular**) provided that $\mathbf{x}\cdot\mathbf{y} = 0$. Observe that the angle of intersection between orthogonal vectors is $\cos^{-1}(0) = 90°$.

Example 1.2.6. Consider the vectors $\mathbf{x} = [1, 1, -1]$, $\mathbf{y} = [1, 2, 3]$, and $\mathbf{z} = [0, -2, -2]$ in $\mathbb{R}^3$. By definition of the dot product, we obtain the following identities.

$$\begin{aligned}
\mathbf{x} \cdot \mathbf{x} &= (1)(1) + (1)(1) + (-1)(-1) = 3 \\
\mathbf{x} \cdot \mathbf{y} &= (1)(1) + (1)(2) + (-1)(3) = 0 \\
\mathbf{x} \cdot \mathbf{z} &= (1)(0) + (1)(-2) + (-1)(-2) = 0 \\
\mathbf{y} \cdot \mathbf{z} &= (1)(0) + (2)(-2) + (3)(-2) = -10
\end{aligned}$$

Consequently, it follows that $\mathbf{x}$ is orthogonal to both $\mathbf{y}$ and $\mathbf{z}$, but $\mathbf{x}$ is not orthogonal to itself and $\mathbf{y}$ is not orthogonal to $\mathbf{z}$. Even more, we have that $\mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2$.

Example 1.2.7. Consider the vectors $\mathbf{x} = [1, 2, 0, 2]$ and $\mathbf{y} = [-3, 1, 1, 5]$ in $\mathbb{R}^4$. Even though we cannot visualize $\mathbf{x}$ and $\mathbf{y}$ as rays emitting from the origin because they exist in real 4-space, we can find their angle θ of intersection. By definition of vector magnitude, we have that

$$\begin{aligned}
\|\mathbf{x}\| &= \sqrt{1^2 + 2^2 + 0^2 + 2^2} = \sqrt{9} = 3 \text{ and} \\
\|\mathbf{y}\| &= \sqrt{(-3)^2 + 1^2 + 1^2 + 5^2} = \sqrt{36} = 6.
\end{aligned}$$

By definition of the dot product, it follows that $\mathbf{x} \cdot \mathbf{y} = (1)(-3) + (2)(1) + (0)(1) + (2)(5) = 9$. Consequently, we conclude by the geometric representation of the dot product that

$$\theta = \cos^{-1}\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}\right) = \cos^{-1}\left(\frac{9}{(3)(6)}\right) = \cos^{-1}\left(\frac{1}{2}\right) = 60^\circ.$$

Considering that the dot product of vectors in real n-space is determined by the coordinates of the underlying vectors, the following proposition is unsurprising and straightforward to prove.

Proposition 1.2.8 (Properties of the Dot Product in Real n-Space)**.** *Consider any vectors* $\mathbf{x}$, $\mathbf{y}$, *and* $\mathbf{z}$ *lying in standard position in real n-space and any real number α. We have that*

(1.) *the dot product is commutative, i.e.,* $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$*;*

(2.) *the dot product is distributive across vector addition, i.e.,* $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$*;*

(3.) *the dot product is* **homogeneous***, i.e.,* $(\alpha\mathbf{x}) \cdot \mathbf{y} = \alpha(\mathbf{x} \cdot \mathbf{y}) = \mathbf{x} \cdot (\alpha\mathbf{y})$*; and*

(4.) *the dot product is* **non-degenerate***, i.e.,* $\mathbf{x} \cdot \mathbf{x}$ *is nonzero if and only if* $\mathbf{x}$ *is nonzero.*

Even more, the dot product in real n-space satisfies the **Law of Cosines**

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta).$$

Proof. Each of the above properties can be verified directly by listing the coordinates of the vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$, computing the dot product, and appealing to familiar properties of real numbers.

Even more, the commutativity and distributivity of the dot product yield that

$$\|\mathbf{x} - \mathbf{y}\|^2 = (\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) = \mathbf{x} \cdot \mathbf{x} + \mathbf{y} \cdot \mathbf{y} - 2(\mathbf{x} \cdot \mathbf{y}) = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2(\mathbf{x} \cdot \mathbf{y}).$$

By the geometric representation of the dot product, we find that $\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\|\cos(\theta)$ for the angle θ of intersection between $\mathbf{x}$ and $\mathbf{y}$. Considering that the vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{x} - \mathbf{y}$ induce a triangle of side lengths $\|\mathbf{x}\|$, $\|\mathbf{y}\|$, and $\|\mathbf{x} - \mathbf{y}\|$, respectively, such that the side of length $\|\mathbf{x} - \mathbf{y}\|$ lies opposite the angle θ of intersection between $\mathbf{x}$ and $\mathbf{y}$, the Law of Cosines holds in this case. $\square$

By applying the Properties of the Dot Product in Real n-Space in the case of orthogonal vectors, we can prove the following important properties of orthogonal vectors.

Proposition 1.2.9 (Properties of Orthogonal Vectors in Real n-Space)**.** *Consider any vectors* $\mathbf{x}$, $\mathbf{y}$, *and* $\mathbf{z}$ *lying in standard position in real* n*-space.*

1.) *If* $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$ *and* $\mathbf{z}$, *then* $\mathbf{x}$ *is orthogonal to* $\mathbf{y} + \mathbf{z}$.

2.) *If* $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$, *then* $\mathbf{x}$ *is orthogonal to* $\alpha\mathbf{y}$ *for all real numbers* α.

3.) *If* $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$, *then their angle of intersection is* $90°$, *i.e.,* $\mathbf{x}$ *and* $\mathbf{y}$ *are perpendicular.*

4.) *If* $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$, *then* $\|\mathbf{x}-\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, *i.e., the* **Pythagorean Theorem** *holds.*

Proof. 1.) By definition, if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal and $\mathbf{x}$ and $\mathbf{z}$ are orthogonal, then $\mathbf{x} \cdot \mathbf{y} = 0$ and $\mathbf{x} \cdot \mathbf{z} = 0$. By Proposition 1.2.8, it follows that $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z} = 0$.

2.) By Proposition 1.2.8, we have that $\mathbf{x} \cdot (\alpha\mathbf{y}) = \alpha(\mathbf{x} \cdot \mathbf{y}) = 0$ for all real numbers α.

3.) By Proposition 1.2.5, if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal vectors lying in standard position in real n-space, then the angle θ of intersection between the vectors $\mathbf{x}$ and $\mathbf{y}$ is given by $\theta = \cos^{-1}(0) = 90°$.

4.) By the Law of Cosines and its proof in Proposition 1.2.8, we have that

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2(\mathbf{x} \cdot \mathbf{y}).$$

Considering that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, we conclude that $2(\mathbf{x} \cdot \mathbf{y}) = 0$, as desired. □

Example 1.2.10. We determine in this example a unit vector perpendicular to $\mathbf{x} = [-1, 3, 4]$. By definition, we seek a real vector $\mathbf{y} = [y_1, y_2, y_3]$ such that $\mathbf{x} \cdot \mathbf{y} = 0$ and $\|\mathbf{y}\| = 1$. Computing the dot product of $\mathbf{x}$ and $\mathbf{y}$, we find that $0 = \mathbf{x} \cdot \mathbf{y} = -y_1 + 3y_2 + 4y_3$. We have three variables and only one equation, hence there must be two free variables that we are allowed to set equal to anything that is convenient. We will choose $y_1 = 0$ and $y_2 = -4$; the resulting equation is $3(-4) + 4y_3 = 0$ so that $4y_3 = 3(4)$ and $y_3 = 3$. Consequently, the vector $\mathbf{y} = [0, -4, 3]$ is orthogonal to $\mathbf{x}$; however, its magnitude is $\sqrt{0^2 + (-4)^2 + 3^2} = 5$, so it is not a unit vector. By Proposition 1.2.3, we find that $\frac{1}{5}\mathbf{y}$ is a unit vector; it is orthogonal to $\mathbf{x}$ by Proposition 1.2.9 because $\mathbf{y}$ is orthogonal to $\mathbf{x}$.

Example 1.2.11. We determine in this example a unit vector perpendicular to $\mathbf{x} = [-1, 3, 4]$ and $\mathbf{y} = [2, 1, -1]$. Like before in Example 1.2.10, we must solve the following system of equations.

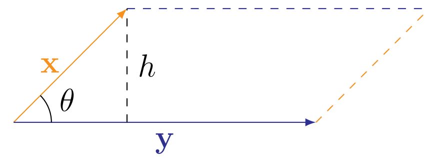
$$\begin{aligned} 0 &= \mathbf{x} \cdot [z_1, z_2, z_3] = -z_1 + 3z_2 + 4z_3 \\ 0 &= \mathbf{y} \cdot [z_1, z_2, z_3] = 2z_1 + z_2 - z_3 \end{aligned}$$

By adding twice the first equation to the second equation, we find that $7z_2 + 7z_3 = 0$ or $z_3 = -z_2$. We have two equations in three unknowns, so we will have at least one free variable; however, as the arithmetic bears out, we find that z_3 depends on z_2, hence z_2 is a second free variable. By setting $z_1 = 0$ and $z_2 = 1$, we find that $z_3 = -1$ and $\mathbf{z} = [0, 1, -1]$ is orthogonal to $\mathbf{x}$ and $\mathbf{y}$. Considering that $\|\mathbf{z}\| = \sqrt{0^2 + 1^2 + (-1)^2} = \sqrt{2}$, we conclude that $\frac{1}{\sqrt{2}}\mathbf{z}$ is a unit vector orthogonal to $\mathbf{x}$ and $\mathbf{y}$.

Geometrically, we have seen to our pleasant surprise that the dot product in real n-space enjoys many nice properties. But perhaps one of its most astounding features is the following.

Proposition 1.2.12. *Given any nonzero, non-parallel vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *lying in standard position in real* n*-space, the area of the* **parallelogram** *spanned by* $\mathbf{x}$ *and* $\mathbf{y}$ *is* $\|\mathbf{x}\|\|\mathbf{y}\|\sin(\theta)$.

Proof. Pictorially, the parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$ can be determined as follows.



Observe that the angle θ of intersection between $\mathbf{x}$ and $\mathbf{y}$ satisfies that $h = \|\mathbf{x}\|\sin(\theta)$. Because the area of a parallelogram is the product of its base and its height, it is $h\|\mathbf{y}\| = \|\mathbf{x}\|\|\mathbf{y}\|\sin(\theta)$. □

Before we conclude this section, we state and prove two inequalities regarding vectors in real n-space. Crucially, the following proposition provides a purely algebraic foundation for the geometry of the dot product in real n-space that we have as yet taken for granted (cf. Proposition 1.2.5).

Theorem 1.2.13 (Cauchy-Schwarz Inequality)**.** *Given any vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$, *we have that*

$$|\mathbf{x}\cdot\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|.$$

Consequently, the inverse cosine of $\mathbf{x}\cdot\mathbf{y}/\|\mathbf{x}\|\|\mathbf{y}\|$ *is well-defined, and Proposition 1.2.5 is valid.*

Proof. Clearly, if one of $\mathbf{x}$ or $\mathbf{y}$ is zero, then $\mathbf{x}\cdot\mathbf{y} = 0$ and $\|\mathbf{x}\|\|\mathbf{y}\| = 0$, hence the inequality holds. Consequently, we may assume that neither $\mathbf{x}$ nor $\mathbf{y}$ is zero so that $\mathbf{y}\cdot\mathbf{y}$ is nonzero by the Properties of the Dot Product in Real n-Space. Even more, for any real numbers α and β, we have that

$$\|\alpha\mathbf{x}+\beta\mathbf{y}\|^2 = (\alpha\mathbf{x}+\beta\mathbf{y})\cdot(\alpha\mathbf{x}+\beta\mathbf{y}) = \alpha^2(\mathbf{x}\cdot\mathbf{x}) + 2\alpha\beta(\mathbf{x}\cdot\mathbf{y}) + \beta^2(\mathbf{y}\cdot\mathbf{y})$$

is non-negative. By the above identity with $\alpha = \mathbf{y}\cdot\mathbf{y}$ and $\beta = -(\mathbf{x}\cdot\mathbf{y})$, we find that

$$(\mathbf{x}\cdot\mathbf{x})(\mathbf{y}\cdot\mathbf{y})^2 - (\mathbf{x}\cdot\mathbf{y})^2(\mathbf{y}\cdot\mathbf{y}) \geq 0.$$

Considering that $\mathbf{y}\cdot\mathbf{y}$ is nonzero by assumption, it must be a positive real number. Cancelling one factor of $\mathbf{y}\cdot\mathbf{y}$ from both sides of the above inequality yields that $(\mathbf{x}\cdot\mathbf{y})^2 \leq (\mathbf{x}\cdot\mathbf{x})(\mathbf{y}\cdot\mathbf{y})$. □

Theorem 1.2.14 (Triangle Inequality)**.** *Given any vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$, *we have that*

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

Proof. By Proposition 1.2.2, we have that $\|\mathbf{x}+\mathbf{y}\|$, $\|\mathbf{x}\|$, and $\|\mathbf{y}\|$ are each non-negative real numbers, hence the desired inequality holds if and only if the inequality $\|\mathbf{x}+\mathbf{y}\|^2 \leq (\|\mathbf{x}\|+\|\mathbf{y}\|)^2$ holds. By Proposition 1.2.8, the left-hand side of this inequality is $(\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y}) = \|\mathbf{x}\|^2 + 2(\mathbf{x}\cdot\mathbf{y}) + \|\mathbf{y}\|^2$. By the Cauchy-Schwarz Inequality, it follows that $2(\mathbf{x}\cdot\mathbf{y}) \leq 2\|\mathbf{x}\|\|\mathbf{y}\|$, hence the inequality holds.

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x}\cdot\mathbf{y}) + |\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\|+\|\mathbf{y}\|)^2$$ □

1.3 Matrices and Matrix Operations

We will continue to assume throughout this chapter that m and n are positive integers. We refer to a visual representation of any collection of data arranged into m rows and n columns as an $m \times n$ **array**. Each entry of an $m \times n$ array A is a **component** 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 that lies in the ith row and jth column. Often, we will refer to a_{ij} as the (i,j)th **entry** of the array A. Collectively, therefore, we may view the array A as **indexed** by its entries 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 each of the following examples.

Example 1.3.1. 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 A.

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

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

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

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

Example 1.3.3. Consider once again the scenario of Example 1.3.1. 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 1.3.1 yield new matrices that we could call "skill matrices" as follows.

$$S = \begin{bmatrix} 88 & 72 \\ 95 & 90 \end{bmatrix} \qquad S^T = \begin{bmatrix} 88 & 95 \\ 72 & 90 \end{bmatrix}$$

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

We refer to matrices with only one row as **row vectors**; matrices with only one column are called **column vectors**. We are familiar with some notion of vectors from our study of real n-space in Section 1.1. We may also use the terms (horizontal) n-**tuples** for row vectors with n columns (i.e., $1 \times n$ matrices) and (vertical) m-tuples for column vectors with m rows (i.e., $m \times 1$ matrices).

Like we mentioned in the first paragraph of this section, an $m \times n$ matrix A is uniquely determined by the entry a_{ij} in its ith row and jth column for each pair of integers $1 \leq i \leq m$ and $1 \leq j \leq n$. For instance, the matrix of Example 1.3.4 is the unique matrix with one row whose jth column consists of the integer j for each integer $1 \leq j \leq 5$. Under this identification, we will adopt the one-line notation $A = \big[a_{ij}\big]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ for the $m \times n$ matrix A with a_{ij} in its ith row and jth column.

Example 1.3.5. Consider the 2×3 matrix whose ith row and jth column consists of the sum $i + j$. We may write this symbolically (in one-line notation) as $\big[i + j\big]_{\substack{1 \leq i \leq 2 \\ 1 \leq j \leq 3}}$ or expanded as follows.

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

Example 1.3.6. 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** $O_{m \times n}$. Explicitly, we have the following examples.

$$O_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad O_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad O_{3\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad O_{3\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Often, it is most convenient to simply write O for the zero matrix with the understanding that the number of rows and columns of O is contingent upon the context in which it is discussed.

Example 1.3.7. We refer to the matrix $I_{m \times n} = \big[\delta_{ij}\big]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ 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.

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

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. Like with the zero matrix, we will write I for the square identity matrix of the appropriate size.

Example 1.3.8. Given any $m \times n$ matrix $A = \big[a_{ij}\big]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, 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 $A^T = \big[a_{ji}\big]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$. Put another way, the (i, j)th entry of A^T is the (j, i)th entry of A, hence the ith row of A^T is precisely the ith column of A. Explicitly, for the matrix A defined in Example 1.3.5, we have the following.

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix} \qquad A^T = \begin{bmatrix} 2 & 3 \\ 3 & 4 \\ 4 & 5 \end{bmatrix}$$

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 $I_{m\times n}^T = I_{n\times m}$, hence we have that $I_{n\times n}^T = I_{n\times n}$ or $I^T = I$.

Definition 1.3.9. 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 1.3.10. Given any $m \times n$ matrices $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$ and $B = \begin{bmatrix} b_{ij} \end{bmatrix}_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$, the **matrix sum** of A and B is the $m \times n$ matrix $A+B = \begin{bmatrix} a_{ij} + b_{ij} \end{bmatrix}_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$. Put in words, the matrix sum $A+B$ is the $m \times n$ matrix whose (i, j)th entry is the sum of the (i, j)th entries of A and B.

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

Example 1.3.11. 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 1.3.12. If A is any $m \times n$ matrix, then we have that $A + O_{m\times n} = A = O_{m\times n} + A$. Consequently, we may view $O_{m\times n}$ as the **additive identity** among all $m \times n$ matrices.

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

Example 1.3.13. Given any $m \times n$ matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$, we will write $-A = \begin{bmatrix} -a_{ij} \end{bmatrix}_{\substack{1\leq i\leq m \\ 1\leq j\leq n}}$. We have that $A + (-A) = O_{m\times n} = -A + A$, and we say that $-A$ is the **additive inverse** of A.

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

Proposition 1.3.14. *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 1.3.10, 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 1.3.8, 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 $(A+B)^T = A^T + B^T$. $\square$

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 the product of the matrices A and B is defined as follows.

Definition 1.3.15. Given any $m \times n$ matrix $A = \big[a_{ij}\big]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ and any $n \times r$ matrix $B = \big[a_{ij}\big]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq r}}$, the (left) **matrix product** of A and B is the $m \times r$ matrix AB whose (i, j)th entry is given by

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}.$$

Put in words, the matrix product AB is the $m \times r$ matrix whose (i, j)th entry is the sum of the product 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, matrix multiplication is not commutative, i.e., the order of the matrices 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 1.3.16. Consider the following real matrices.

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

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.

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

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

On the other hand, neither of the matrix products AC or BD exist; however, the matrices CA and DB can be computed because A and B have as many rows as C and D have columns, respectively.

Example 1.3.17. Consider the following real matrices.

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

Considering that A and B are both 2×2 matrices, the 2×2 matrices AB and BA can be formed.

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

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

Crucially, we note that AB and BA are not equal as matrices, i.e., we have that $AB \neq BA$.

Remark 1.3.18. Example 1.3.16 motivates the following definition of matrix multiplication. Consider a $1 \times n$ row vector $\mathbf{v} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \end{bmatrix}$ and the following $n \times 1$ column vector.

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

We define the **vector dot product** $\mathbf{v} \cdot \mathbf{w}$ of the vectors $\mathbf{v}$ and $\mathbf{w}$ as the 1×1 matrix $\mathbf{v}\mathbf{w}^T$, i.e.,

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

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

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

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

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

Example 1.3.19. 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 performing live 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 must be 10% certain that she will not attend. On the other hand, Bob is 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 show, 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.

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

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; the resulting model is referred to as a **Markov process**.

Remark 1.3.20. Example 1.3.19 illustrates that if $\mathbf{x}$ is an $n \times 1$ column vector and A is an $m \times n$ matrix, then the $m \times 1$ column vector $A\mathbf{x}$ is simply a linear combination of the columns of A.

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

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

Proposition 1.3.21 (Matrix Multiplication Is Associative). *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 equal.*

Proof. By Definition 1.3.15, we have that BC is an $n \times s$ matrix, hence the matrix product $A(BC)$ is well-defined because the number of columns of A is equal to the number of rows of BC; a similar argument shows that $(AB)C$ is well-defined, hence it suffices to prove that $A(BC) = (AB)C$. By the same definition, the (i, j)th entry of $A(BC)$ is the sum of the products of the (i, k)th entry of A and the (k, j)th entry of BC for all integers $1 \leq k \leq n$, and the (k, j)th entry of BC is the sum of the products of the (k, ℓ)th entry of B and the (ℓ, j)th entry of C for all integers $1 \leq \ell \leq r$. Put into symbols, the previous sentence can be expressed as the double summation identity

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

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

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

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

Proposition 1.3.22 (Matrix Multiplication Is Distributive). *If A is any $m \times n$ matrix and B and C are any $n \times r$ matrices, then $A(B + C) = AB + AC$ and $A(cB) = c(AB)$ for all scalars c.*

Proof. By Definition 1.3.10, 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 1.3.15, the (i, j)th entry of $A(B + C)$ is the sum of the products of the (i, k)th entry of A and the $(k, j)$the entry of BC for all integers $1 \leq k \leq n$; the latter is by Definition 1.3.10 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 1.3.21. $\square$

Ultimately, Proposition 1.3.22 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 1.3.23. Given any $n \times n$ matrix A, 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 1.3.21, 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 1.3.24. *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 1.3.8, the (i,j)th entry of $(AB)^T$ is the (j,i)th AB. By Definition 1.3.15, 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 conclude with a summary of the matrix operations proved in the previous propositions.

Proposition 1.3.25 (Properties of Matrix Addition, Multiplication, and Transposition)**.** *Consider any matrices A, B, and C such that the following matrix sums and matrix products are well-defined.*

1.) *Matrix addition is associative, i.e., $(A+B)+ = A+(B+C)$.*
2.) *Matrix addition is commutative, i.e., $A+B = B+A$.*
3.) *The zero matrix O is the additive identity, i.e., $A+O = A$.*
4.) *The additive inverse of A is $-A$, i.e., $A+(-A) = O$.*
5.) *Matrix multiplication is associative, i.e., $(AB)C = A(BC)$.*
6.) *Matrix multiplication is distributive, i.e., $A(B+C) = AB+AC$ and $(A+B)C = AC+BC$.*
7.) *The multiplicative identity is the identity matrix, i.e., $IA = A$ and $BI = B$.*
8.) *Matrix transposition is distributive across matrix addition, i.e., $(A+B)^T = A^T + B^T$.*
9.) *Matrix transposition is order-reversing, i.e., $(AB)^T = B^T A^T$.*
10.) *Scalar multiplication is associative, i.e., $r(sA) = (rs)A$.*
11.) *Scalar multiplication is distributive across matrix addition, i.e., $r(A+B) = rA+rB$.*
12.) *Scalar multiplication is distributive across scalar addition, i.e., $(r+s)A = rA+sA$.*
13.) *Scalar multiplication is homogeneous, i.e., $(rA)B = r(AB) = A(rB)$.*

1.4 Linear Systems of Equations and Gaussian Elimination

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

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

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

Example 1.4.1. On 10 June 2022, in Game Four of the 2022 NBA Finals, Stephen Curry scored 43 points. Let x_1 be the number of one-pointers made; let x_2 be the number of two-pointers made; and let x_3 be the number of three-pointers made by Curry in this appearance. Observe that Curry's point total is given by the 1×3 (integer) linear equation $x_1 + 2x_2 + 3x_3 = 43$.

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

Example 1.4.2. One can find many solutions to the matrix equation of Example 1.4.1. 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, we can reduce the number of possible solutions in Example 1.4.1. Using the fact that Curry made seven three-pointers, 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.

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

We may determine the values of x_1 and x_2 that solve the system: 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 $x_1 = 8$.

Examples 1.4.1 and 1.4.2 highlight the differences between the **general solution** of a system of linear equations as opposed to a **particular solution**. Explicitly, the 1×3 system of equations

$$x_1 + 2x_2 + 3x_3 = 43$$

admits infinitely many solutions: by solving this equation for x_1 in terms of x_2 and x_3, we find that $x_1 = -2x_2 - 3x_3 + 43$, hence the general solution to this system of equations is given by

$$\boldsymbol{\xi} = [-2x_2 - 3x_3 + 43, x_2, x_3] = x_2[-2, 1, 0] + x_3[-3, 0, 1] + [43, 0, 0].$$

Consequently, any choice of real numbers x_2 and x_3 determine a particular solution to this system of linear equations. We will soon revisit this distinction with more sophisticated tools.

Example 1.4.3. Geometrically, linear equations encode lines, planes, and hyperplanes. Explicitly, for any real numbers a and b (not both of which are zero) and any real number c, the solutions of the linear equation $ax + by = c$ form a line (e.g., $2x + y = 3$ is a line with y-intercept $(0, 3)$ and slope -2). Likewise, it is not difficult to verify that for any real numbers a, b, and c (not all of which are zero) and any real number d, the solutions of the linear equation $ax + by + cz = d$ form a plane: indeed, if a is nonzero, then solving for x in this linear equation yields that

$$x = -\frac{b}{a}y - \frac{c}{a}z + d,$$

hence (x, y, z) is a translation of a point lying in the plane spanned by $[0, 1, 0]$ and $[0, 0, 1]$.

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

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

Conversely, the aforementioned matrices A, $\mathbf{x}$, and $\mathbf{b}$ satisfy that $A\mathbf{x} = \mathbf{b}$. We refer to the equation $A\mathbf{x} = \mathbf{b}$ as a (real) $m \times n$ **matrix equation**. Often, the $m \times n$ matrix A and the $m \times 1$ matrix $\mathbf{b}$ are known while the $n \times 1$ matrix $\mathbf{x}$ 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.

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

Example 1.4.4. We will convert the data of Examples 1.4.1 and 1.4.2 into the language of matrix equations. Consider the matrix $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ whose jth column is the point value of a j-pointer; the matrix $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^T$ whose jth row is the number of j-pointers made by Curry; and the matrix $\mathbf{b} = \begin{bmatrix} 43 \end{bmatrix}$ consisting of the total points made by Curry. Observe that the linear equation $x_1 + 2x_2 + 3x_3 = 43$ is in one-to-one correspondence with the matrix equation $A\mathbf{x} = \mathbf{b}$.

We say that a (real) $n \times 1$ matrix $\boldsymbol{\xi}$ forms a **solution** to the matrix equation $A\mathbf{x} = \mathbf{b}$ if it holds that $A\boldsymbol{\xi} = \mathbf{b}$; this is a direct analog of a solution of the $m \times n$ system of linear equations.

Example 1.4.5. Rephrasing the results of 1.4.2, the matrices $\boldsymbol{\xi}_1 = \begin{bmatrix} 43 & 0 & 0 \end{bmatrix}$ and $\boldsymbol{\xi}_2 = \begin{bmatrix} 41 & 1 & 0 \end{bmatrix}$ give rise to two distinct solutions of the matrix equation of Example 1.4.4. On the other hand, put into the language of matrix equations, the information that $22 = x_1 + 2x_2$ and $15 = x_1 + x_2$ can be most efficiently synthesized by viewing the coefficients of these linear equations as rows of a matrix. Explicitly, we construct a matrix A whose first row is $\begin{bmatrix} 1 & 2 \end{bmatrix}$, corresponding to the respective coefficients of x_1 and x_2 in the equation $22 = x_1 + 2x_2$; the second row of the matrix A is $\begin{bmatrix} 1 & 1 \end{bmatrix}$, corresponding to the respective coefficients of x_1 and x_2 in the equation $15 = x_1 + x_2$. Once again, the column vector $\mathbf{x}$ consists of the variables x_1 and x_2 in distinct rows, and the column vector $\mathbf{b}$ consists of the integers 22 and 15 in distinct rows. Ultimately, yields the matrix equation

$$A\mathbf{x} = \mathbf{b} \text{ or } \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 22 \\ 15 \end{bmatrix}.$$

Once we have extracted an $m \times n$ matrix equation $A\mathbf{x} = \mathbf{b}$ from a (real) $m \times n$ system of linear equations, our immediate objective is to determine the matrix analog of solving the system. Before we do this, we declare the following three valid operations for systems of linear equations.

Definition 1.4.6 (Elementary Row Operations)**.** *Given any (real) $m \times n$ system of linear equations, the following arithmetic operations are permissible to perform on the system.*

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

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

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

Consequently, we are looking for matrix analogs of the above three arithmetic operations. Considering that the coefficients of ith equation are encoded in the ith row of the matrix A and the ith row of the matrix $\mathbf{b}$, we may rather consider the **augmented matrix** $\begin{bmatrix} A \mid \mathbf{b} \end{bmatrix}$. By definition, this is simply the matrix A with one additional column in the form of $\mathbf{b}$. We use the bar $\mid$ notation to emphasize that $\mathbf{b}$ is appended as the rightmost column of the matrix A and 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 (i, i)th entry c; 1 in all other entries of the main diagonal; and 0s elsewhere.

1.) Multiplication of the ith row of an $m \times n$ system of linear equations by a scalar c corresponds to left multiplication of the $m \times (n+1)$ augmented matrix $\begin{bmatrix} A \mid \mathbf{b} \end{bmatrix}$ 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 1.4.7. We obtain the following augmented matrix for the matrices of Example 1.4.5.

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

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

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

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

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

Example 1.4.8. Consider the augmented matrix $\left[A \mid \mathbf{b}\right]$ of Example 1.4.7. Observe that in order to subtract the first equation $x_1 + 2x_2 = 22$ from the second equation $x_1 + x_2 = 15$, it suffices to add -1 times the first equation to the second equation. By the previous observation, this can be achieved on the level of matrices by performing the following matrix multiplication.

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

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

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

Example 1.4.9. Once again, consider the augmented matrix $\left[A \mid \mathbf{b}\right]$ of Example 1.4.7. We may interchange the first equation $x_1 + 2x_2 = 22$ and the second equation $x_1 + x_2 = 15$ as follows.

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

Collectively, we refer to the operations of Definition 1.4.6 as **elementary row operations**; the matrices defined by operations (1.), (2.), and (3.) are therefore called the $m \times m$ **elementary row matrices**. Explicitly, an elementary row matrix is an $m \times m$ matrix obtain by from the $m \times m$ identity matrix I_m by (1.) multiplying any row of I_m by a nonzero scalar c; (2.) adding c times the ith row of I_m to the jth row of I_m; or (3.) interchanging rows i and j of I_m.

Likewise, the operations of Definition 1.4.6 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."

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 1.4.8.

Example 1.4.10. Consider the augmented matrix $\left[A \mid \mathbf{b}\right]$ of Example 1.4.8. 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 and only if A can be reduced to B via a sequence of elementary row operations if and only if 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. We will write $A \sim B$ if A and B are either row or column equivalent.

Example 1.4.11. By Example 1.4.8 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 1.4.10, 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: in this case, the row of B consisting of a single nonzero entry can be used to further reduce A to a matrix possessing more zero entries, as we illustrate next.

Example 1.4.12. Consider the row equivalent matrices A and B of Example 1.4.11. Observe that if we add twice the second row of B to the first row of B, then we obtain the matrix

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

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

We say that a row of an $m \times n$ matrix A is **nonzero** if it contains (at least) one nonzero entry.

Definition 1.4.13. We say that a (real) $m \times n$ 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 that A lies in row echelon form, the first nonzero entry of a nonzero row of A is a **pivot**.

Example 1.4.14. Consider the following real matrices.

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 4 \\ 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad C = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}$$

Both A and B lie in row echelon form; however, C does not lie in row echelon form because the first nonzero entry of its second row lies in the column directly below the first nonzero of its first row.

We have encountered other instances of matrices in row echelon form, as well: the matrices B of Example 1.4.11 and C of Example 1.4.12 lie in row echelon form; however, the matrix A of Example 1.4.11 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 1.4.15. *Every real matrix is row equivalent to a real matrix in row echelon form.*

Proof. Consider any 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, we may find the first nonzero row of A for which the entry in first column of A is nonzero. By interchanging this row with the first row of A, we may ultimately assume that our $m \times n$ matrix A has the form

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

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

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

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

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

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

Definition 1.4.16. We say that a matrix lies in **reduced row echelon form** if and only if

1.) it lies in row echelon form;

2.) its pivots are all 1; and

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

Corollary 1.4.17. *Every real matrix is row equivalent to a real matrix in reduced row echelon form.*

Proof. By Theorem 1.4.15, every real matrix A is row equivalent to a real matrix B in row echelon form. By multiplying each nonzero row of B by the multiplicative inverse of its pivot, we obtain a row equivalent matrix C whose pivots are all 1. Last, we must ensure that the only nonzero entry of any column containing a pivot is the pivot itself. Observe that if c_{ij} is nonzero and the jth column of C contains a pivot in row k, then we may add $-c_{ij}$ times the kth row of C to the ith row of C to obtain 0 in the ith row and jth column of C. Continuing in this manner yields the result. □

Essentially, the proofs of Theorem 1.4.15 and Corollary 1.4.17 outline the method of **Gaussian Elimination** in systems of linear equations; for completeness, we summarize the results below.

Algorithm 1.4.18 (Gaussian Elimination)**.** Given any nonzero real $m \times n$ matrix A, the following steps will reduce the matrix A to a row equivalent matrix B in reduced row echelon form.

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

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

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

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

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

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

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

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

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

One of the best ways to understand the method of Gaussian Elimination is to practice using it. We illustrate the technique and its applications in the following several examples.

Example 1.4.19. Let us convert the following matrix to reduced row echelon form.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Before we conclude this section, we provide two examples that illustrate how all of the topics we have discussed in this section come to bear on the theory of systems of linear equations.

Example 1.4.20. Consider the following real 3×4 system of linear equations.

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 3 \\ x_1 + 2x_3 + 3x_4 &= 4 \\ x_2 + x_4 &= 5 \end{aligned}$$

Converting this system of linear equations into a matrix equation by taking the coefficients of each linear equation as the entries of a 3×4 matrix A, expressing the variables $x_1, \dots, x_4$ as the rows of a 4×1 column vector, and writing the right-hand side as a 3×1 column vector yields the following.

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 3 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

Consequently, in order to solve this system of linear equations, it suffices to convert the following 3×5 augmented matrix into its reduced row echelon form by the method of Gaussian Elimination.

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 3 \\ 1 & 0 & 2 & 3 & 4 \\ 0 & 1 & 0 & 1 & 5 \end{array}\right] \overset{R_2-R_1 \mapsto R_2}{\sim} \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 3 \\ 0 & -1 & 1 & 2 & 1 \\ 0 & 1 & 0 & 1 & 5 \end{array}\right] \overset{R_2 \leftrightarrow R_3}{\sim} \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 3 \\ 0 & 1 & 0 & 1 & 5 \\ 0 & -1 & 1 & 2 & 1 \end{array}\right] \overset{R_3+R_2 \mapsto R_3}{\sim} \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 3 \\ 0 & 1 & 0 & 1 & 5 \\ 0 & 0 & 1 & 3 & 6 \end{array}\right]$$

$$\overset{R_1-R_3 \mapsto R_1}{\sim} \left[\begin{array}{cccc|c} 1 & 1 & 0 & -2 & -3 \\ 0 & 1 & 0 & 1 & 5 \\ 0 & 0 & 1 & 3 & 6 \end{array}\right]$$

$$\overset{R_1-R_2 \mapsto R_1}{\sim} \left[\begin{array}{cccc|c} 1 & 0 & 0 & -3 & -8 \\ 0 & 1 & 0 & 1 & 5 \\ 0 & 0 & 1 & 3 & 6 \end{array}\right]$$

Consequently, the 3×4 system is equivalent to the following system in reduced row echelon form.

$$\begin{aligned} x_1 - 3x_4 &= -8 \\ x_2 + x_4 &= 5 \\ x_3 + 3x_4 &= 6 \end{aligned}$$

We obtain the general solution of this system by expressing each of the three variables x_1, x_2, and x_3 in terms of the **free variable** x_4. Crucially, observe that the general solution is given by

$$\boldsymbol{\xi} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3x_4 - 8 \\ -x_4 + 5 \\ -3x_4 + 6 \\ x_4 \end{bmatrix} = x_4 \begin{bmatrix} 3 \\ -1 \\ -3 \\ 1 \end{bmatrix} + \begin{bmatrix} -8 \\ 5 \\ 6 \\ 1 \end{bmatrix}.$$

Consequently, for each assignment of a real number to the free variable x_4, we obtain a unique solution $\boldsymbol{\xi}$. Ultimately, this system of linear equations admits infinitely many solutions, and each solution is determined by the value of x_4 by the above equation. Below are two particular solutions.

$$\text{If } x_4 = 0, \text{ then the particular solution to the system is given by } \boldsymbol{\xi} = \begin{bmatrix} -8 \\ 5 \\ 6 \\ 1 \end{bmatrix}.$$

$$\text{If } x_4 = 3, \text{ then the particular solution to the system is given by } \boldsymbol{\xi} = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 3 \end{bmatrix}.$$

Considering that this system of linear equations admits a solution, we say the system is **consistent**.

Example 1.4.21. Consider the following real 4×3 system of linear equations.

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 0 \\ 4x_1 + 5x_2 + 6x_3 &= 1 \\ 7x_1 + 8x_2 + 9x_3 &= 0 \end{aligned}$$

We obtain an augmented matrix $\left[A \mid \mathbf{b}\right]$ called the **coefficient matrix** corresponding to this system of linear equations by writing down the coefficients of the variables. Each equation is a distinct row. Each variable induces a distinct column. Explicitly, we obtain the following coefficient matrix.

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 1 \\ 7 & 8 & 9 & 0 \end{array}\right]$$

We proceed to convert the matrix to reduced row echelon form via Gaussian Elimination.

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 1 \\ 7 & 8 & 9 & 0 \end{array}\right] \overset{\substack{R_2 - 4R_1 \mapsto R_2 \\ R_3 - 7R_1 \mapsto R_3}}{\sim} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -3 & -6 & 1 \\ 0 & -6 & -12 & 0 \end{array}\right] \overset{R_3 - 2R_2 \mapsto R_3}{\sim} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -3 & -6 & 1 \\ 0 & 0 & 0 & -2 \end{array}\right]$$

We note that from this step, it can be determined that this system of linear equations is **inconsistent**, i.e., it has no solution. Explicitly, observe that the third row of the above augmented matrix implies (on the level of linear equations) that $0 = 0x_1 + 0x_2 + 0x_3 = -2$ — a contradiction.

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -3 & -6 & 1 \\ 0 & 0 & 0 & -2 \end{array}\right] \overset{\substack{-\frac{1}{2}R_3 \mapsto R_3 \\ -\frac{1}{3}R_2 \mapsto R_2}}{\sim} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & -\frac{1}{3} \\ 0 & 0 & 0 & 1 \end{array}\right] \overset{R_2 + \frac{1}{3}R_3 \mapsto R_2}{\sim} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \overset{R_1 - 2R_2 \mapsto R_1}{\sim} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

1.5 Inverses of Square 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 $LA = I$, where we denote by I 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 $AR = I$. 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 $AB = I = BA$. Our next proposition shows that a two-sided inverse of a matrix A is unique.

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

Proof. Consider any $n \times n$ matrices L and R such that $LA = I = AR$. By Proposition 1.3.21, we have that $R = IR = (LA)R = L(AR) = LI = L$. Consequently, L is a two-sided inverse of A. Even more, if L' is any two-sided inverse of A, then it is a right inverse of A so that $L' = L$. □

Consequently, if an $n \times n$ matrix A admits a (two-sided) inverse, then it is unique, and we may denote it by A^{-1}. We will also say in this case that A is **invertible** (or **non-singular**). Certainly, the zero matrix does not possess an inverse, hence some (and in fact many) matrices are not invertible. We explore next how matrix inverses behave with respect to the matrix operations of Section 1.3.

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

Proof. By Proposition 1.3.24, it follows that $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$, and we conclude that $(A^T)^{-1} = (A^{-1})^T$ by the uniqueness of the matrix inverse guaranteed by Proposition 1.5.1. □

Proposition 1.5.3. *If $A_1, \dots, A_k$ are any invertible $n \times n$ matrices, then*

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

Put another way, the product of invertible $n \times n$ matrices is an invertible matrix, and the matrix inverse of the product is the product of the matrix inverses in reverse order.

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

Corollary 1.5.4. *If A is an invertible $n \times n$ matrix, then A^k is invertible for all integers $k \geq 0$.*

Proof. By Proposition 1.5.3, it follows that A^k is invertible with $(A^k)^{-1} = (A^{-1})^k$. □

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, an $n \times n$ matrix A is row equivalent to the matrix B if and only if there exist elementary row matrices $E_1, \dots, E_k$ such that $B = E_k \cdots E_1 A$. Considering that $A = E_1^{-1} \cdots E_k^{-1} B$, we conclude that A is invertible if and only if B is invertible by Propositions 1.5.3 and 1.5.6. □

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. Explicitly, we have that A and B are row equivalent if and only if there exist elementary row matrices $E_1, \dots, E_k$ such that $B = E_k \cdots E_1 A$. Elementary row matrices are precisely those $n \times n$ matrices obtained from the $n \times n$ identity matrix by performing (at most) one of the following matrix operations.

1.) We may multiply any row of I by a nonzero scalar c.

2.) We may add c times the ith row of I to the jth row of I.

3.) We may interchange any pair of rows i and j of I.

We refer to the above operations as the Elementary Row Operations.

Proposition 1.5.6. *Every elementary row matrix is invertible.*

Proof. Let E be an elementary row matrix. Consider the following three cases.

1.) If E is obtained from I by multiplying the ith row of I by a nonzero scalar c, then E^{-1} is obtained from I by multiplying the ith row of I by the nonzero scalar c^{-1}.

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

3.) If E is obtained from I by interchanging rows i and j of I, then E is its own inverse. $\square$

Before we provide several equivalent criteria for the invertibility of a square matrix or establish how to compute a matrix inverse, it is imperative to discuss how theory of systems of linear equations comes to bear on the theory of invertible matrices. Consider the matrix equation $A\mathbf{x} = \mathbf{b}$ for some real $n \times n$ matrix A, the real $n \times 1$ column vector $\mathbf{x}$ whose ith row is a variable x_i, and some real $n \times 1$ column vector $\mathbf{b}$. Crucially, we note that if A is row equivalent to the $n \times n$ identity matrix I, then the matrix equation $A\mathbf{x} = \mathbf{b}$ is consistent (i.e., it admits a solution): indeed, if there exist elementary row matrices $E_1, \dots, E_k$ such that $E_k \cdots E_1 A = I$, then we have that

$$\mathbf{x} = I\mathbf{x} = (E_k \cdots E_1 A)\mathbf{x} = E_k \cdots E_1 (A\mathbf{x}) = E_k \cdots E_1 \mathbf{b}.$$

Conversely, if the matrix equation $A\mathbf{x} = \mathbf{b}$ admits a solution, then A must be row equivalent to the identity matrix. We establish this as follows using a proof by contrapositive.

Theorem 1.5.7. *Given any real $n \times n$ matrix A, the matrix equation $A\mathbf{x} = \mathbf{b}$ admits a solution for every real $n \times 1$ matrix $\mathbf{b}$ if and only if A is row equivalent to the $n \times n$ identity matrix.*

Proof. By the paragraph preceding the statement of the theorem, if A is row equivalent to the $n \times n$ identity matrix, then the matrix equation $A\mathbf{x} = \mathbf{b}$ admits a unique solution for every $n \times 1$ matrix $\mathbf{b}$. Conversely, we will assume that A is not row equivalent to the $n \times n$ identity matrix. Consequently, the nth row of the reduced row echelon form $\operatorname{RREF}(A)$ of the matrix A must be zero. Even more, there exist elementary row matrices $E_1, \dots, E_k$ such that $\operatorname{RREF}(A) = E_k \cdots E_1 A$. By Proposition

1.5.6, each of the $n \times n$ matrices $E_1, \dots, E_k$ is invertible, hence their product $E_k \cdots E_1$ is invertible by Proposition 1.5.3. Consider the real $n \times 1$ matrix $\mathbf{b} = (E_k \dots E_1)^{-1}\mathbf{e}_n$ for the nth standard basis vector $\mathbf{e}_n$ that consists of zeros in each of the first $n-1$ rows and 1 in the nth row. Observe that the matrix equation $A\mathbf{x} = \mathbf{b}$ has no solution: indeed, by construction, we have that

$$\text{RREF}(A)\mathbf{x} = (E_k \cdots E_1 A)\mathbf{x} = E_k \cdots E_1(A\mathbf{x}) = E_k \cdots E_1 \mathbf{b} = \mathbf{e}_n.$$

Considering that the nth row of $\text{RREF}(A)\mathbf{x}$ is 0 and the nth row of $\mathbf{e}_n$ is 1, we have established that it is impossible to obtain a real $n \times 1$ matrix $\mathbf{x}$ for which the matrix equation $A\mathbf{x} = \mathbf{b}$ holds. □

By virtue of Theorem 1.5.7, it follows that any left inverse of an $n \times n$ matrix must be a right inverse, as well. Consequently, the invertibility of a square matrix can be determined by checking whether the matrix can be reduced to the identity matrix. Even more, the unique matrix inverse of a matrix that is row equivalent to the identity matrix is simply the product of the elementary matrices required to put the matrix in reduced row echelon form. We prove this as follows.

Theorem 1.5.8. *Given any $n \times n$ matrices A and B, we have that $AB = I$ if and only if $BA = I$. Explicitly, any left inverse of a square matrix is the unique inverse of the matrix.*

Proof. We will assume first that $AB = I$, and we will demonstrate that $BA = I$. Conversely, we may simply reverse the roles of A and B to find that if $BA = I$, then $AB = I$. Given any $n \times 1$ matrix $\mathbf{b}$, the matrix equation $A\mathbf{x} = \mathbf{b}$ admits a solution $\boldsymbol{\xi} = B\mathbf{b}$: indeed, we have that

$$A\boldsymbol{\xi} = A(B\mathbf{b}) = (AB)\mathbf{b} = I\mathbf{b} = \mathbf{b}.$$

By Theorem 1.5.7, it follows that A is row equivalent to the $n \times n$ identity matrix, hence there exist elementary row matrices $E_1, \dots, E_k$ such that $E_k \cdots E_1 A = I$. By Proposition 1.5.1, in view of the fact that $E_k \cdots E_1$ is a left inverse of A, it follows that $E_k \cdots E_1$ is the unique inverse of A. □

Conversely, we demonstrate that every invertible matrix is row equivalent to the identity matrix. By Corollary 1.4.17, a matrix is row equivalent to its reduced row echelon form. By Corollary 1.5.5, an $n \times n$ matrix 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 the underlying matrix A cannot be invertible. On the other hand, we will establish that if all rows of $\text{RREF}(A)$ 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 the (i, j)th component of the matrix is zero for all integers $1 \leq i < j \leq n$. Put another way, all entries below the main diagonal of an upper-triangular matrix are zero.

Theorem 1.5.9. *Every upper-triangular matrix with nonzero diagonal elements is invertible.*

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

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

By hypothesis that a_{ii} is nonzero for each integer $1 \leq i \leq n$, we may multiply the ith row of the above matrix by a_{ii}^{-1} to obtain an upper-triangular matrix whose pivots are all 1. Each of these products corresponds to multiplication of U (on the left) by an elementary row matrix, hence this process does not come to bear on the existence of an inverse of U. Consequently, we may assume from the beginning that this is the case, i.e., we may restrict our attention to the following situation.

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

By Corollary 1.5.5, it suffices to demonstrate that U is row equivalent to the invertible $n \times n$ identity matrix I. We achieve this by furnishing some elementary row operations that reduces U to I. Observe that if we add $-a_{in}$ times the last row of U to the ith row of U, then we obtain a 0 in the (i, n)th component of the resulting matrix. Continuing in this way, we may reduce the nth column of U to zero except in the bottom right-hand corner. Considering that adding any scalar multiple of a row of U to another row of U is a row equivalence, we conclude that U is row equivalent to this matrix. Continuing in this way for each column of U from right to left, it follows that U is row equivalent to the identity matrix. By Theorem 1.5.8, we conclude that U is invertible. □

Corollary 1.5.10 (Invertibility Criterion)**.** *Given any $n \times n$ matrix A, we have that A is invertible if and only if it is row equivalent to the $n \times n$ identity matrix.*

Proof. By Theorems 1.5.7 and 1.5.8, a matrix that is row equivalent to the identity matrix must be invertible. Conversely, by Proposition 1.5.5, Theorem 1.5.9, and the paragraph that precedes the theorem, an $n \times n$ matrix A is invertible if and only if the upper-triangular matrix $\operatorname{RREF}(A)$ is invertible if and only if $\operatorname{RREF}(A) = I$. Consequently, if A is an invertible $n \times n$ matrix, then there exist elementary row matrices $E_1, \dots, E_k$ such that $E_k \cdots E_1 A = I$, from which we conclude by Theorem 1.5.7 that the unique inverse of A is given by $A^{-1} = E_k \cdots E_1$. □

Corollary 1.5.11. *Every invertible $n \times n$ matrix is a product of elementary row matrices.*

Proof. By the proof of Corollary 1.5.10, every invertible $n \times n$ matrix A admits some elementary row matrices $E_1, \dots, E_k$ such that $E_k \cdots E_1 A = I$. 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 1.5.6, 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. □

Generally, the method of Gaussian Elimination can in practice be implemented to determine if a square matrix is invertible and to explicitly produce the inverse of such a matrix. Observe that if A is an $n \times n$ matrix, then we may construct the augmented matrix $\left[A \mid I\right]$ by adjoining the $n \times n$ identity matrix I on the right-hand side of A. By performing elementary row operations, we may reduce A to its reduced row echelon form $\operatorname{RREF}(A)$. Consequently, if A is invertible, this will reduce A to I and simultaneously convert I to A^{-1}. Explicitly, this process yields that $\left[A \mid I\right] \sim \left[I \mid A^{-1}\right]$.

Example 1.5.12. Consider the following 2×2 matrix A and the augmented matrix $\left[A \mid I\right]$.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \end{bmatrix} \qquad \left[A \mid I\right] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 5 & 0 & 1 \end{array}\right]$$

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

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 5 & 0 & 1 \end{array}\right] \overset{R_2 - 3R_1 \mapsto R_2}{\sim} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -1 & -3 & 1 \end{array}\right] \overset{-R_2 \mapsto R_2}{\sim} \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & 3 & -1 \end{array}\right] \overset{R_1 - 2R_2 \mapsto R_1}{\sim} \left[\begin{array}{cc|cc} 1 & 0 & -5 & 2 \\ 0 & 1 & 3 & -1 \end{array}\right]$$

Consequently, we find that A is an invertible 2×2 matrix with the following matrix inverse.

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

Example 1.5.13. Consider the following 3×3 matrix A and the augmented matrix $\left[A \mid I\right]$.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix} \qquad \left[A \mid I\right] = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{array}\right]$$

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

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

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

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

By the paragraph preceding Example 1.5.12, we conclude that the inverse of A is given as follows.

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

Example 1.5.14. Let us determine a numerical criterion for which a real 2×2 matrix is invertible by performing Gaussian Elimination to obtain the reduced row echelon form. Consider any 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.

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

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

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

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

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

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

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

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

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

Equivalently, the displayed matrix above is $E_3 E_2 E_1 A$ 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 $E_3 E_2 E_1 A$ on the left by

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

then we obtain $E_4 E_3 E_2 E_1 A = I_{2 \times 2}$ so that $A^{-1} = E_4 E_3 E_2 E_1$. 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.

1.6 Real Vector Subspaces and Bases

Consider any real $m \times n$ matrix A, any real $n \times 1$ matrix $\mathbf{x}$, and any real $m \times 1$ matrix $\mathbf{b}$. We say that the matrix equation $A\mathbf{x} = \mathbf{b}$ is **homogeneous** if $\mathbf{b} = \mathbf{0}$; otherwise, this matrix equation is **non-homogeneous**. Every homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$ is consistent: indeed, the zero vector $\mathbf{0}$ satisfies that $A\mathbf{0} = \mathbf{0}$, hence it is a **trivial solution** of the matrix equation. Elsewhere, the nonzero solutions of the matrix equation $A\mathbf{x} = \mathbf{0}$ are referred to as **non-trivial solutions**. Crucially, any linear combination of non-trivial solutions of a homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$ forms a solution of the matrix equation: for if $A\boldsymbol{\xi}_1 = \mathbf{0}$ and $A\boldsymbol{\xi}_2 = \mathbf{0}$, then for any real numbers α_1 and α_2, it follows by the Distributive Law for Matrix Multiplication (Proposition 1.3.22) that

$$A(\alpha_1\boldsymbol{\xi}_1 + \alpha_2\boldsymbol{\xi}_2) = \alpha_1(A\boldsymbol{\xi}_1) + \alpha_2(A\boldsymbol{\xi}_2) = \alpha_1\mathbf{0} + \alpha_2\mathbf{0} = \mathbf{0}.$$

We summarize the above exposition in the following proposition.

Proposition 1.6.1. *Every linear combination of solutions of a homogeneous matrix equation constitutes a solution of the matrix equation. Explicitly, if $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are solutions to the matrix equation $A\mathbf{x} = \mathbf{0}$, then for any real numbers α_1 and α_2, it follows that $\alpha_1\boldsymbol{\xi}_1 + \alpha_2\boldsymbol{\xi}_2$ is another solution.*

Collections of vectors with the property that any linear combination of vectors in the collection is itself a member of the collection form an important class of objects in linear algebra.

Definition 1.6.2. Given any nonempty collection W of vectors in real n-space, we will say that W forms a **subspace** of real n-space if both of the following conditions hold.

1.) We have that W is **closed under addition**, i.e., $\mathbf{v} + \mathbf{w}$ lies in W if $\mathbf{v}$ and $\mathbf{w}$ lie in W.

2.) We have that W is **closed under scalar multiplication**, i.e., $\alpha\mathbf{v}$ lies in W if $\mathbf{v}$ lies in W.

Example 1.6.3. One can readily verify that the **zero subspace** $\{\mathbf{0}\}$ is a subspace of real n-space and the totality of real n-space $\mathbb{R}^n$ is a subspace of real n-space. We refer to these as the **trivial subspaces** of real n-space because they represent most extreme subspaces of real n-space.

Example 1.6.4. By Proposition 1.6.1, for any real $m \times n$ matrix A, the collection of solutions of the homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$ constitutes a subspace of real n-space. We refer to this as the **null space** of the matrix A, and we denote this by $\operatorname{null}(A) = \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \mathbf{0}\}$.

Example 1.6.5. Consider the collection $W = \{[x, -x] \mid x \in \mathbb{R}\}$ of vectors in real 2-space such that the second coordinate is equal in absolute value to the first coordinate but opposite in sign.

1.) Given any real numbers x and y, observe that

$$[x, -x] + [y, -y] = [x + y, -x - y] = [x + y, -(x + y)].$$

Consequently, it follows that W is closed under addition.

2.) Given any real numbers x and α, observe that

$$\alpha[x, -x] = [\alpha x, \alpha(-x)] = [\alpha x, -(\alpha x)].$$

Consequently, it follows that W is closed under scalar multiplication.

By Definition 1.6.2, we conclude that W is a subspace of real 2-space.

Example 1.6.6. Consider the collection $W = \{[x, 2x - 3] \mid x \in \mathbb{R}\}$ of vectors in real 2-space lying in standard position and coinciding with the line $y = 2x - 3$. Observe that the vectors $[0, -3]$ and $[1, -1]$ lie in W because they terminate at a point on the line $y = 2x - 3$; however, their sum $[1, -4]$ does not lie in W because it is not true that $y = 2x - 3$, so W is not a subspace of real 2-space.

Even more, our next proposition illustrates that the span of any collection of vectors in real n-space forms a subspace of real n-space. We may then view Example 1.6.5 as a special case of this: indeed, the real vectors $[x, -x] = x[1, -1]$ are precisely the vectors in $\operatorname{span}\{[1, -1]\}$.

Proposition 1.6.7 (Subspace Property of Span)**.** *Given any vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ in real n-space, we have that $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is a subspace of real n-space.*

Proof. Conventionally, the span of the empty set is the zero vector, hence the span of no vectors in real n-space is the zero subspace. Consequently, we may assume that $n \geq 1$. By definition of the span of vectors, every vector of W is of the form $\mathbf{u} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_k \mathbf{v}_k$ for some real numbers $\alpha_1, \alpha_2, \dots, \alpha_k$, hence for any real number α, we have that

$$\alpha \mathbf{u} = \alpha(\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_k \mathbf{v}_k) = (\alpha\alpha_1)\mathbf{v}_1 + (\alpha\alpha_2)\mathbf{v}_2 + \cdots + (\alpha\alpha_k)\mathbf{v}_k$$

lies in W so that W is closed under scalar multiplication. Likewise, for any vector of W of the form $\mathbf{w} = \beta_1 \mathbf{v}_1 + \beta_2 \mathbf{v}_2 + \cdots + \beta_k \mathbf{v}_k$, the vector $\mathbf{u} + \mathbf{w}$ satisfies that

$$\mathbf{u} + \mathbf{w} = (\alpha_1 + \beta_1)\mathbf{v}_1 + (\alpha_2 + \beta_2)\mathbf{v}_2 + \cdots + (\alpha_k + \beta_k)\mathbf{v}_k,$$

hence W is closed under addition. We conclude by Definition 1.6.2 that W is a subspace of $\mathbb{R}^n$. $\square$

Every subspace of real n-space can in fact be realized as the span of some vectors in real n-space (or possibly none at all). We will not prove this yet, but we use it as our guide in the following.

Example 1.6.8. Compute the null space of the following real matrix.

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 8 & 7 & 6 \\ 5 & 4 & 3 & 2 \end{bmatrix}$$

By definition, the null space of the above matrix consists of all vectors $\mathbf{x} = [x_1, x_2, x_3, x_4]$ in real 4-space satisfying the homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$, hence $\operatorname{null}(A)$ consists of all solutions of the following homogeneous real system of linear equations.

$$\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 &= 0 \\ 5x_1 + 6x_2 + 7x_3 + 8x_4 &= 0 \\ 9x_1 + 8x_2 + 7x_3 + 6x_4 &= 0 \\ 5x_1 + 4x_2 + 3x_3 + 2x_4 &= 0 \end{aligned}$$

Bearing this in mind, the usefulness of the null space is apparent. Computing the reduced row echelon form of the matrix A using Gaussian Elimination allows us to determine the null space as follows. Crucially, we need not consider the augmented matrix $\left[A \mid \mathbf{0}\right]$ because the last column of this augmented matrix consists of zeros and remains unaffected by elementary row operations.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 8 & 7 & 6 \\ 5 & 4 & 3 & 2 \end{bmatrix} \overset{\substack{R_2 - 5R_1 \mapsto R_2 \\ R_3 - 9R_1 \mapsto R_3 \\ R_4 - 5R_1 \mapsto R_4}}{\sim} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \\ 0 & -10 & -20 & -30 \\ 0 & -6 & -12 & -18 \end{bmatrix} \overset{\substack{-\frac{1}{4}R_2 \mapsto R_2 \\ -\frac{1}{10}R_3 \mapsto R_3 \\ -\frac{1}{6}R_4 \mapsto R_4}}{\sim} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 \end{bmatrix} \overset{\substack{R_3 - R_2 \mapsto R_3 \\ R_4 - R_2 \mapsto R_4}}{\sim} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We are now one elementary row operation from the reduced row echelon form of the matrix A.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \overset{R_1 - 2R_2 \mapsto R_1}{\sim} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Considering this matrix in the context of the underlying system of equations, it follows that x_3 and x_4 are free variables with which we can express x_1 and x_2 in terms of x_3 and x_4 as follows.

$$\begin{aligned} x_1 - x_3 - 2x_4 &= 0 \\ x_2 + 2x_3 + 3x_4 &= 0 \end{aligned}$$

By solving these equations in terms of x_3 and x_4, it follows that $x_1 = x_3 + 2x_4$ and $x_2 = -2x_3 - 3x_4$. Last, expressing these solutions as column vectors, we obtain the null space as a span of two vectors.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_3 + 2x_4 \\ -2x_3 - 3x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_3 \\ -2x_3 \\ x_3 \\ 0 \end{bmatrix} + \begin{bmatrix} 2x_4 \\ -3x_4 \\ 0 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$$

Consequently, it follows that $\operatorname{null}(A) = \operatorname{span}\{[1, -2, 1, 0]^T, [2, -3, 0, 1]^T\}$.

Even more, by the Subspace Property of Span, every real $m \times n$ matrix A induces two more subspaces of real n-space: the **row space** of A is the subspace $\operatorname{row}(A)$ spanned by the rows of the matrix A. Likewise, the **column space** of A is the subspace $\operatorname{col}(A)$ spanned by the columns of A.

Example 1.6.9. Consider the following real 2×2 matrix and its reduced row echelon form.

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

By definition, the row space of A is spanned by the rows of A; however, both of the rows of A are the same vector, hence we have that $\operatorname{row}(A) = \operatorname{span}\{[1, 0]\}$. On the other hand, the column space of A is spanned by the columns; however, since there is a zero column of A and the zero vector never contributes to the span, we have that $\operatorname{col}(A) = \operatorname{span}\{[1, 1]^T\}$. Considering the reduced row echelon form of A in the context of the homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$, it follows that $x_1 = 0$ and x_2 is a free variable. Consequently, the null space of A consists of all real vectors of the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

We conclude therefore that $\operatorname{null}(A) = \operatorname{span}\{[0, 1]^T\}$.

Example 1.6.10. Consider the real 3×3 identity matrix below.

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

Crucially, the rows and columns of I are the standard basis vectors $\mathbf{e}_1 = [1, 0, 0]$, $\mathbf{e}_2 = [0, 1, 0]$, and $\mathbf{e}_3 = [0, 0, 1]$ of real 3-space, hence we have that $\operatorname{row}(I) = \operatorname{col}(I) = \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\} = \mathbb{R}^3$. On the other hand, we have that $I\mathbf{v} = \mathbf{0}$ if and only if $\mathbf{v} = \mathbf{0}$ so that $\operatorname{null}(I) = \{\mathbf{0}\}$.

Example 1.6.11. Consider the following real 3×4 matrix and its reduced row echelon form.

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

By definition, the row space and column space of A are given as follows.

$$\operatorname{row}(A) = \operatorname{span}\{[1,-1,-1,0],[1,1,-1,0],[1,1,1,6]\}$$
$$\operatorname{col}(A) = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}\right\}$$

Considering the reduced row echelon form of A in the context of the homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$, we obtain the following homogeneous system of linear equations with free variable x_4.

$$\begin{aligned} x_1 + 3x_4 &= 0 \\ x_2 &= 0 \\ x_3 + 3x_4 &= 0 \end{aligned}$$

Consequently, the null space of A consists of all real vectors of the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -3x_4 \\ 0 \\ -3x_4 \\ x_4 \end{bmatrix} = x_4 \begin{bmatrix} -3 \\ 0 \\ -3 \\ 1 \end{bmatrix}.$$

We conclude therefore that $\operatorname{null}(A) = \operatorname{span}\{[-3,0,-3,1]^T\}$.

Generally, if a collection of vectors W in real n-space is the span of some vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$, we might naturally seek to determine if every vector in W can be written uniquely as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$. Consider the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ of real n-space: every vector in real n-space can be written uniquely as a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ since

$$\mathbf{x} = [x_1, x_2, \dots, x_n] = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$$

and the coordinates x_i uniquely determine the vector $\mathbf{x}$. Generalizing to $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$, if the coefficients $\alpha_1, \alpha_2, \dots, \alpha_k$ of a vector $\mathbf{w} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k$ in W are unique, then we say that the coordinates of $\mathbf{w}$ are $(\alpha_1, \alpha_2, \dots, \alpha_k)$ with respect to $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$, and we refer to the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ as a **basis** of the subspace W of real n-space. We point out that the terminology of "standard basis vectors" of real n-space falls under this broader definition.

Consequently, in order to determine if a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ in real n-space form a basis for $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$, it suffices to explore the notion of "uniqueness" of the coefficients of a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$. We will say that the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are **linearly independent** if and only if $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$ implies that $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$, i.e., the only linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ that is the zero vector is the linear combination with all coefficients of zero. Conversely, if there exist scalars $\alpha_1, \dots, \alpha_n$ not all of which are zero such that $\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_k \mathbf{v}_k = \mathbf{0}$, then we say that $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are **linearly dependent**. Observe that in this case, there exists a nonzero scalar α_i such that $\alpha_i \mathbf{v}_i = -\alpha_1 \mathbf{v}_1 - \alpha_2 \mathbf{v}_2 - \cdots - \alpha_k \mathbf{v}_k$ and $\mathbf{v}_i = -\alpha_1 \alpha_i^{-1} \mathbf{v}_1 - \alpha_2 \alpha_i^{-1} \mathbf{v}_2 - \cdots - \alpha_k \alpha_i^{-1} \mathbf{v}_k$, i.e., the vector $\mathbf{v}_i$ can be written as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ excluding $\mathbf{v}_i$. Consequently, any collection of vectors including the zero vector $\mathbf{0}$ is linearly dependent, and we may restrict our attention to nonzero vectors.

Example 1.6.12. We outline a method for determining the linear independence of vectors as follows. Consider the vectors $\mathbf{v} = [1, 1]$ and $\mathbf{w} = [-3, 2]$ of real 2-space. By definition, $\mathbf{v}$ and $\mathbf{w}$ are linearly independent if and only if $\alpha \mathbf{v} + \beta \mathbf{w} = \mathbf{0}$ implies that $\alpha = \beta = 0$. Expanding this equation by componentwise addition, 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 homogeneous 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 $\mathbf{v}$ and $\mathbf{w}$; 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 $\mathbf{0}$ is written as a column vector equal to this matrix product. Consequently, if the matrix whose columns are $\mathbf{v}$ and $\mathbf{w}$ is row equivalent to the 2×2 identity matrix I, then it will follow that $\alpha = \beta = 0$, i.e., $\mathbf{v}$ and $\mathbf{w}$ will be linearly independent. By the method of Gaussian Elimination, we obtain the unique reduced row echelon form as follows.

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

We conclude therefore that $\mathbf{v} = [1, 1]$ and $\mathbf{w} = [-3, 2]$ are linearly independent.

Our previous example gives rise to the following general method for determining all linearly independent vectors among a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ in real n-space.

Algorithm 1.6.13 (Linear Independence Algorithm)**.** Consider any collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ in real n-space. Carry out the following steps to find a (not necessarily unique) collection of linearly independent vectors of largest size among the vectors collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$. (Generally, the vectors produced by this algorithm will depend on the order of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$.)

(1.) Construct the real $m \times n$ matrix A whose jth column is the $m \times 1$ column vector $\mathbf{v}_j$.

(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 vector that is linearly independent from all other vectors. Every column that does not possess a pivot corresponds to a vector that can be written as a nonzero linear combination of some vectors.

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

Example 1.6.14. We will use the Linear Independence Algorithm to determine the linearly independent vectors among the vectors $\mathbf{v}_1 = [1, 1, 1]$, $\mathbf{v}_2 = [-1, 1, 1]$, $\mathbf{v}_3 = [-1, -1, 1]$, and $\mathbf{v}_4 = [0, 0, 6]$. We must construct the 3×4 matrix whose jth column is $\mathbf{v}_j^T$; then, we must subsequently convert this matrix into its unique reduced row echelon form. We illustrate this process this as follows.

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

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

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

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

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

Consequently, the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent and $\mathbf{v}_4 = 3\mathbf{v}_1 + 0\mathbf{v}_2 + 3\mathbf{v}_3$.

Example 1.6.15. We will demonstrate that the real vectors $\mathbf{v}_1 = [1, 2, 3, 4]$, $\mathbf{v}_2 = [5, 6, 7, 8]$, and $\mathbf{v}_3 = [6, 8, 10, 12]$ are not linearly independent.

Other benefits of the Linear Independence Algorithm include its indispensable utility in determining linearly independent spanning sets for subspaces of real n-space.

Definition 1.6.16. Given any subspace W of real n-space, we will say that a collection of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ forms a **basis** for W if both of the following two conditions hold.

1.) We have that $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$.

2.) We have that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent.

Example 1.6.17. Consider the real vectors $\mathbf{v} = [1, 1]$ and $\mathbf{w} = [-3, 2]$ of Example 1.6.12. We have already demonstrated that these vectors are linearly independent, hence in order to conclude that they form a basis for real 2-space, it suffices to prove that they span real 2-space. We will achieve this by finding the coordinates α and β of any vector $[a, b]$ with respect to $\mathbf{v}$ and $\mathbf{w}$. 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 1.6.12 exhibits elementary row operations to convert the matrix on the left to reduced row echelon form; to find α and β, we carry out these operations on the following augmented matrix.

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

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

Example 1.6.18. Consider the real vectors $\mathbf{v}_1 = [2, 1, -3]$, $\mathbf{v}_2 = [4, 0, 2]$, and $\mathbf{v}_3 = [2, -1, 3]$. We wish to determine if these vectors form a basis for the subspace of real 3-space that they span. We achieve this by carrying out the steps outlined in the Linear Independence Algorithm.

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

(1.) We employed the elementary row operation $R_1 \leftrightarrow R_2$.

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

(3.) We employed the elementary row operations $R_2 \leftrightarrow R_3$, $\frac{1}{2}R_2 \mapsto R_2$, and $\frac{1}{4}R_3 \mapsto R_3$.

(4.) We employed the elementary row operation $R_3 - R_2 \mapsto R_3$.

Considering that each column of the row echelon form of the above matrix admits a pivot, it follows that the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent, hence they form a basis for the subspace they span. Even more, we claim that this subspace is indeed the totality of real 3-space. Carrying out one final elementary row operation puts the matrix in reduced row echelon form.

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{(5.)}{\sim} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(5.) We employed the elementary row operation $R_1 + R_3 \mapsto R_1$.

Consequently, the matrix A whose columns are the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ is invertible. Explicitly, if we perform these elementary row operations on the 3×3 identity matrix, we will obtain A^{-1}.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{(1.)}{\sim} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \overset{(2.)}{\sim} \begin{bmatrix} 0 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 3 & 1 \end{bmatrix} \overset{(3.)}{\sim} \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & 0 \end{bmatrix} \overset{(4.)}{\sim} \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ \frac{1}{4} & -2 & -\frac{1}{2} \end{bmatrix} \overset{(5.)}{\sim} \begin{bmatrix} \frac{1}{4} & -1 & -\frac{1}{2} \\ 0 & \frac{3}{2} & \frac{1}{2} \\ \frac{1}{4} & -2 & -\frac{1}{2} \end{bmatrix}$$

Given any vector $\mathbf{x}$ in real 3-space, it follows that $\mathbf{x} = I\mathbf{x} = (AA^{-1})\mathbf{x} = A(A^{-1}\mathbf{x})$. Considering that $A(A^{-1}\mathbf{x})$ is a linear combination of the columns of A, we conclude that $\mathbf{x}$ lies in the span of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$; therefore, since $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent, every vector in real 3-space can be written as a unique linear combination of these vectors, so they form a basis for real 3-space.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \left(\frac{1}{4}x_1 - x_2 - \frac{1}{2}x_3\right) \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} + \left(\frac{3}{2}x_2 + \frac{1}{2}x_3\right) \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix} + \left(\frac{1}{4}x_1 - 2x_2 - \frac{1}{2}x_3\right) \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

Both of the previous examples are indicative of a general phenomenon that neatly relates many of the concepts that we have studied in this chapter. We conclude this section with a discussion of the connections between homogeneous systems of linear equations and matrix equations.

Theorem 1.6.19 (Fundamental Theorem of Linear Systems of Equations)**.** *Consider any consistent $m \times n$ system of linear equations with coefficient matrix A, indeterminate matrix $\mathbf{x}$, and target matrix $\mathbf{b}$. If $n > m$, the system admits infinitely many solutions; otherwise, the following are equivalent.*

1.) *We can obtain a unique solution for the matrix equation* $A\mathbf{x} = \mathbf{b}$.

2.) *We can row reduce* A *to the* $n \times n$ *identity matrix followed by* $m - n$ *rows of zero.*

3.) *We can obtain a basis for the column space of* A *via the columns of* A.

Theorem 1.6.20 (Fundamental Theorem of Consistent Linear Systems of Equations)**.** *Consider any consistent* $m \times n$ *system of linear equations with coefficient matrix* A, *indeterminate matrix* $\mathbf{x}$, *and target matrix* $\mathbf{b}$. *Given any solution* $\boldsymbol{\xi}$ *of the matrix equation* $A\mathbf{x} = \mathbf{b}$ *and any solution* $\boldsymbol{\eta}$ *of the homogeneous matrix equation* $A\mathbf{x} = \mathbf{0}$, *we have that* $\boldsymbol{\xi} + \boldsymbol{\eta}$ *is a solution of the matrix equation* $A\mathbf{x} = \mathbf{b}$. *Even more, every solution of the matrix equation* $A\mathbf{x} = \mathbf{b}$ *is of the form* $\boldsymbol{\xi} + \boldsymbol{\eta}$.

Proof. Certainly, if $A\boldsymbol{\xi} = \mathbf{b}$ and $A\boldsymbol{\eta} = \mathbf{0}$, then by the Distributive Law, we have that

$$A(\boldsymbol{\xi} + \boldsymbol{\eta}) = A\boldsymbol{\xi} + A\boldsymbol{\eta} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

Conversely, suppose that $\mathbf{x}$ satisfies that $A\mathbf{x} = \mathbf{b}$. We seek vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ such that $A\boldsymbol{\xi} = \mathbf{b}$ and $A\boldsymbol{\eta} = \mathbf{0}$ and $\mathbf{x} = \boldsymbol{\xi} + \boldsymbol{\eta}$. By assumption that $A\mathbf{x} = \mathbf{b}$ is consistent, there exists a vector $\boldsymbol{\xi}$ other than $\mathbf{x}$ such that $A\boldsymbol{\xi} = \mathbf{b}$ and $A(\mathbf{x} - \boldsymbol{\xi}) = A\mathbf{x} - A\boldsymbol{\xi} = \mathbf{b} - \mathbf{b} = \mathbf{0}$. We conclude that $\mathbf{x} - \boldsymbol{\xi}$ lies in the null space of A, hence there exists a vector $\boldsymbol{\eta}$ such that $A\boldsymbol{\eta} = \mathbf{0}$ and $\mathbf{x} - \boldsymbol{\xi} = \boldsymbol{\eta}$, as desired. □

1.7 Linear Independence and Dimension

We have as yet discussed many of the concepts that we seek to employ in this section. Considering their paramount importance, we recall several of these facts below in preparation for what follows. We will assume to this end that $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are vectors in real n-space.

a.) We say that the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ satisfy a **dependence relation** if there exist scalars $\alpha_1, \alpha_2, \dots, \alpha_k$ not all of which are zero such that $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$. Given that this is the case, the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are said to be **linearly dependent**; otherwise, these vectors are called **linearly independent**. Consequently, any collection of vectors that contains the zero vector $\mathbf{0}$ is linearly dependent: indeed, we may obtain a dependence relation by taking the coefficients of all nonzero vectors as 0 and the coefficient of $\mathbf{0}$ as nonzero.

b.) We say that $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ form a **basis** for a subspace W of real n-space provided that

1.) we have that $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ and

2.) the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are linearly independent.

Consequently, a basis for a subspace is a linearly independent system of generators.

c.) Given any subspace W of real n-space that is spanned by $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$, one may carry out the Linear Independence Algorithm in order to determine a basis for W. Crucially, it is precisely the pivots of the row echelon form of the matrix whose columns are the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ that correspond to the linearly independent vectors among $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$.

Our first objective in this section is to demonstrate that if the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ form a basis for a subspace W of real n-space, then the non-negative integer k is unique. We refer to this number as the **dimension** of W, and we write in this case that $\dim(W) = k$. Essentially, this fact follows as a corollary of the following proposition that states that if some nonzero vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ span W, then any collection of linearly independent vectors consists of no more than k vectors.

Proposition 1.7.1. *Given any nonzero vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ in real n-space, consider the subspace $W = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$. Every collection of $\ell > k$ vectors $\mathbf{w}_1, \dots, \mathbf{w}_\ell$ in W is linearly dependent.*

Proof. Considering that $W = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$, for every collection of nonzero vectors $\mathbf{w}_1, \dots, \mathbf{w}_\ell$ in W, there exist scalars $\alpha_{11}, \dots, \alpha_{1k}, \dots, \alpha_{\ell 1}, \dots, \alpha_{\ell k}$ such that the following equations hold.

$$\begin{aligned} \mathbf{w}_1 &= \alpha_{11}\mathbf{v}_1 + \cdots + \alpha_{1k}\mathbf{v}_k \\ &\vdots \\ \mathbf{w}_\ell &= \alpha_{\ell 1}\mathbf{v}_1 + \cdots + \alpha_{\ell k}\mathbf{v}_k \end{aligned}$$

Consider the $\ell \times k$ 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 $\ell > k$, 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 ℓ). Consequently, there exist scalars $\beta_1, \dots, \beta_\ell$ such that $\beta_1\mathbf{w}_1 + \cdots + \beta_\ell\mathbf{w}_\ell = \mathbf{0}$. □

Corollary 1.7.2. *Given any pair of bases $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ and $\{\mathbf{w}_1, \dots, \mathbf{w}_\ell\}$ of any subspace W of real n-space, we must have that $k = \ell$. Consequently, the dimension $\dim(W)$ of W is well-defined.*

Proof. By Proposition 1.7.1, we must have that $\ell \leq k$ because W is spanned by $\mathbf{v}_1, \dots, \mathbf{v}_k$ and $\mathbf{w}_1, \dots, \mathbf{w}_\ell$ are linearly independent. Conversely, we must have that $k \leq \ell$ because W is spanned by $\mathbf{w}_1, \dots, \mathbf{w}_\ell$ and $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent. We conclude that $k = \ell$, as desired. □

Definition 1.7.3. Given any subspace W of real n-space, the unique number of elements in a basis for W is the **dimension** of W; this number is a non-negative integer denoted by $\dim(W)$.

Example 1.7.4. Consider the zero subspace $\{\mathbf{0}\}$ consisting only of the zero vector $\mathbf{0}$. observe that there are no linearly independent vectors in this subspace, hence its dimension is zero. Even more, this is the only dimension zero subspace of real n-space: for any subspace W that contains a nonzero vector contains a linearly independent vector, so there must be at least one vector in a basis for W.

Example 1.7.5. Considering that the totality of real n-space $\mathbb{R}^n$ is spanned by the linearly independent vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ by the exposition following Example 1.6.11, we find that $\dim(\mathbb{R}^n) = n$. Crucially, we note that the dimension of real n-space as a vector space is equal to the intuitive dimension of real n-space. Explicitly, we exist in real 3-space, and we perceive real 4-space through the passage of time. We can move in three directions (east-west, north-south, up-down), so it makes sense that the dimension of any space in which there are n directions we can move must be n.

Example 1.7.6. By the Linear Independence Algorithm, we can obtain a basis from any collection of vectors that span a subspace of real n-space by reducing to a collection of linearly independent vectors. Explicitly, consider the subspace W of real 4-space spanned by the following vectors.

$$\mathbf{v}_1 = [1, 2, 3, 4] \qquad \mathbf{v}_3 = [3, 2, 1, 0]$$
$$\mathbf{v}_2 = [2, 2, 2, 2] \qquad \mathbf{v}_4 = [4, 3, 2, 1]$$

(1.) We employed elementary row operations $R_2 - 2R_1 \mapsto R_2$, $R_3 - 3R_1 \mapsto R_3$, and $R_4 - 4R_1 \mapsto R_1$.

(2.) We employed elementary row operations $R_3 - 2R_2 \mapsto R_3$ and $R_4 - 3R_2 \mapsto R_4$.

We proceed to row reduce the real 4×4 matrix whose jth column is the vector $\mathbf{v}_j$.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 2 & 2 & 3 \\ 3 & 2 & 1 & 2 \\ 4 & 2 & 0 & 1 \end{bmatrix} \overset{(1.)}{\sim} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -5 \\ 0 & -4 & -8 & -10 \\ 0 & -6 & -12 & -15 \end{bmatrix} \overset{(2.)}{\sim} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -2 & -4 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Considering that the first two columns of the row echelon form of the above matrix contain pivots, we conclude that $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent vectors that span W. Consequently, it follows that $W = \operatorname{span}\{[1, 2, 3, 4], [2, 2, 2, 2]\}$, $\dim(W) = 2$, and $\{[1, 2, 3, 4], [2, 2, 2, 2]\}$ is a basis for W.

Algorithm 1.7.7 (Basis Algorithm)**.** Given any real vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ of real n-space that span a subspace $W = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$, carry out the following algorithm to determine a basis for W.

1.) Construct the real $n \times k$ matrix A whose jth column is $\mathbf{v}_j$.

2.) Use the method of Gaussian Elimination to obtain the row echelon form of A.

3.) Each column that contains a pivot corresponds to a basis vector in the following sense: if each of the columns $i_1, i_2, \dots, i_\ell$ contains a pivot, then it follows that $W = \operatorname{span}\{\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \dots, \mathbf{v}_{i_\ell}\}$, $\dim(W) = \ell$, and $\{\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \dots, \mathbf{v}_{i_\ell}\}$ constitutes a basis for W.

Caution: one must use the columns of the original matrix A; neither the rows of the matrix A nor the columns of the row echelon form of A is guaranteed to correspond to a basis for W.

Consequently, the Basis Algorithm yields a systematic method to reduce any spanning set of a subspace W of real n-space to a basis for W. Conversely, it is always possible to extend any collection of linearly independent vectors in real n-space to a basis for real n-space as follows.

Proposition 1.7.8. *Consider any linearly independent vectors* $\mathbf{v}_1, \dots, \mathbf{v}_k$ *lying in a subspace* W *of real* n*-space with the property that* $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}$ *are linearly dependent for all vectors* $\mathbf{w}$ *in* W*. We must have that* $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ *constitutes a basis for* W*. Put another way, the largest number of linearly independent vectors in a subspace* W *of real* n*-space is the dimension* $\dim(W)$ *of* W*.*

Proof. By definition of a basis, it suffices to demonstrate that $\mathbf{v}_1, \dots, \mathbf{v}_k$ span W. Given any vector $\mathbf{w}$ in W, there exist scalars $\alpha_1, \dots, \alpha_n, \alpha$ not all of which are zero with $\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_k + \alpha w = \mathbf{0}$ by hypothesis that $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}$ are linearly dependent. On the other hand, the linear independence of $\mathbf{v}_1, \dots, \mathbf{v}_k$ implies that if $\alpha = 0$, then $\alpha_1 = \cdots = \alpha_k = 0$. Consequently, we must have that α is nonzero so that $\mathbf{w} = \alpha_1\alpha^{-1}\mathbf{v}_1 + \cdots + \alpha_n\alpha^{-1}\mathbf{v}_k$. We conclude that $W = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$. $\square$

Corollary 1.7.9. *Given any linearly independent vectors* $\mathbf{v}_1, \dots, \mathbf{v}_k$ *lying in a subspace* W *of real* n*-space such that* $\dim(W)$ *is finite, there exist nonzero vectors* $\mathbf{v}_{k+1}, \dots, \mathbf{v}_\ell$ *in* W *such that* $\{\mathbf{v}_1, \dots, \mathbf{v}_\ell\}$ *constitutes a basis for* W*. Put another way, every linearly independent collection of vectors lying in a nonzero subspace* W *of real* n*-space can be extended (or enlarged) to a basis for* W*.*

Proof. Begin with a collection of linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$. By Proposition 1.7.8, if $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}$ are linearly dependent for all vectors $\mathbf{w}$ in W, then $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ constitutes a basis for W; otherwise, there exists a nonzero vector $\mathbf{v}_{k+1}$ in W such that $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$ are linearly independent. Continuing in this manner yields nonzero vectors $\mathbf{v}_{k+1}, \dots, \mathbf{v}_\ell$ in W such that $\mathbf{v}_1, \dots, \mathbf{v}_\ell$ are linearly independent and $\mathbf{v}_1, \dots, \mathbf{v}_\ell, \mathbf{w}$ are linearly dependent for all vectors $\mathbf{w}$ in W by Proposition 1.7.1. Consequently, Proposition 1.7.8 assures that $\mathbf{v}_1, \dots, \mathbf{v}_\ell$ form a basis for W, as desired. □

We prove at last that every subspace of real n-space has finite dimension, hence every subspace of real n-space admits a basis. Bearing this in mind, it follows immediately that every subspace of real n-space can be realized as the span of finitely many vectors in real n-space. Geometrically, therefore, the following theorem ensures that every nonzero subspace of real n-space is a hyperplane.

Theorem 1.7.10. *Every subspace of real n-space has finite dimension. Explicitly, if W is a subspace of real n-space, then we must have that $0 \leq \dim(W) \leq n$. Consequently, every subspace of real n-space can be realized as the span of finitely many linearly independent vectors in real n-space.*

Proof. By Proposition 1.7.8, we have that $\dim(W) = 0$ if and only if W contains no linearly independent vectors if and only if W contains no nonzero vectors if and only if $W = \{\mathbf{0}\}$. Consequently, it suffices to establish that $1 \leq \dim(W) \leq n$ for every nonzero subspace W of real n-space. Begin with a nonzero vector $\mathbf{v}_1$ in W. By Proposition 1.7.8, if $\mathbf{v}_1$ and $\mathbf{w}$ are linearly dependent for every vector $\mathbf{w}$ in W, then $\mathbf{v}_1$ forms a basis for W; otherwise, there exists a nonzero vector $\mathbf{v}_2$ in W such that $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. Continuing in this manner yields nonzero vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ in W such that $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent and $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}$ are linearly dependent for all vectors $\mathbf{w}$ in W. Explicitly, by viewing the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}$ as vectors in real n-space, we may appeal to Proposition 1.7.1 because $\mathbb{R}^n$ has dimension n. Consequently, we conclude by Proposition 1.7.8 that the linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ form a basis for W and $\dim(W) = k$. Even more, we must have that $k \leq n$ by Proposition 1.7.1. Last, if $\dim(W) = n$, then a basis for W must be a basis for $\mathbb{R}^n$. Explicitly, if there were a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of W that were not a basis for $\mathbb{R}^n$, then there would exist a vector $\mathbf{v}$ in $\mathbb{R}^n$ that were not a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_n$, i.e., the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n, \mathbf{v}$ would be linearly independent. But this contradicts Proposition 1.7.8. □

Considering that the preceding four statements are so important, we collect them below.

Theorem 1.7.11 (Fundamental Theorem of Subspaces of Real n-Space)**.** *Consider any subspace W of real n-space. Each of the following statements regarding W holds.*

1.) *We may realize W as the span of some linearly independent vectors, i.e., W admits a basis.*

2.) *We have that $0 \leq \dim(W) \leq n$. Even more, we have that $\dim(W) = 0$ if and only if $W = \{\mathbf{0}\}$ and $\dim(W) = n$ if and only if $W = \mathbb{R}^n$.*

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

4.) *Every collection of linearly independent vectors of W can be enlarged to a basis for W.*

5.) *Every collection of $\dim(W)$ vectors that span W constitutes a basis for V.*

6.) *Every collection of $\dim(W)$ linearly independent vectors of W constitutes a basis for W.*

Example 1.7.12. Considering that $\dim(\mathbb{R}^5)$, every collection of fewer than five vectors in $\mathbb{R}^5$ can be enlarged to basis for $\mathbb{R}^5$ by the Fundamental Theorem of Subspaces of Real n-Space. We illustrate this fact as follows. Consider the following three vectors in real 5-space.

$$\mathbf{v}_1 = [1,2,3,4,5] \qquad \mathbf{v}_2 = [9,8,7,6,5] \qquad \mathbf{v}_3 = [1,0,1,0,1]$$

We will construct a basis for $\mathbb{R}^5$ that includes the above three vectors. We begin by fortifying the collections of vectors with the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_5$. By the Basis Algorithm, we will next construct the real 5×8 matrix whose columns are the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_5$; the pivots of the row echelon form of this matrix correspond to the vectors in a basis for $\mathbb{R}^5$.

$$\begin{bmatrix} 1 & 9 & 1 & 1 & 0 & 0 & 0 & 0 \\ 2 & 8 & 0 & 0 & 1 & 0 & 0 & 0 \\ 3 & 7 & 1 & 0 & 0 & 1 & 0 & 0 \\ 4 & 6 & 0 & 0 & 0 & 0 & 1 & 0 \\ 5 & 5 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \overset{(1.)}{\sim} \begin{bmatrix} 1 & 9 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -10 & -2 & -2 & 1 & 0 & 0 & 0 \\ 0 & -20 & -2 & -3 & 0 & 1 & 0 & 0 \\ 0 & -30 & -4 & -4 & 0 & 0 & 1 & 0 \\ 0 & -40 & -4 & -5 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\overset{(2.)}{\sim} \begin{bmatrix} 1 & 9 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -10 & -2 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 & -3 & 0 & 1 & 0 \\ 0 & 0 & 4 & 3 & -4 & 0 & 0 & 1 \end{bmatrix}$$

$$\overset{(3.)}{\sim} \begin{bmatrix} 1 & 9 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -10 & -2 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -2 & 0 & 1 \end{bmatrix}$$

$$\overset{(4.)}{\sim} \begin{bmatrix} 1 & 9 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -10 & -2 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix}$$

1.) We employed the elementary row operations $R_i - iR_1 \mapsto R_i$ for each integer $2 \leq i \leq 5$.

2.) We employed the elementary row operations $R_i - iR_2 \mapsto R_i$ for each integer $3 \leq i \leq 5$.

3.) We employed the elementary row operations $R_4 - R_3 \mapsto R_4$ and $R_5 - 2R_3 \mapsto R_5$.

4.) We employed the elementary row operation $R_5 - R_4 \mapsto R_5$.

Considering that the first five columns of the row echelon form of the above matrix contain pivots, we conclude that $\{[1,2,3,4,5], [9,8,7,6,5], [1,0,1,0,1], [1,0,0,0,0], [0,1,0,0,0]\}$ is a basis for $\mathbb{R}^5$.

1.8 Rank of a Matrix

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 $\mathbf{c}_1, \dots, \mathbf{c}_n$ are the $m \times 1$ column vectors such that $\mathbf{c}_j$ corresponds to the jth column of A. By definition, the span of these column vectors is the collection of all possible linear combinations of the vectors $\mathbf{c}_1, \dots, \mathbf{c}_n$, i.e., we have that $\operatorname{span}\{\mathbf{c}_1, \dots, \mathbf{c}_n\} = \{\alpha_1 \mathbf{c}_1 + \cdots + \alpha_n \mathbf{c}_n \mid \alpha_1, \dots, \alpha_n \text{ are scalars}\}$. We refer to the vector space $\operatorname{span}\{\mathbf{c}_1, \dots, \mathbf{c}_n\}$ as the **column space** of A; its dimension 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 $A\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is any $n \times 1$ vector of the form $[\alpha_1, \dots, \alpha_n]^T$. Explicitly, we have that $A\boldsymbol{\alpha} = \alpha_1 \mathbf{c}_1 + \cdots + \alpha_n \mathbf{c}_n$, as illustrated in Remark 1.3.20.

Example 1.8.1. Considering that the columns of the real 3×3 identity matrix I are simply the real 3×1 vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ with $\mathbf{e}_1 = [1, 0, 0]$, $\mathbf{e}_2 = [0, 1, 0]$, and $\mathbf{e}_3$ written as row vectors, it follows that the column space of the 3×3 identity matrix I consists of all real vectors of the form $\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{e}_3 = [\alpha_1, \alpha_2, \alpha_3]$ such that α_1, α_2, and α_3 are real numbers. By the exposition following Example 1.6.11, this forms the totality of real 3-space, hence the column space of the 3×3 identity matrix is $\mathbb{R}^3$. Considering that $\dim(\mathbb{R}^3) = 3$ by Example 1.7.5, the column rank of I is 3. One can readily extend this argument to see that the column rank of the $n \times n$ identity matrix is n.

Example 1.8.2. Consider the following real 2×2 matrix.

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

By definition, the column space of A consists of all possible linear combinations of the columns of A, hence every vector in the column space of A is of the form

$$\mathbf{v} = \alpha_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \alpha_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

for some real number α_1. Consequently, the column space of A is simply the span of the nonzero vector $[1, 1]^T$, hence the column rank of A is one. Observe that the reduced row echelon form

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

of A has column space spanned by the nonzero vector $[1, 0]^T$, hence its column rank is also one.

Example 1.8.3. Consider the real 3×4 matrix of Example 1.6.14 in reduced row echelon form.

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

Previously, we illustrated that the column vectors $[1, 1, 1]^T$, $[-1, 1, 1]^T$, and $[-1, -1, 1]^T$ are linearly independent. Considering that $\mathbb{R}^3$ has dimension three by Example 1.7.5, we conclude by Theorem 1.7.11(6.) that these vectors form a basis for $\mathbb{R}^3$, hence they form a basis for the column space of A. Consequently, the column rank of A is three. Likewise, the column rank of $\operatorname{RREF}(A)$ is three by the same rationale because the vectors $[1, 0, 0]^T$, $[0, 1, 0]^T$, and $[0, 0, 1]^T$ are linearly independent.

Each of the previous three examples exhibit matrices whose column rank coincides with the column rank of its reduced row echelon form. We prove next that this is no coincidence: in fact, the column rank of a matrix is always equal to the column rank of its reduced row echelon form.

Proposition 1.8.4. *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 $\mathbf{A}_1, \dots, \mathbf{A}_n$ and $\mathbf{R}_1, \dots, \mathbf{R}_n$ 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 $\mathbf{R}_j = E\mathbf{A}_j$ for each integer $1 \leq j \leq n$. Observe that if there exist scalars $c_1, \dots, c_n$ such that $c_1\mathbf{R}_1 + \cdots + c_n\mathbf{R}_n = \mathbf{0}$, then multiplying both sides of this vector equation on the left by E yields that $c_1\mathbf{A}_1 + \cdots + c_n\mathbf{A}_n = \mathbf{0}$. Conversely, if there exist scalars $d_1, \dots, d_n$ such that $d_1\mathbf{A}_1 + \cdots + d_n\mathbf{A}_n = \mathbf{0}$, then multiplying both sides of this vector equation on the left by E^{-1} yields that $d_1\mathbf{R}_1 + \cdots + d_n\mathbf{R}_n = \mathbf{0}$. We conclude therefore that the columns $\mathbf{A}_{i_1}, \dots, \mathbf{A}_{i_k}$ of A are linearly independent if and only if the columns $\mathbf{R}_{i_1}, \dots, \mathbf{R}_{i_k}$ are linearly independent. By Proposition 1.7.8 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 $\mathbf{r}_1, \dots, \mathbf{r}_m$ of an $m \times n$ matrix A, i.e., the $1 \times n$ row vectors $\mathbf{r}_i$ corresponding to the ith row of A. We define the **row rank** of A to be the dimension of the **row space** of A, i.e., the vector space $\operatorname{span}\{\mathbf{r}_1, \dots, \mathbf{r}_m\} = \{\alpha_1\mathbf{r}_1 + \cdots + \alpha_m\mathbf{r}_m \mid \alpha_1, \dots, \alpha_m \text{ are scalars}\}$.

Example 1.8.5. Like before, the rows of the real 3×3 identity matrix I are the linearly independent real 3×1 vectors $\mathbf{e}_1 = [1, 0, 0]$, $\mathbf{e}_2 = [0, 1, 0]$, and $\mathbf{e}_3 = [0, 0, 1]$; these vectors span the totality of real 3-space $\mathbb{R}^3$, so the row space of I is $\mathbb{R}^3$ and the row rank of the 3×3 identity matrix is $\dim(\mathbb{R}^3) = 3$. Once again, the same argument shows that the row rank of the $n \times n$ identity matrix is n.

Example 1.8.6. Consider the following real 2×2 matrix of Example 1.8.2.

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

By definition, the row space of A consists of all possible linear combinations of the rows of A, hence every vector in the row space of A is of the form

$$\mathbf{v} = \alpha_1[1, 0] + \alpha_2[1, 0] = (\alpha_1 + \alpha_2)[1, 0]$$

for some real numbers α_1 and α_2. Considering that every real number can be written as the sum of two real numbers (take one of them as zero), it follows that the row space of A is the span of the nonzero vector $[1, 0]$, hence the row rank of A is one. Observe that the reduced row echelon form

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

of A has the same row space. Consequently, we find that A and $\operatorname{RREF}(A)$ have the same row rank.

Example 1.8.7. Consider the real 3×4 matrix of Example 1.8.3 and its reduced row echelon form.

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

Consider the row vectors $\mathbf{r}_1 = [1, -1, -1, 0]$, $\mathbf{r}_2 = [1, 1, -1, 0]$, and $\mathbf{r}_3 = [1, 1, 1, 6]$. Certainly, the vector $\mathbf{r}_3$ is linearly independent from the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ because it has a nonzero entry in its fourth column, and the fourth column of $\mathbf{r}_1$ and $\mathbf{r}_2$ is zero. Likewise, the vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are linearly independent: indeed, if we take any scalars α_1 and α_2 such that $\alpha_1\mathbf{r}_1 + \alpha_2\mathbf{r}_2 = \mathbf{0}$, then it follows that $(\alpha_1, -\alpha_1, -\alpha_1, 0) + (\alpha_2, \alpha_2, -\alpha_2, 0) = (0, 0, 0, 0)$ so that $\alpha_1 + \alpha_2 = 0$ and $-\alpha_1 + \alpha_2 = 0$. By adding the first equation to the second, we find that $2\alpha_2 = 0$ or $\alpha_2 = 0$, from which it follows that $\alpha_1 = 0$. Ultimately, we conclude that the row rank of A is three, and the row space of A is

$$\operatorname{span}\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3\} = \{[\alpha_1 + \alpha_2 + \alpha_3, -\alpha_1 + \alpha_2 + \alpha_3, -\alpha_1 - \alpha_2 + \alpha_3, 6\alpha_3] \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}\}.$$

Likewise, the row rank of $\operatorname{RREF}(A)$ is three because the vectors $\mathbf{r}_1 = [1, 0, 0, 3]$, $\mathbf{r}_2 = [0, 1, 0, 0]$, and $\mathbf{r}_3[0, 0, 1, 3]$ are linearly independent: indeed, we have that $\alpha_1\mathbf{r}_1 + \alpha_2\mathbf{r}_2 + \alpha_3\mathbf{r}_3 = \mathbf{0}$ if and only if $(\alpha_1, \alpha_2, \alpha_3, 3\alpha_1 + 3\alpha_3) = (0, 0, 0, 0)$ if and only $\alpha_1 = \alpha_2 = \alpha_3 = 0$. Last, the row space of $\operatorname{RREF}(A)$ consists of all vectors of the form $[\alpha_1, \alpha_2, \alpha_3, 3\alpha_1 + 3\alpha_3]$ for some real numbers α_1, α_2, and α_3.

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 1.8.8. *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 $\mathbf{A}_1, \dots, \mathbf{A}_m$ and $\mathbf{R}_1, \dots, \mathbf{R}_m$ denote the rows of A and R, respectively. Certainly, it does not affect the row space of A to interchange any number of rows of A because this amounts to relabelling the indices of some row vectors $\mathbf{A}_i$ and $\mathbf{A}_j$, and the indices of the vectors in a span do not matter by definition. Likewise, taking any nonzero scalar multiple c of any row $\mathbf{A}_i$ of A does not affect the span of $\mathbf{A}_1, \dots, \mathbf{A}_m$ because any vector $c_1\mathbf{A}_1 + \cdots + c_m\mathbf{A}_m$ in the span of $\mathbf{A}_1, \dots, \mathbf{A}_m$ is now given by $c_1\mathbf{A}_1 + \cdots + (c_i c^{-1})c\mathbf{A}_i + \cdots + c_m\mathbf{A}_m$. Last, replacing any row $\mathbf{A}_j$ of A by the linear combination $c\mathbf{A}_i + \mathbf{A}_j$ for any scalar c and any integer $1 \leq i \leq m$ does not affect the span of $\mathbf{A}_1, \dots, \mathbf{A}_m$ because any vector $c_1\mathbf{A}_1 + \cdots + c_m\mathbf{A}_m$ in the span of $\mathbf{A}_1, \dots, \mathbf{A}_m$ can be realized as the linear combination

$$c_1\mathbf{A}_1 + \cdots + (c_i - c_j c)\mathbf{A}_i + \cdots + c_j(c\mathbf{A}_i + \mathbf{A}_j) + \cdots + c_m\mathbf{A}_m.$$

Consequently, every vector in the span of $\mathbf{A}_1, \dots, \mathbf{A}_m$ lies in the span of $\mathbf{R}_1, \dots, \mathbf{R}_m$. Conversely, every row of R is a linear combination of some rows of A, hence every vector in the span of $\mathbf{R}_1, \dots, \mathbf{R}_m$ lies in the span of $\mathbf{A}_1, \dots, \mathbf{A}_m$. We conclude that $\operatorname{span}\{\mathbf{A}_1, \dots, \mathbf{A}_m\} = \operatorname{span}\{\mathbf{R}_1, \dots, \mathbf{R}_m\}$, i.e., the row spaces of A and R are equal; thus, the row rank of A and the row rank of R are equal. □

Corollary 1.8.9. *Elementary column operations do not affect column rank.*

Proposition 1.8.10. *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; thus, according to Proposition 1.8.4, elementary row operations on the matrix transpose do not affect the column rank of the matrix transpose, so elementary column operations do not affect the row rank of the matrix. □

Proposition 1.8.11. *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 1.8.8, the row rank of R is equal to the row rank of A, hence there are precisely r nonzero rows of R, where r is the row rank of A. Each of the r nonzero rows of R possesses a pivot of 1 in some column, and all other entries of any column containing a pivot are zero. By successively interchanging the columns of R, we obtain a matrix with the $r \times r$ identity matrix in the top left-hand corner and zeros in all subsequent rows. By construction of R, there exists a sequence of elementary row operations that reduce A to R, so in conjunction with the aforementioned column interchanges, we obtain a sequence of elementary row and column operations that reduces A to a matrix containing the $r \times r$ identity matrix in the top-left hand corner and whose subsequent rows are all zero. Considering that adding a scalar multiple of one column to another column is an elementary column operation, we can reduce any nonzero columns strictly to the right of column r to zero. Explicitly, if a is the (i,j)th component of the matrix and $1 \leq i \leq r$ and $r+1 \leq j \leq n$, then $C_j - cC_i \mapsto C_j$ yields a 0 in the (i,j)th component of the resulting matrix. Each of these is an elementary column operation, so after a sequence of elementary column operations, we obtain the desired matrix of the proposition statement. Last, neither elementary row operations nor elementary column operations affect column rank by Propositions 1.8.4 and 1.8.9, hence the column rank of A is equal to the column rank of this matrix, which equals the row rank of the matrix, i.e., the row rank of A. □

Consequently, by Proposition 1.8.11, 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 1.8.12. *The rank of a matrix is equal to the number of pivots of its row echelon form.*

Corollary 1.8.13. *An $n \times n$ matrix A is invertible if and only if* $\operatorname{rank}(A) = n$.

Example 1.8.14. Consider the following real 2×2 matrix.

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

By Corollary 1.8.12, 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.

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

We have obtained a pivot in the first row of the matrix. Consequently, the rank of A is one. We note that if the matrix A had a pivot in each of its two rows, then it would be row equivalent to the 2×2 identity matrix, hence A would be invertible by the Invertibility Criterion.

Example 1.8.15. Consider the following real 3×3 matrix.

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

By Corollary 1.8.12, in order to find the rank of A, it suffices to find the row echelon form for A. We accomplish this by performing elementary row operations on A as follows.

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

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

Before we conclude this section, we provide the following algorithm for determining bases for the row space, column space, and null space of a matrix; then, we provide several examples to illustrate the technique. Example 1.6.4 provides the definition of the null space of a real $m \times n$ matrix as the collection of vectors $\mathbf{v}$ in real n-space such that the matrix product $A\mathbf{v} = \mathbf{0}$. Consequently, we may view the null space of a matrix as all solutions of the homogeneous matrix equation $A\mathbf{x} = \mathbf{0}$.

Algorithm 1.8.16 (Constructing Bases for the Row Space, Column Space, and Null Space)**.** Consider any real $m \times n$ matrix A that is row equivalent to a matrix R in row echelon form.

1.) By definition, the row space of A is the subspace of $\mathbb{R}^n$ spanned by the rows of A. By Proposition 1.8.8, elementary row operations do not affect the row space of A. Even more, the proof of this fact illustrates that the nonzero rows of R form a basis for the row space of A.

2.) Likewise, the column space of A is the subspace of $\mathbb{R}^m$ spanned by the columns of A. Proposition 1.8.4 shows that elementary row operations do not affect the column space; the columns of R with pivots yield a basis for the column space of A by the Linear Independence Algorithm.

3.) Last, the null space of A is the subspace of $\mathbb{R}^n$ formed by the vectors $\mathbf{v}$ satisfying that $A\mathbf{v} = \mathbf{0}$. Elementary row operations correspond to left multiplication by elementary row matrices; these matrices are invertible by Proposition 1.5.6, so the vector $\mathbf{v}$ lies in the null space of A if and only if $\mathbf{v}$ lies in the null space of R. Consequently, the null space of A is the same as the null space of R, so a basis for the null space of A is provided by a basis for the null space of R. One advantage of this is that the null space of R is easily determined via back substitution.

Even more, the rank of the matrix A is equal to the number of pivots of R by Corollary 1.8.12.

Example 1.8.17. Consider the following real 2×4 matrix A.

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

By Algorithm 1.8.16, to determine the row space, column space, null space, and rank of A, we convert A to row echelon form. We require only the elementary row operation $R_2 - 5R_1 \mapsto R_2$.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -12 \end{bmatrix}$$

Consequently, there are pivots in the first and second columns of A, so the following are immediate.

$$\operatorname{row}(A) = \operatorname{span}\{[1,2,3,4],[0,-4,-8,-12]\}$$

$$\operatorname{col}(A) = \operatorname{span}\left\{\begin{bmatrix}1\\5\end{bmatrix},\begin{bmatrix}2\\6\end{bmatrix}\right\}$$

Even more, we have that $\operatorname{rank}(A) = 2$ because there are two pivots in the row echelon form of A. We obtain a basis for the null space of A by performing further elementary row operations to obtain the reduced row echelon form of A. We achieve this in the following two steps.

$$\begin{bmatrix}1 & 2 & 3 & 4\\0 & -4 & -8 & -12\end{bmatrix} \overset{(1.)}{\sim} \begin{bmatrix}1 & 2 & 3 & 4\\0 & 1 & 2 & 3\end{bmatrix} \overset{(2.)}{\sim} \begin{bmatrix}1 & 0 & -1 & -2\\0 & 1 & 2 & 3\end{bmatrix}$$

1.) We performed the elementary row operation $-\frac{1}{4}R_2 \mapsto R_2$.

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

Consequently, the null space of A consists of vectors $\mathbf{v} = [v_1, v_2, v_3, v_4]^T$ with $v_1 - v_3 - v_4 = 0$ and $v_2 + 2v_3 + 3v_4 = 0$. Explicitly, we have that $v_1 = v_3 + 2v_4$ and $v_2 = -2v_3 - 3v_4$, hence we find that

$$\mathbf{v} = \begin{bmatrix}v_1\\v_2\\v_3\\v_4\end{bmatrix} = \begin{bmatrix}v_3+2v_4\\-2v_3-3v_4\\v_3\\v_4\end{bmatrix} = \begin{bmatrix}v_3\\-2v_3\\v_3\\0\end{bmatrix} + \begin{bmatrix}2v_4\\-3v_4\\0\\v_4\end{bmatrix} = v_3\begin{bmatrix}1\\-2\\1\\0\end{bmatrix} + v_4\begin{bmatrix}2\\-3\\0\\1\end{bmatrix}.$$

We conclude that the null space is spanned by the two column vectors above so that

$$\operatorname{null}(A) = \operatorname{span}\left\{\begin{bmatrix}1\\-2\\1\\0\end{bmatrix},\begin{bmatrix}2\\-3\\0\\1\end{bmatrix}\right\}.$$

Example 1.8.18. Consider the following real 3×2 matrix A.

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

Converting to the row echelon form of A, we obtain bases for the row space and column space. Even more, if we wish to determine the null space of A, we will find the reduced row echelon form of A.

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

1.) We performed the elementary row operations $R_2 - R_1 \mapsto R_2$ and $R_3 - 2R_1 \mapsto R_3$.

2.) We performed the row operations $-R_3 \mapsto R_3$, $R_2 + R_3 \mapsto R_2$, $R_1 - R_3 \mapsto R_1$, and $R_2 \leftrightarrow R_3$.

Consequently, the rank of A is two, and the content of Proposition 1.8.11 is illustrated. We conclude that $\text{row}(A) = \mathbb{R}^2$ because the nonzero rows of $\text{RREF}(A)$ are precisely the standard basis vectors of $\mathbb{R}^2$. We use the columns of A corresponding to the columns of the row echelon form of A with pivots as a basis for the column space of A according to the Linear Independence Algorithm.

$$\text{col}(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Last, the null space of A consists of vectors $\mathbf{v} = [v_1, v_2]^T$ with $v_1 = 0$ and $v_2 = 0$ because the reduced row echelon form of A has pivots in the first two rows followed by a row of zeros. Consequently, the null space of A is the zero subspace of $\mathbb{R}^2$, i.e., we conclude that $\text{null}(A) = \{[0, 0]\}$.

Crucially, the previous examples suggest that the dimension of the null space of a real $m \times n$ matrix A is simply the number of free variables in the system of linear equations induced by the matrix equation $A\mathbf{x} = \mathbf{0}$. Call the dimension of $\text{null}(A)$ the **nullity** of A and denote it by $\text{nullity}(A)$. By Proposition 1.8.11, this is the number of pivot-free columns of the row echelon form of A. Even more, the number of zero rows of $\text{RREF}(A)$ and the rank of A sum to the number of columns of A.

Theorem 1.8.19 (Rank Equation)**.** *Given any real $m \times n$ matrix A with row echelon form R,*

1.) *the nullity of A is the number of free variables in the system of linear equations $A\mathbf{x} = \mathbf{0}$, i.e.,* $\text{nullity}(A)$ *is the number of pivot-free columns in the row echelon form R of A;*

2.) *the rank of A is the number of pivots in the row echelon form R of A; and*

3.) *the rank and nullity of A sum to the number of columns of A, i.e.,* $\text{rank}(A) + \text{nullity}(A) = n$.

1.9 Real n-Space, Revisited

Earlier in this chapter, we explored the geometry of real n-space using the language of vectors and the vector dot product. Explicitly, we established the following properties of vectors in real n-space.

a.) Each vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ in real n-space admits a non-negative length (or magnitude)

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Even more, the magnitude is **non-degenerate** in the sense that $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$, and for any real number α, we have that $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ (Proposition 1.2.2).

b.) Every pair of vectors $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ and $\mathbf{y} = [y_1, y_2, \ldots, y_n]$ in real n-space induce a real number called the dot product of $\mathbf{x}$ and $\mathbf{y}$ and defined by

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Crucially, the dot product is commutative, distributive across vector addition, homogeneous, non-degenerate in the sense that $\mathbf{x} \cdot \mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$, and satisfies the Law of Cosines

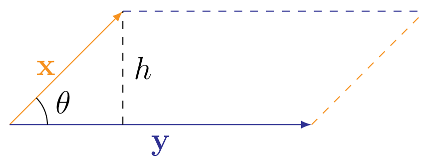
$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\| \cos(\theta)$$

for the angle θ between $\mathbf{x}$ and $\mathbf{y}$ (Proposition 1.2.8). Consequently, the dot product satisfies

$$\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos(\theta).$$

c.) Generalizing the notion of perpendicular vectors in the plane, we say that a pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in real n-space are orthogonal if $\mathbf{x} \cdot \mathbf{y} = 0$. Equivalently, we note that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if the angle θ between $\mathbf{x}$ and $\mathbf{y}$ is $90°$ (Proposition 1.2.9).

d.) Given any nonzero, non-parallel vectors $\mathbf{x}$ and $\mathbf{y}$ lying in standard position in real n-space, the parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$ can be pictured as follows.



Observe that the angle θ of intersection between $\mathbf{x}$ and $\mathbf{y}$ satisfies that $h = \|\mathbf{x}\| \sin(\theta)$. Because the area of a parallelogram is the product of its base and its height, it is $h\|\mathbf{y}\| = \|\mathbf{x}\|\|\mathbf{y}\| \sin(\theta)$. Consequently, the area of the parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$ is $\|\mathbf{x}\|\|\mathbf{y}\| \sin(\theta)$.

Consider the second and fourth points above. Given that $\mathbf{x}$ and $\mathbf{y}$ are nonzero, non-parallel vectors lying in standard position in real n-space, they span a parallelogram of area $a = \|\mathbf{x}\|\|\mathbf{y}\| \sin(\theta)$ for the angle θ between $\mathbf{x}$ and $\mathbf{y}$. Considering that $\sin^2(\theta) + \cos^2(\theta) = 1$, the following hold.

$$\begin{aligned}
a^2 &= \|\mathbf{x}\|^2\|\mathbf{y}\|^2 \sin^2(\theta) \\
a^2 &= \|\mathbf{x}\|^2\|\mathbf{y}\|^2 [1 - \cos^2(\theta)] \\
a^2 &= \|\mathbf{x}\|^2\|\mathbf{y}\|^2 - \|\mathbf{x}\|^2\|\mathbf{y}\|^2 \cos^2(\theta) \\
a^2 &= \|\mathbf{x}\|^2\|\mathbf{y}\|^2 - (\mathbf{x} \cdot \mathbf{y})^2
\end{aligned}$$

Restricting our attention to vectors in the plane, we may assume that $\mathbf{x} = [x_1, x_2]$ and $\mathbf{y} = [y_1, y_2]$. By definition, we have that $\|\mathbf{x}\|^2 = x_1^2 + x_2^2$, $\|\mathbf{y}\|^2 = y_1^2 + y_2^2$, and $\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2$ so that

$$\begin{aligned}
a^2 &= \|\mathbf{x}\|^2\|\mathbf{y}\|^2 - (\mathbf{x} \cdot \mathbf{y})^2 \\
&= (x_1^2y_1^2 + x_1^2y_2^2 + x_2^2y_1^2 + x_2^2y_2^2) - (x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2) \\
&= x_1^2y_2^2 - 2x_1x_2y_1y_2 + x_2y_1^2 \\
&= (x_1y_2 - x_2y_1)^2.
\end{aligned}$$

Consequently, the area of the parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$ in the plane is $a = |x_1y_2 - x_2y_1|$. We refer to the real number $x_1y_2 - x_2y_1$ as the **determinant** of the vectors $\mathbf{x} = [x_1, x_2]$ and $\mathbf{y} = [y_1, y_2]$. Considering these vectors as the rows of a 2×2 matrix, we define the determinant of a 2×2 matrix

$$\det\left(\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}\right) = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = x_1y_2 - x_2y_1$$

as the difference of the product of the diagonal and **antidiagonal** entries of the matrix.

Example 1.9.1. Consider the following real 2×2 matrix.

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

We have that $\det(A) = (1)(4) - (2)(3) = -2$, so the determinant of A is -2.

Example 1.9.2. Consider the following real 2×2 matrix.

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

We have that $\det(A) = (1)(6) - (2)(3) = 0$. Crucially, we note that the vectors $[1, 2]$ and $[3, 6]$ in real 2-space are parallel because $[3, 6] = 2[1, 2]$, hence they do not span a parallelogram.

Example 1.9.3. Consider the parallelogram in real 2-space with vertices $(0, 0)$, $(2, 1)$, $(5, 1)$, and $(3, 0)$. Graphing the parallelogram in the plane, we find that it is spanned by the vectors $\mathbf{x} = [2, 1]$ and $\mathbf{y} = [3, 0]$. Consequently, the area of the parallelogram is the absolute value of the determinant

$$\begin{vmatrix} 2 & 1 \\ 3 & 0 \end{vmatrix} = (2)(0) - (1)(3) = -3.$$

We conclude that the area of the parallelogram is 3 units^2. Crucially, we note that this is exactly

$$\begin{vmatrix} 3 & 0 \\ 2 & 1 \end{vmatrix} = (3)(1) - (0)(2) = 3,$$

hence swapping the rows of the matrix simply changed the sign of the determinant.

Given any vectors $\mathbf{x} = [x_1, x_2, x_3]$ and $\mathbf{y} = [y_1, y_2, y_3]$ in real 3-space, the **cross product**

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = (x_2y_3 - x_3y_2)\mathbf{e}_1 - (x_1y_3 - x_3y_1)\mathbf{e}_2 + (x_1y_2 - x_2y_1)\mathbf{e}_3$$

of the vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as the **symbolic determinant** of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{e}_3$ with the vectors $\mathbf{x}$ and $\mathbf{y}$ expressed as the second and third rows of a matrix, respectively. Crucially, observe that $\mathbf{x} \times \mathbf{y}$ is in fact a vector in real 3-space satisfying the following properties.

Proposition 1.9.4 (Properties of the Cross Product)**.** *Consider any vectors* $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ *in real 3-space.*

1.) *We have that* $\mathbf{x} \times \mathbf{y} = -(\mathbf{y} \times \mathbf{x})$.

2.) *We have that* $\mathbf{x} \times (\alpha\mathbf{x}) = 0$ *for all real numbers* α.

3.) *We have that* $\mathbf{x} \times (\alpha\mathbf{y}) = \alpha(\mathbf{x} \times \mathbf{y})$ *for all real numbers* α.

4.) *We have that* $(\mathbf{x} + \mathbf{y}) \times \mathbf{z} = (\mathbf{x} \times \mathbf{z}) + (\mathbf{y} \times \mathbf{z})$ *and* $\mathbf{x} \times (\mathbf{y} + \mathbf{z}) = (\mathbf{x} \times \mathbf{z}) + (\mathbf{y} \times \mathbf{z})$.

5.) *We have that* $(\mathbf{x} \times \mathbf{y}) \cdot \mathbf{z} = (\mathbf{z} \times \mathbf{x}) \cdot \mathbf{y} = (\mathbf{y} \times \mathbf{z}) \cdot \mathbf{x}$.

6.) *We have that* $(\mathbf{x} \times \mathbf{y}) \cdot \mathbf{x} = 0$ *and* $(\mathbf{x} \times \mathbf{y}) \cdot \mathbf{y} = 0$.

Proof. Each of the first three properties follows immediately from Theorem 1.10.13 because the cross product is defined by a 3×3 determinant. Likewise, the fourth property follows from Proposition 1.10.7 because the cross product $(\mathbf{x}+\mathbf{y})\times\mathbf{z}$ is determined by the matrix whose second row is the sum of the second row of the matrices that determine $\mathbf{x}\times\mathbf{z}$ and $\mathbf{y}\times\mathbf{z}$. Consequently, it suffices to prove the fifth, sixth, and seventh properties of the cross product. We will assume that $\mathbf{x} = [x_1, x_2, x_3]$, $\mathbf{y} = [y_1, y_2, y_3]$, and $\mathbf{z} = [z_1, z_2, z_3]$. Computing the cross products yields the following.

$$\mathbf{x}\times\mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = (x_2y_3 - x_3y_2)\mathbf{e}_1 - (x_1y_3 - x_3y_1)\mathbf{e}_2 + (x_1y_2 - x_2y_1)\mathbf{e}_3$$

$$\mathbf{y}\times\mathbf{z} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix} = (y_2z_3 - y_3z_2)\mathbf{e}_1 - (y_1z_3 - y_3z_1)\mathbf{e}_2 + (y_1z_2 - y_2z_1)\mathbf{e}_3$$

By subsequently taking the dot products, we obtain the following identities.

$$\begin{aligned}(\mathbf{x}\times\mathbf{y})\cdot\mathbf{z} &= (x_2y_3 - x_3y_2)z_1 - (x_1y_3 - x_3y_1)z_2 + (x_1y_2 - x_2y_1)z_3 \\ &= x_1y_2z_3 - x_1y_3z_2 + x_2y_3z_1 - x_2y_1z_3 + x_3y_1z_2 - x_3y_2z_1 \\ &= (y_2z_3 - y_3z_2)x_1 - (y_1z_3 - y_3z_1)x_2 + (y_1z_2 - y_2z_1)x_3 = (\mathbf{y}\times\mathbf{z})\cdot\mathbf{x}\end{aligned}$$

Changing the names of the vectors establishes that each of these quantities is equal to $(\mathbf{z}\times\mathbf{x})\cdot\mathbf{y}$, so we omit the details. Consequently, the fifth property of the cross product is established. Combining the first and fifth properties above, we conclude that $(\mathbf{x}\times\mathbf{y})\cdot\mathbf{x} = (\mathbf{y}\times\mathbf{x})\cdot\mathbf{x} = -(\mathbf{x}\times\mathbf{y})\cdot\mathbf{x}$, hence we must have that $(\mathbf{x}\times\mathbf{y})\cdot\mathbf{x} = 0$. Likewise, it follows that $(\mathbf{x}\times\mathbf{y})\cdot\mathbf{y} = 0$, as desired. □

Consequently, by the sixth part of Properties of the Cross Product, the cross product yields a tried-and-true method to construct nonzero vectors orthogonal to a given pair of vectors $\mathbf{x}$ and $\mathbf{y}$.

Example 1.9.5. Constructing a nonzero vector that is orthogonal to the vectors $\mathbf{x} = [-1, 3, 4]$ and $\mathbf{y} = [-2, -1, 3]$ in real 3-space amount to determining the cross product $\mathbf{x}\times\mathbf{y}$.

$$\mathbf{x}\times\mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ -1 & 3 & 4 \\ -2 & -1 & 3 \end{vmatrix} = [(3)(3) - (4)(-1)]\mathbf{e}_1 - [(-1)(3) - (4)(-2)]\mathbf{e}_2 + [(-1)(-1) - (3)(-2)]\mathbf{e}_3$$

Computing each of the coefficients yields that $\mathbf{x}\times\mathbf{y} = [13, -5, 7]$. We can rest assured by the theory that $\mathbf{x}\times\mathbf{y}$ is in fact orthogonal to $\mathbf{x}$ and $\mathbf{y}$, but we may compute the dot products for verification.

$$(\mathbf{x}\times\mathbf{y})\cdot\mathbf{x} = [13, -5, 7]\times[-1, 3, 4] = (13)(-1) + (-5)(3) + (7)(4) = -13 - 15 + 28 = 0$$
$$(\mathbf{x}\times\mathbf{y})\cdot\mathbf{y} = [13, -5, 7]\times[-2, -1, 3] = (13)(-2) + (-5)(-1) + (7)(3) = -26 + 5 + 21 = 0$$

Example 1.9.6. Consider the plane determined by the non-collinear points $(1, 2, 3)$, $(2, 5, 0)$, and $(-1, 0, 3)$ in real 3-space. Considering the point $(1, 2, 3)$ as a local origin, we obtain vectors

$$\mathbf{x} = [2, 5, 0] - [1, 2, 3] = [1, 3, -3] \text{ and}$$
$$\mathbf{y} = [-1, 0, 3] - [1, 2, 3] = [-2, -2, 0]$$

that span the plane: indeed, these vectors are linearly independent (the third coordinate of $\mathbf{y}$ is zero), and they lie in the plane by construction, hence they form a basis for the plane that contains them. Consequently, to determine a vector orthogonal to the plane, it suffices by the third and fourth parts of Properties of the Cross Product to determine the cross product of $\mathbf{x}$ and $\mathbf{y}$.

$$\mathbf{x} \times \mathbf{y} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ 1 & 3 & -3 \\ -2 & -2 & 0 \end{vmatrix} = -6\mathbf{e}_1 + 6\mathbf{e}_2 + 4\mathbf{e}_3 = [-6, 6, 4]$$

Considering the point $(1, 2, 3)$ as a local origin, every point in the plane spanned by $\mathbf{x}$ and $\mathbf{y}$ is of the form $(x - 1, y - 2, z - 3)$, hence $\mathbf{x} \times \mathbf{y}$ is orthogonal to the vectors $[x - 1, y - 2, z - 3]$.

$$-6(x - 1) + 6(y - 2) + 4(z - 3) = [-6, 6, 4] \cdot [x - 1, y - 2, z - 3] = (\mathbf{x} \times \mathbf{y}) \cdot [x - 1, y - 2, z - 3] = 0$$

One can simplify the left-hand side to obtain the equation of the plane $-3x + 3y + 2z = 9$.

Geometrically, the magnitude of the cross product of a pair of vectors has a nice interpretation.

Proposition 1.9.7. *Given any nonzero, non-parallel vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *lying in standard position in real 3-space, the area of the parallelogram spanned by* $\mathbf{x}$ *and* $\mathbf{y}$ *is* $\|\mathbf{x} \times \mathbf{y}\|$.

Proof. By the exposition preceding the list provided at the beginning of the section, the square of the area a of the parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$ satisfies that

$$a^2 = \|\mathbf{x}\|^2 \|\mathbf{y}\|^2 - (\mathbf{x} \cdot \mathbf{y})^2.$$

Consider the vectors $\mathbf{x} = [x_1, x_2, x_3]$ and $\mathbf{y} = [y_1, y_2, y_3]$ according to their coordinates. By definition of the magnitude and the dot product, we have that $\|\mathbf{x}\|^2 = x_1^2 + x_2^2 + x_3^2$, $\|\mathbf{y}\|^2 = y_1^2 + y_2^2 + y_3^2$, $\mathbf{x} \cdot \mathbf{y} = x_1y_1 + x_2y_2 + x_3y_3$, and $\|\mathbf{x} \times \mathbf{y}\|^2 = (x_2y_3 - x_3y_2)^2 + (x_1y_3 - x_3y_1)^2 + (x_1y_2 - x_2y_1)^2$. Expanding and simplifying the identity for a^2 in terms of the coordinates of $\mathbf{x}$ and $\mathbf{y}$ completes the proof. □

1.10 Determinants of $n \times n$ Matrices

Back in Section 1.9, we defined the **determinant** of a real 2×2 matrix as the real number

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

Explicitly, the determinant of a real 2×2 matrix is the difference of the product $a_{11}a_{22}$ of its diagonal entries and the product $a_{12}a_{21}$ of its **antidiagonal** entries. Generally, the determinant of an $n \times n$ matrix can be defined **recursively** for any positive integer n. Beyond special cases, we will not typically concern ourselves with determinants of matrices with more than three rows and columns, so it suffices to define the determinant of a real 3×3 matrix. Out of desire for notational convenience, we will seldom use the $\det(-)$ notation for a matrix whose components we wish to display explicitly; rather, we will denote the determinant using vertical bars as follows.

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

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

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

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

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

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

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

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

Definition 1.10.1. Given any $n \times n$ matrix A, let A_{ij} denote the $(n - 1) \times (n - 1)$ submatrix of A obtained by deleting the ith row and jth column of A. We define the **determinant** of A by

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

Example 1.10.2. 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 1.10.3. By the recursive definition of the determinant, we obtain the following.

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

Earlier in this 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 1.10.4. *Given any $n \times n$ matrix A and any scalar α, consider the $n \times n$ matrix B obtained from A by multiplying any row of A by α. We have that* $\det(B) = \alpha \det(A)$.

Proof. We will assume that B is obtained from A by multiplying the ith row of A by α. Consider the $(n-1) \times (n-1)$ matrix A_{ij} obtained from A by deleting the ith row and jth column of A. By hypothesis, we have that $b_{ij} = \alpha a_{ij}$ and $B_{ij} = A_{ij}$ for each integer $1 \leq j \leq n$. By Definition 1.10.1, we conclude that $\det(B) = \sum_{j=1}^n (-1)^{i+j} b_{ij} \det(B_{ij}) = \alpha\Big(\sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij})\Big) = \alpha \det(A)$. □

Corollary 1.10.5. *Given any $n \times n$ matrix A with a zero row, we have that* $\det(A) = 0$.

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

Corollary 1.10.6. *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. □

Proposition 1.10.7. *Given any $n \times n$ matrices A and B that are equal except in one row, consider the $n \times n$ matrix C obtained from A and B by adding the two rows of A and B that are distinct and including all of the rows of A and B that are equal. We have that* $\det(C) = \det(A) + \det(B)$.

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

so that $\det(C_{ik}) = \det(A_{ik}) = \det(B_{ik})$ for all integers $1 \leq i \leq k$. We conclude the result as follows.

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

□

Proposition 1.10.8. *Given any $n \times n$ matrix A with two equal rows, we have that* $\det(A) = 0$.

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

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

Consequently, we may assume inductively that the result holds for some integer $n \geq 3$. We may assume that the ith row of A and the jth row of A are equal for some integers $1 \leq i < j \leq n$. Consider the $n \times n$ matrix $A_{k\ell}$ obtained from A by deleting the kth row and ℓth column of A for some integer $1 \leq k \leq n$ that is distinct from both i and j. We may find such an integer k by assumption that $n \geq 3$. Crucially, we note that the ith row of $A_{k\ell}$ and the jth row of $A_{k\ell}$ are equal for all integers $1 \leq \ell \leq n$, hence by induction, it follows that $\det(A_{k\ell}) = 0$ for all integers $1 \leq \ell \leq n$. By Definition 1.10.1, we conclude the desired result that $\det(A) = \sum_{\ell=1}^{n} (-1)^{k+\ell} a_{k\ell} \det(A_{k\ell}) = 0$. □

Proposition 1.10.9. *Given any $n \times n$ matrix A, any scalar α, and any integers $1 \leq i < j \leq n$, consider the $n \times n$ matrix B obtained from A by replacing the jth row of A with the sum of α times the ith row and the jth row of A. We have that* $\det(B) = \det(A)$. *Put another way, if we add any scalar multiple of a row of an $n \times n$ matrix to any other row, the determinant does not change.*

Proof. By definition of B, we have that $b_{k\ell} = a_{k\ell}$ for all integers $1 \leq k \leq n$ such that $k \neq j$ and $b_{j\ell} = \alpha a_{i\ell} + a_{j\ell}$ for all integers $1 \leq \ell \leq n$. Consider the $n \times n$ matrix C obtained from A by replacing the jth row of A with α times the ith row of A. Crucially, observe that B is obtained from A and C by including all common rows of A and C and taking the sum of the jth rows of A and C as the jth row of B. Consequently, by Proposition 1.10.7, we have that $\det(B) = \det(A) + \det(C)$. Consider the $n \times n$ matrix D obtained from A by replacing the jth row of A with the ith row of A. Explicitly, we note that C is obtained from D by multiplying the jth row of D by α. By Proposition 1.10.4, we have that $\det(C) = \alpha \det(D)$. Considering that the ith and jth rows of D are equal, it follows from Proposition 1.10.8 that $\det(D) = 0$ so that $\det(B) = \det(A) + \det(C) = \det(A) + \alpha \det(D) = \det(A)$. □

Corollary 1.10.10. *Given any $n \times n$ matrix A, if some row of A can be written as a linear combination of some other rows of A, then we have that* $\det(A) = 0$.

Proof. We will denote by A_i the ith row of A. Consider the case that $A_i = \alpha_1 A_{i_1} + \cdots + \alpha_k A_{i_k}$ for some integers $1 \leq i_1 < \cdots < i_k \leq n$ and some scalars $\alpha_1, \dots, \alpha_k$. 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_jth row of A from the ith row of A for each integer $1 \leq j \leq k$ to reduce the ith row of A to zero. By Proposition 1.10.9, this process does not change the determinant of A; on the other hand, the determinant of the resulting matrix is zero by Corollary 1.10.5 so that $\det(A) = 0$. □

Proposition 1.10.11. *Given any $n \times n$ matrix A, consider the $n \times n$ matrix B obtained from A by interchanging any pair of rows of A. We have that* $\det(B) = -\det(A)$. *Put another way, swapping any pair of rows of an $n \times n$ matrix alters the sign of the determinant.*

Proof. Certainly, if any pair of rows of A are equal, then we have that $\det(B) = 0 = -0 = -\det(A)$. Consequently, we may assume that all rows of A are distinct. Crucially, we may obtain B from A by a sequence of operations that alter the determinant in exactly the manner claimed. Begin with the matrix C that is obtained from A by replacing the ith row of A with the sum of the ith and jth rows of A. By Propositions 1.10.7 and 1.10.8, it follows that $\det(C) = \det(A)$. Consider next the matrix D that is obtained from C by subtracting the ith row of C from the jth row of C so that the jth row of D is the ith row of A with the opposite sign. By Proposition 1.10.9, it follows that $\det(D) = \det(C) = \det(A)$. Last, we notice that B can be obtained from D by multiplying the jth row of D by -1; then, Proposition 1.10.4 yields that $\det(B) = -\det(D) = -\det(A)$. □

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

Proposition 1.10.12. *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. □

Theorem 1.10.13 (Properties of the Determinant)**.** *Consider any $n \times n$ matrix A.*

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

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

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

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

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

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

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

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

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

Example 1.10.14. 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 1.10.10 that $\det(A) = 0$. One could make a similar argument with the first and third rows of A.

Example 1.10.15. 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 1.10.9, 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 1.10.3 and Proposition 1.10.11, we conclude the following.

$$\det(A) = -\det(C) = -\begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} = 2$$

Example 1.10.16. 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 Theorem 1.10.13, we have that $\det(A) = -\det(I_{3\times 3})$. Last, observe the following.

$$\det(A) = -\det(I_{3\times 3}) = -\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -1\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = -1$$

1.11 The Adjugate of a Matrix

Every square matrix possesses a numerical invariant called its determinant. We will come to understand throughout this course that the determinant of a matrix contains a wealth of information about the properties of the matrix (e.g., Properties of the Determinant). Computing the determinant of a square matrix amounts to recursively expanding the matrix about some row or column by multiplying each subsequent entry a_{ij} of the specified row or column of the matrix by the determinant of the submatrix obtained by deleting the ith row and column jth column of the matrix.

One other way to obtain the determinant of an $n \times n$ matrix A is as the coefficient of the **scalar matrix** $\det(A)I$. We achieve this by taking the product of A with its **adjugate matrix** $\operatorname{adj}(A)$. We note that the adjugate matrix can also be encountered under the name of the **classical adjoint** (cf. [HK71, Exercise 5.2.3]); however, we will not adopt such terminology here because it is often associated with another object related to linear transformations. Like with the determinant, the adjugate matrix is defined recursively beginning with the case of 2×2 matrices as follows.

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

Explicitly, the adjugate matrix of any 2×2 matrix is obtained by swapping the elements on the main diagonal and changing the signs of the elements on the antidiagonal. Observe the following.

$$\operatorname{adj}(A)A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} = \begin{bmatrix} \det(A) & 0 \\ 0 & \det(A) \end{bmatrix} = \det(A)I$$

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

We will soon verify that this rationale is much more general and applies to square matrices of all sizes. Before we are able to do this, we must define the adjugate of any $n \times n$ matrix.

Definition 1.11.1. Given any $n \times n$ matrix A, let A_{ij} denote the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the ith row and jth column of A. We refer to the scalar $\mu_{ij} = \det(A_{ij})$ used in the definition of the determinant of A as the (i, j)th **minor** of the matrix A.

Definition 1.11.2. Given any $n \times n$ matrix A, let μ_{ij} denote the (i, j)th minor of A, i.e., μ_{ij} is the determinant of the $(n-1) \times (n-1)$ submatrix of A obtained by deleting the ith row and jth column of A. We refer to the scalar $\gamma_{ij} = (-1)^{i+j}\mu_{ij}$ as the (i, j)th **cofactor** of the matrix A.

Definition 1.11.3. Given any $n \times n$ matrix A, let γ_{ij} denote the (i, j)th cofactor of A, i.e., suppose that $\gamma_{ij} = (-1)^{i+j} \mu_{ij} = (-1)^{i+j} \det(A_{ij})$, where A_{ij} is the matrix obtained from A by deleting its ith row and jth column. We refer to the matrix $\Gamma = \left[\gamma_{ij}\right]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$ as the **cofactor matrix** of A.

Definition 1.11.4. Given any $n \times n$ matrix A, let Γ denote the $n \times n$ cofactor matrix of A. We refer to the $n \times n$ matrix $\operatorname{adj}(A) = \Gamma^T$ as the **adjugate** (or adjugate matrix) of A.

One thing to notice is that the adjugate matrix can be defined for any square matrix over any ring because it only involves the operations of multiplication and subtraction; we will see that this provides a drastic improvement to the method of Gaussian Elimination we used previously to detect if a matrix is invertible. Explicitly, the process of Gaussian Elimination is only defined for matrices over fields because division is sometimes necessary to find the reduced row echelon form of a matrix.

Example 1.11.5. 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 1.10.3, we have that $\det(A) = -2$. We will verify that $\operatorname{adj}(A)A = -2I = \det(A)I$. By Definition 1.11.4, we note that $\operatorname{adj}(A)$ is given by the transpose of the cofactor matrix Γ of A. By Definition 1.11.3, the (i, j)th component of the cofactor matrix Γ is the (i, j)th cofactor γ_{ij} of A. By Definition 1.11.2, 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 1.11.1, each minor μ_{ij} is given by the determinant of the 2×2 matrix A_{ij} obtained from A by deleting its ith row and jth column. Consequently, we find the following minors.

$$\mu_{11} = \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1 \qquad \mu_{21} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 \qquad \mu_{31} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

$$\mu_{12} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \qquad \mu_{22} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \qquad \mu_{32} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1$$

$$\mu_{13} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \qquad \mu_{23} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1 \qquad \mu_{33} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

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

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

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

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

We are now in position to form the 3×3 cofactor matrix Γ as follows.

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

Observe that in this case, Γ is a symmetric matrix because each row of Γ is equal to the corresponding column of Γ. Consequently, we have that $\operatorname{adj}(A) = \Gamma^T = \Gamma$. Even more, the following holds.

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

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

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

Example 1.11.6. 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 1.11.4, we have that $\operatorname{adj}(A)$ is equal to the transpose of the cofactor matrix Γ of A. By Definition 1.11.3, we construct the cofactor matrix Γ by finding each of the cofactors γ_{ij} of A. By Definition 1.11.2, 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 1.11.1, each minor μ_{ij} is given by the determinant of the 2×2 matrix A_{ij} obtained from A by deleting its ith row and jth column. Consequently, we find the following minors.

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

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

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

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

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

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

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

We are now in position to form the 3×3 cofactor matrix Γ as follows.

$$\Gamma = \begin{bmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{bmatrix}$$

Observe that in this case, Γ is a symmetric matrix because each row of Γ is equal to the corresponding column of Γ. Consequently, we have that $\operatorname{adj}(A) = \Gamma^T = \Gamma$. Even more, the following holds.

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

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 1.11.7. *Given any $n \times n$ matrix A, we have that* $\operatorname{adj}(A)A = \det(A)I$.

Proof. By Definition 1.11.4, we have that $\operatorname{adj}(A) = \Gamma^T$, where Γ is the cofactor matrix of A. By Definition 1.11.3, the (i, j)th component of Γ is the (i, j)th cofactor γ_{ij} of A. By Definition 1.11.2, it follows that $\gamma_{ij} = (-1)^{i+j} \det(A_{ij})$, where A_{ij} is the $(n-1)\times(n-1)$ submatrix of A obtained from A by deleting the ith row and jth column of A. Consequently, the (i, j)th component of $\operatorname{adj}(A)$ is the (j, i)th component of Γ, i.e., the (j, i)th cofactor $\gamma_{ji} = (-1)^{i+j} \det(A_{ji})$ of A. By Definition 1.3.15, we note that the (i, j)th component of $\operatorname{adj}(A)A$ is the sum of the products of the (i, k)th component of $\operatorname{adj}(A)$ and the (k, j)th component of A for each integer $1 \leq k \leq n$, i.e., the (i, j)th component of $\operatorname{adj}(A)A$ is $\sum_{k=1}^{n}(-1)^{i+k}a_{kj} \det(A_{ki})$. By Definition 1.10.1, we conclude that the (i, i)th components of $\operatorname{adj}(A)A$ are exactly $\det(A)$ because these are obtained from the aforementioned sum by setting $i = j$. Consequently, it suffices to prove that $\sum_{k=1}^{n}(-1)^{i+k}a_{kj} \det(A_{ki}) = 0$ whenever $i \neq j$.

Consider the $n \times n$ matrix B obtained from A by replacing the ith column of A with the jth column of A. Observe that for each integer $1 \leq k \leq n$, we have that $b_{ki} = a_{kj}$ because the ith column of B is equal to the jth column of A. Even more, we have that $B_{ki} = A_{ki}$ for all integers $1 \leq k \leq n$ because A and B only differ in the ith column. By Theorem 1.10.13, we have that

$$0 = \det(B) = \sum_{k=1}^{n}(-1)^{i+k}b_{ki} \det(B_{ki}) = \sum_{k=1}^{n}(-1)^{i+k}a_{kj} \det(A_{ki}).$$

We conclude therefore that the non-diagonal components of $\operatorname{adj}(A)A$ are zero, as desired. □

Proposition 1.11.8. *Given any $n \times n$ matrix A, we have that* $\operatorname{adj}(A^T) = \operatorname{adj}(A)^T$. *Put another way, the adjugate of the transpose is the transpose of the adjugate.*

Proof. Crucially, observe that deleting the ith row and jth column of A^T is the same as deleting the ith column and jth row of A and taking its transpose because the ith row of A^T is the ith column of A and the jth column of A^T is the jth row of A. Consequently, we have that $(A^T)_{ij} = (A_{ji})^T$. By the underlying definitions of the adjugate, the (i, j)th component of $\operatorname{adj}(A^T)$ is $(-1)^{i+j} \det((A^T)_{ij})$, hence by our opening remarks, the (i, j)th component of $\operatorname{adj}(A^T)$ is exactly $(-1)^{i+j} \det((A_{ji})^T)$. By Proposition 1.10.12, it follows that the (i, j)th component of $\operatorname{adj}(A^T)$ is $(-1)^{i+j} \det(A_{ji})$. Considering that this is the (j, i)th component of $\operatorname{adj}(A)$ by definition, we conclude that the (i, j)th component of $\operatorname{adj}(A^T)$ is the (i, j)th component of $\operatorname{adj}(A)^T$, hence the two matrices in consideration are equal. □

Corollary 1.11.9. *Given any $n \times n$ matrix A, we have that $A \operatorname{adj}(A) = \det(A)I$.*

Proof. By Proposition 1.11.7, we have that $\operatorname{adj}(A^T)A^T = \det(A^T)I$. By Proposition 1.10.12, we have that $\det(A^T) = \det(A)$ so that $\operatorname{adj}(A^T)A^T = \det(A)I$. By Proposition 1.11.8, we have that $\operatorname{adj}(A^T) = \operatorname{adj}(A)^T$ so that $\operatorname{adj}(A)^T A^T = \det(A)I$. Last, by Proposition 1.3.24, we conclude that

$$\det(A)I = \det(A)I^T = (\det(A)I)^T = (\operatorname{adj}(A)^T A^T)^T = (A^T)^T(\operatorname{adj}(A)^T)^T = A \operatorname{adj}(A). \qquad \square$$

Theorem 1.11.10. *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 1.11.7 and 1.11.9 imply that

$$\left(\frac{1}{\det(A)} \operatorname{adj}(A)\right) A = I = A\left(\frac{1}{\det(A)} \operatorname{adj}(A)\right)$$

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

Example 1.11.11. By Example 1.10.2, the following 3×3 matrix is not invertible.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

Example 1.11.12. By Example 1.10.3, the following 3×3 matrix is invertible.

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

Example 1.11.13. By Example 1.10.14, 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 1.11.14. By Example 1.10.15, the following 3×3 matrix is invertible.

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

Example 1.11.15. By Example 1.10.16, the following 3×3 matrix is invertible.

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

We could have also noticed that it is row equivalent to the 3×3 identity matrix.

Before we conclude this section, we state a critically important property of determinants.

Theorem 1.11.16. *Given any $n \times n$ matrices A and B, we have that* $\det(AB) = \det(A)\det(B)$.

Proof. Consider the unique reduced row echelon form $R = \text{RREF}(A)$ for A. By Theorem 1.10.13, there exists a scalar α that is uniquely determined by the elementary row operations $E_1, \dots, E_k$ that are used to convert R to A such that $\det(A) = \alpha \det(R)$ and $E_k \cdots E_1 R = A$. Either R has a row consisting of zeros, or it is the $n \times n$ identity matrix. By the aforementioned corollary, if R has a row consisting of zeros, then $\det(R) = 0$ so that $\det(A) = \alpha \det(R) = 0$ and $\det(A)\det(B) = 0$. By Theorem 1.11.10, we have that $\det(AB)$ is nonzero if and only if AB is invertible if and only if RB is invertible. By assumption that R has a row consisting of zeros, it follows that RB is not invertible because it has a column consisting of zeros, and we conclude that $\det(AB) = 0$. Conversely, if R is the $n \times n$ identity matrix, then $\det(A) = \alpha \det(R) = \alpha$ and $A = E_k \cdots E_1 R = E_k \cdots E_1$, from which we conclude that $\det(A)\det(B) = \alpha \det(B) = \det(E_k \cdots E_1 B) = \det(E_k \cdots E_1 R B) = \det(AB)$. □

1.12 Chapter Overview

This section is currently under construction.

Chapter 2

Canonical Forms of Matrices

We introduced in the first chapter the notion of vectors in real n-space and their geometry. Even more importantly, we discussed matrices, their arithmetic, and numerous important properties of them. Essentially, the theory of matrices vastly simplifies the algebra of large sets of data. We will soon demonstrate that the collection of all real $m \times n$ matrices forms an algebraic structure called a vector space; vector spaces are ubiquitous throughout mathematics, so it is critical to understand their properties. We will soon define functions (linear transformations) between vector spaces study certain vector spaces called the kernel and the range associated to a linear transformation. Ultimately, we will establish 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 Characteristic and Minimal Polynomials

We introduce in this section two polynomial invariants of an $n \times n$ matrix. Both of these polynomials are related to the determinant of a matrix associated with the given square matrix. Explicitly, suppose that A is any $n \times n$ matrix. We will adopt the shorthand I for the $n \times n$ identity matrix. Given any indeterminate x, we refer to the matrix $xI - A$ as the **characteristic matrix** of A. Both A and I are by assumption $n \times n$ matrices, hence the characteristic matrix $xI - A$ is likewise an $n \times n$ matrix. Even more, we note that diagonal of $xI - A$ consists of $x - a_{ii}$ for each integer $1 \leq i \leq n$ and the off-diagonal components of $xI - A$ are the off-diagonal components of A with the opposite sign. Explicitly, we have that $xI - A = \begin{bmatrix} x\delta_{ij} - a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$ for the Kronecker delta δ_{ij}.

Example 2.1.1. Consider the following 2×2 matrix A and its characteristic matrix $xI - A$.

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad\qquad xI - A = \begin{bmatrix} x-1 & -2 \\ -2 & x-1 \end{bmatrix}$$

We note that $\det(xI - A) = (x-1)(x-1) - (-2)(-2) = x^2 - 2x - 3 = (x-3)(x+1)$.

Example 2.1.2. Consider the following 3×3 matrix A and its characteristic matrix $xI - A$.

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \qquad xI - A = \begin{bmatrix} x-1 & -1 & 0 \\ -1 & x & -1 \\ 0 & -1 & x-1 \end{bmatrix}$$

We note that $\det(xI-A) = (x-1)[x(x-1)-(-1)(-1)]-(-1)[(-1)(x-1)-(-1)(0)]$. By simplifying this, we obtain that $\det(xI-A) = (x-1)(x^2-x-1)-(x-1)$, hence factoring by grouping yields that $\det(xI-A) = (x-1)(x^2-x-1-1) = (x-1)(x^2-x-2) = (x-1)(x-2)(x+1)$.

Considering that we may always expand the determinant of the $n \times n$ characteristic matrix $xI-A$ along the first row, it follows that $\chi_A(x) = \det(xI-A)$ must be a polynomial in indeterminate x of degree n because the product of the diagonal elements of $xI-A$ form a polynomial in indeterminate x of degree n. (Concretely, one can prove this by induction.) Consequently, we refer to the determinant $\det(xI-A)$ of the characteristic matrix of A as the **characteristic polynomial** of A. One of the first observations that we can make regarding the characteristic polynomial is the following.

Proposition 2.1.3. *Given any $n \times n$ matrix A with characteristic polynomial $\chi(x)$, we have that $\det(A) = (-1)^n \chi(0)$. Put another way, the constant term of $\chi(x)$ is $(-1)^n \det(A)$.*

Proof. By definition of the characteristic polynomial, we have that $\chi(0) = \det(0I - A) = \det(-A)$. Consequently, by Proposition 1.10.6, it follows that $\chi(0) = (-1)^n \det(A)$, hence the result can be obtained by multiplying both sides of this identity by $(-1)^n$ and using the fact that $(-1)^{2n} = 1$. $\square$

Example 2.1.4. Given any 2×2 matrix A with characteristic polynomial $\chi(x) = x^2 - 2x + 1$, we must have that $\det(A) = (-1)^2(0^2 - 2(0) + 1) = 1$.

Example 2.1.5. Given any 3×3 matrix A with characteristic polynomial $\chi(x) = x^3 - ex^2 + \pi$, we must have that $\det(A) = (-1)^3(0^3 - e(0)^2 + \pi) = -\pi$.

Given any polynomial $p(x) = c_k x^k + \cdots + c_1 x + c_0$, we can "plug in" any $n \times n$ matrix A to the polynomial $p(x)$ to obtain a **matrix polynomial** $p(A) = c_k A^k + \cdots + c_1 A + c_0 I$. Explicitly, the matrices A^i for each integer $1 \leq i \leq k$ are given by the i-fold product of the matrix A with itself, and the constant term c_0 of $p(x)$ becomes the scalar matrix $c_0 I$ in the matrix polynomial $p(A)$.

Example 2.1.6. Consider the 2×2 matrix A of Example 2.1.1. Observe that the following hold.

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

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

$$(A - 3I)(A + I) = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Consequently, the matrix polynomial $\chi(A) = (A - 3I)(A + I)$ yields the 2×2 zero matrix.

Example 2.1.7. Consider the 3×3 matrix A of Example 2.1.2. Observe that the following hold.

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

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

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

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

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

Consequently, the matrix polynomial $\chi(A) = (A - I)(A - 2I)(A + I)$ yields the 3×3 zero matrix.

Our next theorem demonstrates that these examples are indicative of a general phenomenon.

Theorem 2.1.8 (Cayley-Hamilton Theorem)**.** *Given any $n \times n$ matrix A with characteristic polynomial $\chi(x)$, it holds that $\chi(A) = O$, i.e., the characteristic polynomial of A annihilates A.*

Proof. Considering that we have the adjugate matrix at our disposal from our discussion in the previous section 1.11, we will incorporate it into this proof; however, there are a wealth of excellent proofs of this fact that the interested reader is encouraged to discover. Considering that the characteristic matrix $xI - A$ of A is an $n \times n$ matrix whose coefficients lie in a polynomial ring, it admits an adjugate matrix $\operatorname{adj}(xI - A)$ such that $\operatorname{adj}(xI - A)(xI - A) = \det(xI - A)I = \chi(x)I$ by Proposition 1.11.7 and the definition of the characteristic polynomial $\chi(x)$. On the other hand, the components of the $n \times n$ matrices $xI - A$, $\operatorname{adj}(xI - A)$, and $\chi(x)I$ are polynomials in indeterminate x, hence these matrices can be written uniquely as formal polynomials with matrix coefficients: we must simply determine the part of the matrices corresponding to each monomial x^i for each integer $0 \leq i \leq n$. Explicitly, the characteristic matrix $xI - A$ is already written as a formal polynomial with matrix coefficients: indeed, the degree-one "coefficient" is the identity matrix I, and the "constant term" is the matrix A. Even more, if we write $\chi(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ for some scalars $c_{n-1}, \ldots, c_1, c_0$, then the unique expression of $\chi(x)I$ as a formal polynomial with matrix coefficients is $\chi(x)I = x^n I + c_{n-1}x^{n-1}I + \cdots + c_1 xI + c_0 I$. Consider the unique $n \times n$ matrices $B_{n-1}, \ldots, B_1, B_0$

such that $\operatorname{adj}(xI - A) = x^{n-1}B_{n-1} + \cdots + xB_1 + B_0$. Expanding the left- and right-hand sides of the identity $\operatorname{adj}(xI - A)(xI - A) = \chi(x)I$ according to our formal polynomial factorizations, we find that $(x^{n-1}B_{n-1} + \cdots + xB_1 + B_0)(xI - A) = x^n I + c_{n-1}x^{n-1}I + \cdots + c_1 xI + c_0 I$. Expanding the product on the left-hand side and comparing the terms with x^i, we obtain the following.

$$\begin{aligned} B_{n-1} &= I && \text{(the coefficient of } x^n\text{)} \\ B_{n-2} - B_{n-1}A &= c_{n-1}I && \text{(the coefficient of } x^{n-1}\text{)} \\ &\vdots \\ B_0 - B_1 A &= c_1 I && \text{(the coefficient of } x\text{)} \\ -B_0 A &= c_0 I && \text{(the constant term)} \end{aligned}$$

Crucially, we may now multiply each subsequent identity from bottom to top by A^i for the integer $0 \leq i \leq n$ corresponding to the monomial x^i to find the following identities.

$$\begin{aligned} B_{n-1}A^n &= A^n && \text{(the coefficient of } x^n\text{)} \\ B_{n-2}A^{n-1} - B_{n-1}A^n &= c_{n-1}A^{n-1} && \text{(the coefficient of } x^{n-1}\text{)} \\ &\vdots \\ B_0 A - B_1 A^2 &= c_1 A && \text{(the coefficient of } x\text{)} \\ -B_0 A &= c_0 I && \text{(the constant term)} \end{aligned}$$

Last, adding the left-hand column yields a telescoping sum that results in the zero matrix; however, the right-hand sums to the $n \times n$ matrix $A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = \chi(A)$. □

One immediate consequence of the Cayley-Hamilton Theorem is that for every $n \times n$ matrix A, there exists a unique **monic** polynomial $\mu_A(x)$ of least degree such that $\mu_A(A) = O$. We refer to this polynomial as the **minimal polynomial** of A. Explicitly, a monic polynomial is one whose leading coefficient is one. By the Cayley-Hamilton Theorem, the characteristic polynomial $\chi_A(x)$ of A is a monic polynomial satisfying that $\chi_A(A) = O$, hence there exists a monic polynomial with the desired property. Consequently, we can find a monic polynomial of least degree that annihilates A by the **Well-Ordering Principle** applied to the nonempty set of positive integers corresponding to the degree of monic polynomials that annihilate A. Even more, the uniqueness of the minimal polynomial comes from the fact that if we take two monic polynomials of least degree that both annihilate A, then each of the polynomials will divide the other, hence they must be equal.

Even if this line of argument is not immediately clear, what matters is the following.

Proposition 2.1.9. *Given any $n \times n$ matrix A, its minimal polynomial $\mu(x)$ divides every polynomial $p(x)$ such that $p(A) = O$. Consequently, the minimal polynomial of A must divide the characteristic polynomial of A, so it is either the characteristic polynomial of A or a proper factor of it.*

Proof. By the Division Algorithm for polynomials, there exist unique polynomials $q(x)$ and $r(x)$ such that $p(x) = q(x)\mu(x) + r(x)$ and the degree of $r(x)$ is strictly smaller than the degree of $\mu(x)$. By assumption, we have that $p(A) = O$. By definition of $\mu(x)$, we have that $\mu(A) = O$. Combined, these observations imply that $O = p(A) = q(A)\mu(A) + r(A) = q(A)O + r(A) = r(A)$. Consequently, we have found a polynomial $r(x)$ of lesser degree than $\mu(x)$ that annihilates A. Even more, if $r(x)$ is nonzero, then we may multiply by the multiplicative inverse of its leading coefficient to obtain a monic polynomial of lesser degree than $\mu(x)$ that annihilates A. Because this is impossible by the definition of $\mu(x)$, we conclude that $r(x)$ must be the zero polynomial so that $\mu(x)$ divides $p(x)$.

By the Cayley-Hamilton Theorem, the characteristic polynomial of A annihilates A, so it must be divisible by the minimal polynomial of A by the argument of the preceding paragraph. $\square$

Example 2.1.10. Consider the 2×2 matrix A of Examples 2.1.1 and 2.1.6. We proved previously that the characteristic polynomial of A is $\chi(x) = (x-3)(x+1)$; neither $x-3$ nor $x+1$ annihilates A by the previous example, hence we conclude by Proposition 2.1.9 that $\mu(x) = \chi(x)$.

Example 2.1.11. Consider the 3×3 matrix A of Examples 2.1.2 and 2.1.7. We proved previously that the characteristic polynomial of A is $\chi(x) = (x-1)(x-2)(x+1)$. Observe that none of $x-1$, $x-2$, or $x+1$ annihilate A by the previous example. Even more, $(x-1)(x-2)$ and $(x-1)(x+1)$ and $(x-2)(x+1)$ do not annihilate A. Consequently, we conclude by Proposition 2.1.9 that $\mu(x) = \chi(x)$.

Example 2.1.12. Consider the 3×3 zero matrix O. Observe that the characteristic polynomial of O is given by $\chi(x) = \det(xI - O) = \det(xI) = x^3 \det(I) = x^3$; however, the minimal polynomial of O is simply $\mu(x) = x$. Generally, this is similarly the case for all $n \times n$ zero matrices.

Even though the minimal polynomial of a matrix is not necessarily the characteristic polynomial of the matrix, we know by Proposition 2.1.9 that the minimal polynomial is always a factor of the characteristic polynomial. Consequently, the **roots** of the minimal polynomial are always among the roots of the characteristic polynomial. Explicitly, for any scalar c such that $\mu(c) = 0$, we must have that $\chi(c) = 0$. We refer to such a scalar c such that $\chi_A(c) = 0$ as an **eigenvalue** of A. We note that the eigenvalues of A are precisely those scalars such that $\det(cI - A) = 0$. Under this identification, we can drastically narrow down the possibilities for the minimal polynomial $\mu_A(x)$.

Proposition 2.1.13. *Given any $n \times n$ matrix A, the characteristic polynomial of A and the minimal polynomial of A have the same roots. Particularly, the minimal polynomial of A is divisible by every irreducible polynomial factor of the characteristic polynomial of A.*

Proof. We will prove that $\mu_A(c) = 0$ if and only if c is a characteristic value of A. By the Factor Theorem, if we assume that $\mu_A(c) = 0$, then $\mu_A(x) = (x-c)q(x)$ for some polynomial $q(x)$ of strictly lesser degree than $\mu_A(x)$. By definition of $\mu_A(x)$, we must have that $q(A)$ is nonzero. Consequently, we have that $O = \mu_A(A) = (A - cI)q(A)$, hence $cI - A$ cannot be invertible because its product with the nonzero matrix $-q(A)$ is the zero matrix. We conclude by Proposition 2.2.5 that c is a characteristic value of A. Conversely, if c is a characteristic value of A, then $cI - A$ is not invertible, hence there exists a nonzero $n \times n$ matrix B such that $(cI - A)B = O$ or $cB = AB$. Crucially, for any integer $1 \leq k \leq n$, we have that $A^k B = A^{k-1}(AB) = A^{k-1}(cB) = c(A^{k-1}B) = \cdots = c^k B$ by Propositions 1.3.21 and 1.3.22. Consequently, it follows that $O = OB = \mu_A(A)B = \mu_A(c)B$. Considering that $\mu_A(c)$ is a scalar and B is a nonzero matrix, this is only possible if $\mu_A(c) = 0$. $\square$

We summarize the content of Propositions 2.1.9 and 2.1.13 in the following algorithm.

Algorithm 2.1.14 (Computing the Characteristic and Minimal Polynomials)**.** Consider any $n \times n$ matrix A. Carry out the following steps to find the characteristic and minimal polynomials of A.

1.) Compute the characteristic matrix $xI - A$.

2.) Compute the characteristic polynomial $\chi_A(x) = \det(xI - A)$.

3.) Completely factor the characteristic polynomial, taking care to account for whether entries of the matrix A are real or complex; this affects the factorization of $\chi_A(x)$.

4.) List all possibilities for the minimal polynomial of A, taking care to account for the fact that $\mu_A(x)$ must be a factor of $\chi_A(x)$ with all of the same roots as $\chi_A(x)$.

Example 2.1.15. Consider the following 3×3 matrix A and its characteristic matrix $xI - A$.

$$A = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad xI - A = \begin{bmatrix} x+1 & 0 & 0 \\ 0 & x-1 & 0 \\ 0 & 0 & x+1 \end{bmatrix}$$

One can readily verify that $\chi(x) = (x+1)^2(x-1)$ is the characteristic polynomial of A. Consequently, by Proposition 2.1.13, we must have that $\mu(x) = \chi(x)$ or $\mu(x) = (x+1)(x-1) = x^2 - 1$. Considering that $A^2 = I$, it follows that $A^2 - I = O$ so that $\mu(x) = x^2 - 1$.

Example 2.1.16. Consider the following 3×3 matrix A and its characteristic matrix $xI - A$.

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

By definition, the characteristic polynomial of A is found by computing the following.

$$\begin{aligned} \chi(x) = \det(xI - A) &= (x-1)[(x-2)(x-3) - 6] + [-2(x-3) - 6] + [6 + 3(x-2)] \\ &= (x-1)(x^2 - 5x + 6 - 6) - (2x - 6 + 6) + (6 - 3x - 6) \\ &= (x-1)(x^2 - 5x) - 5x \\ &= x^3 - 6x^2 \end{aligned}$$

Considering that $\chi(x) = x^3 - 6x^2 = x^2(x-6)$, it follows that $\mu(x) = \chi(x)$ or $\mu(x) = x(x-6)$. We conclude that $\mu(x) = x(x-6)$ because $A(A - 6I) = O$, as the following calculation shows.

$$A(A - 6I) = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} -5 & 1 & 1 \\ 2 & -4 & 2 \\ 3 & 3 & -3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Explicitly, one need only check that the first row is zero because the second and third rows of $A(A - 6I)$ are merely a scalar multiple of the first row of $A(A - 6I)$ by definition of A.

2.2 Eigenvalues and Eigenvectors

Consider any $n \times n$ matrix A. Previously, we have established that the determinant of the characteristic matrix $xI - A$ corresponding to A is a monic polynomial $\chi_A(x) = \det(xI - A)$ of degree n called the characteristic polynomial of A; we refer to the roots of $\chi_A(x)$ as the **eigenvalues** of A.

Example 2.2.1. Consider the 2×2 matrix A of Example 2.1.1.

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

We showed that $\chi(x) = \det(xI - A) = (x-3)(x+1)$, hence the eigenvalues of A are -1 and 3.

Example 2.2.2. Consider the following real diagonal 3×3 matrix.

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

Observe that the characteristic matrix $xI - A$ corresponding to A is a diagonal matrix with diagonal components $x-a$, $x-b$, and $x-c$. One can readily verify that the determinant of a diagonal matrix is the product of its diagonal entries, hence we have that $\chi(x) = \det(xI - A) = (x-a)(x-b)(x-c)$. Consequently, the eigenvalues of A are simply the diagonal entries a, b, and c.

Example 2.2.3. Consider the 3×3 matrix A of Example 2.1.2.

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

Considering that $\chi(x) = \det(xI - A) = (x-1)(x-2)(x+1)$, the eigenvalues of A are -1, 1, and 2.

Example 2.2.4. Consider the 3×3 matrix A of Example 2.1.16.

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

We demonstrated previously that $\chi(x) = \det(xI - A) = x^2(x-6)$, hence the eigenvalues of A are 0 (with multiplicity two) and 6. We will soon return to this notion of multiplicity of eigenvalues.

Crucially, the eigenvalues of a square matrix A determine the invertibility of the characteristic matrix $xI - A$ for each fixed real number x. We outline this as follows.

Proposition 2.2.5. *Given any $n \times n$ matrix A, the following are equivalent.*

1.) *We have that $\chi_A(c) = 0$, i.e., the real number c is an eigenvalue of A.*

2.) *We have that $cI - A$ is not invertible.*

3.) *We have that $A\mathbf{v} = c\mathbf{v}$ for some nonzero vector $\mathbf{v}$ in real n-space.*

Proof. By definition of the characteristic polynomial $\chi_A(x)$ of A, it follows that $\chi_A(c) = 0$ if and only if $\det(cI - A) = 0$. By Proposition 1.11.10, this occurs if and only if $cI - A$ is not invertible, hence the first two criteria are equivalent. By the Rank Equation, there exists a nonzero vector $\mathbf{v}$ in real n-space such that $A\mathbf{v} = c\mathbf{v}$ if and only if $c\mathbf{v} - A\mathbf{v} = \mathbf{0}$ if and only if $(cI - A)\mathbf{v} = \mathbf{0}$ if and only if $\operatorname{nullity}(cI - A) > 0$ if and only if $\operatorname{rank}(cI - A) < n$. Consequently, Proposition 1.8.13 yields the equivalence of criteria (2.) and (3.): $\operatorname{rank}(cI - A) < n$ if and only if $cI - A$ is not invertible. □

Bearing in mind the content of Proposition 2.2.5, we say that a nonzero vector $\mathbf{v}$ in real n-space is an **eigenvector** of an $n \times n$ matrix A corresponding to the eigenvalue c if and only if $A\mathbf{v} = c\mathbf{v}$.

Example 2.2.6. Consider the 2×2 matrix A of Example 2.2.1.

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

We showed that the eigenvalues of A are -1 and 3. We claim that $\mathbf{v}_1 = [-1, 1]$ is an eigenvector of A corresponding to the eigenvalue $c_1 = -1$ and $\mathbf{v}_2 = [1, 1]$ is an eigenvector of A corresponding to the eigenvalue $c_2 = 3$. We bear this out by showing that $A\mathbf{v}_i = c_i\mathbf{v}_i$ for each vector and eigenvalue.

Example 2.2.7. Consider the real diagonal 3×3 matrix of Example 2.2.2.

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

We showed that the eigenvalues of A are simply the diagonal entries $c_1 = a$, $c_2 = b$, and $c_3 = c$. We claim that $\mathbf{v}_i = \mathbf{e}_i$ is an eigenvector of A corresponding to the eigenvalue c_i.

$$A\mathbf{e}_1 = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = a\mathbf{e}_1$$

$$A\mathbf{e}_2 = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ b \\ 0 \end{bmatrix} = b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = b\mathbf{e}_2$$

$$A\mathbf{e}_3 = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ c \end{bmatrix} = c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = c\mathbf{e}_3$$

Example 2.2.8. Consider the 3×3 matrix A of Example 2.2.3.

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

We showed that the eigenvalues of A are -1, 1, and 2. We claim that $\mathbf{v}_1 = [1, -2, 1]$ is an eigenvector of A corresponding to the eigenvalue $c_1 = -1$; $\mathbf{v}_2 = [-1, 0, 1]$ is an eigenvector of A corresponding to the eigenvalue $c_2 = 1$; and $\mathbf{v}_3 = [1, 1, 1]$ is an eigenvector of A corresponding to $c_3 = 2$.

Example 2.2.9. Consider the 3×3 matrix A of Example 2.2.4.

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

We showed that the eigenvalues of A are 0 and 6. We claim that $\mathbf{v}_1 = [-1, 1, 0]$ and $\mathbf{v}_2 = [-1, 0, 1]$ are both eigenvectors of A corresponding to the eigenvalue $c_1 = 0$ and $\mathbf{v}_3 = [1, 2, 3]$ is an eigenvector of A corresponding to the eigenvalue $c_2 = 6$. Be sure to notice that $c_1 = 0$ has two eigenvectors!

Crucially, eigenvectors corresponding to eigenvalues are unique in the following sense.

Proposition 2.2.10 (Uniqueness of Eigenvalues Corresponding to Eigenvectors)**.** *If $\mathbf{v}$ is an eigenvector of an $n \times n$ matrix A corresponding to an eigenvalue α, then α is uniquely determined by its eigenvector $\mathbf{v}$ in the sense that if $Av = \beta \mathbf{v}$ for any scalar β, then we must have that $\beta = \alpha$.*

Proof. On the contrary, we will assume that α and β are distinct scalars. Consequently, we have that $\alpha - \beta$ is a nonzero scalar. By assumption that $A\mathbf{v} = \beta \mathbf{v}$ and by hypothesis that $\mathbf{v}$ is an eigenvector of A corresponding to the eigenvalue α, we have that $\alpha \mathbf{v} = A\mathbf{v} = \beta \mathbf{v}$ so that $\alpha \mathbf{v} - \beta \mathbf{v} = \mathbf{0}$ and $(\alpha - \beta)\mathbf{v} = \mathbf{0}$. Considering that $\alpha - \beta$ is a nonzero scalar, we can multiply both sides of this equation by $(\alpha - \beta)^{-1}$ to obtain that $\mathbf{v} = \mathbf{0}$. But this is impossible: by hypothesis that $\mathbf{v}$ is an eigenvector corresponding to α, we must have that $\mathbf{v}$ is a nonzero vector by definition of an eigenvector. $\square$

Consequently, if a nonzero vector $\mathbf{v}$ of real n-space corresponds to an eigenvalue α, then α is uniquely determined by $\mathbf{v}$, and there cannot exist another scalar β such that $A\mathbf{v} = \beta \mathbf{v}$. Once we have found the eigenvalues of a matrix by computing the roots of its characteristic polynomial, the hunt is on to determine the eigenvectors of A corresponding to these eigenvalues. We remind the reader that if c is an eigenvalue of an $n \times n$ matrix A, then by Proposition 2.2.5, the eigenvectors of A corresponding to the eigenvalue c of A are simply the vectors $\mathbf{v}$ in real n-space satisfying that $A\mathbf{v} = c\mathbf{v}$. Consequently, in practice, the way to find the eigenvectors of an $n \times n$ matrix A corresponding to an eigenvalue c of A is to solve the matrix equation $(cI - A)\mathbf{v} = \mathbf{0}$.

Algorithm 2.2.11 (Constructing the Eigenvalues and Eigenvectors of a Matrix)**.** Consider any $n \times n$ matrix A. Carry out the following steps to find the eigenvalues and eigenvectors of A.

1.) Construct the characteristic matrix $xI - A$.

2.) Construct the characteristic polynomial $\chi_A(x) = \det(xI - A)$.

3.) Compute the roots $c_1, \dots, c_n$ of the characteristic polynomial $\chi_A(x)$. By definition, these roots are the eigenvalues of the matrix A. Be sure to note whether A is a real or complex matrix; this will determine the possible eigenvalues of A as well as the minimal polynomial of A.

4.) Compute the null space of $c_iI - A$ for each distinct eigenvalue c_i of A. By definition, the basis vectors of $\operatorname{null}(c_iI - A)$ are the distinct eigenvectors of A corresponding to the eigenvalue c_i.

Considering that $\operatorname{null}(cI - A)$ is a subspace of real n-space by Example 1.6.4, we refer to the dimension of $\operatorname{null}(cI - A)$ as the **geometric multiplicity** of the eigenvalue c of A. Going forward, we will distinguish the subspace $\operatorname{null}(cI - A)$ as the **eigenspace** of the eigenvalue c.

Example 2.2.12. Consider the 2×2 matrix A of Example 2.2.2 with eigenvalues -1 and 3.

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

By definition, the eigenvectors of A corresponding to the eigenvalue $c_1 = -1$ are the solutions of the homogeneous equation $(-I - A)\mathbf{v} = \mathbf{0}$, so we may reduce $-I - A$ to row echelon form.

$$-I - A = \begin{bmatrix} -2 & -2 \\ -2 & -2 \end{bmatrix} \overset{R_2 - R_1 \mapsto R_2}{\sim} \begin{bmatrix} -2 & -2 \\ 0 & 0 \end{bmatrix}$$

By viewing the first row of the above matrix as a linear equation $-2x_1 - 2x_2 = 0$, it follows that $x_2 = -x_1$ and x_1 is a free variable. We conclude that the eigenspace of $c_1 = -1$ satisfies that

$$\operatorname{null}(-I - A) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}.$$

Likewise, the eigenspace of the eigenvalue $c_2 = 3$ can be determined as follows.

$$3I - A = \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix} \overset{R_2 + R_1 \mapsto R_2}{\sim} \begin{bmatrix} 2 & -2 \\ 0 & 0 \end{bmatrix}$$

By viewing the first row of the above matrix as a linear equation $2x_1 - 2x_2 = 0$, it follows that $x_2 = x_1$ and x_1 is a free variable. We conclude that the eigenspace of $c_2 = 3$ satisfies that

$$\operatorname{null}(-I - A) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Consequently, $\mathbf{v}_1 = \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$ is an eigenvector for $c_1 = -1$ and $\mathbf{v}_2 = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ is an eigenvector for $c_2 = 3$.

Example 2.2.13. Consider the 3×3 matrix A of Example 2.2.3 with eigenvalues -1, 1, and 2.

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

We reduce the matrices $-I - A$, $I - A$, and $2I - A$ to row echelon form to determine the eigenspaces.

$$-I - A = \begin{bmatrix} -2 & -1 & 0 \\ -1 & -1 & -1 \\ 0 & -1 & -2 \end{bmatrix} \overset{R_2 - \frac{1}{2}R_1 \mapsto R_2}{\sim} \begin{bmatrix} -2 & -1 & 0 \\ 0 & -\frac{1}{2} & -1 \\ 0 & -1 & -2 \end{bmatrix} \overset{R_3 - 2R_2 \mapsto R_3}{\sim} \begin{bmatrix} -2 & -1 & 0 \\ 0 & -\frac{1}{2} & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$I - A = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 0 \end{bmatrix} \overset{\substack{R_2 + R_1 \mapsto R_2 \\ R_3 - R_1 \mapsto R_3}}{\sim} \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$2I - A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \overset{R_2 + R_1 \mapsto R_2}{\sim} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix} \overset{R_3 + R_2 \mapsto R_3}{\sim} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

Each of the above matrices corresponds to a homogeneous system of linear equations that is constructed by reading the rows of the matrices. Explicitly, the rows of the matrix corresponding to $-I - A$ satisfy that $-2x_1 - x_2 = 0$ and $-\frac{1}{2}x_2 - x_3 = 0$. Consequently, we find that $x_1 = -\frac{1}{2}x_2$ and $x_3 = -\frac{1}{2}x_2$ and x_2 is a free variable. By judiciously setting $x_2 = 2$, we obtain the eigenspace

$$\operatorname{null}(-I - A) = \operatorname{span}\left\{\begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}\right\} \text{ so that } \mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \text{ is an eigenvector corresponding to } c_1 = -1.$$

Likewise, the matrix corresponding to $I - A$ induces the equations $-x_2 = 0$ and $-x_1 - x_3 = 0$, from which it follows that $x_2 = 0$, $x_1 = -x_3$, and x_3 is a free variable. We obtain the eigenspace

$$\operatorname{null}(I - A) = \operatorname{span}\left\{\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\} \text{ so that } \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ is an eigenvector corresponding to } c_2 = 1.$$

Last, the matrix corresponding to $2I - A$ yields the equations $x_1 - x_2 = 0$ and $x_2 - x_3 = 0$, hence $x_1 = x_2$, $x_3 = x_2$, and x_2 is a free variable. By choosing that $x_2 = 1$, we obtain the eigenspace

$$\operatorname{null}(2I - A) = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\} \text{ so that } \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ is an eigenvector corresponding to } c_3 = 2.$$

Example 2.2.14. Consider the 3×3 matrix A of Example 2.2.4 with eigenvalues 0 and 6.

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

Even though the matrix at hand admits only two distinct eigenvalues (one with multiplicity two), there are in fact three distinct eigenvectors: indeed, the eigenvalue $c_1 = 0$ induces an eigenspace $\operatorname{null}(0I - A)$ of dimension two as follows, hence the geometric multiplicity of $c_1 = 0$ is two.

$$0I - A = \begin{bmatrix} -1 & -1 & -1 \\ -2 & -2 & -2 \\ -3 & -3 & -3 \end{bmatrix} \overset{\substack{R_2 - 2R_1 \mapsto R_2 \\ R_3 - 3R_1 \mapsto R_3 \\ -R_1 \mapsto R_1}}{\sim} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$6I - A = \begin{bmatrix} 5 & -1 & -1 \\ -2 & 4 & -2 \\ -3 & -3 & 3 \end{bmatrix} \overset{\substack{R_3 + R_1 \mapsto R_3 \\ R_3 + R_2 \mapsto R_3}}{\sim} \begin{bmatrix} 5 & -1 & -1 \\ -2 & 4 & -2 \\ 0 & 0 & 0 \end{bmatrix} \overset{R_2 + \frac{2}{5}R_1 \mapsto R_2}{\sim} \begin{bmatrix} 5 & -1 & -1 \\ 0 & \frac{18}{5} & -\frac{12}{5} \\ 0 & 0 & 0 \end{bmatrix}$$

Consequently, for the above matrix corresponding to $0I - A$, we have that $x_1 + x_2 + x_3 = 0$ so that $x_1 = -x_2 - x_3$ and x_2 and x_3 are free variables; thus, every vector of $\operatorname{null}(0I - A)$ is of the form

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_2 - x_3 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ so that } \operatorname{null}(0I - A) = \operatorname{span}\left\{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}.$$

On the other hand, the matrix corresponding to $6I - A$ yields the equations $5x_1 - x_2 - x_3 = 0$ and $18x_2 - 12x_3 = 0$ so that $x_2 = \frac{2}{3}x_3$, $x_1 = \frac{1}{5}x_2 + \frac{1}{5}x_3 = \frac{1}{3}x_3$, and x_3 is a free variable. By judiciously choosing $x_3 = 3$, we obtain the eigenspace of A corresponding to $c_2 = 6$

$$\operatorname{null}(6I - A) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\}.$$

Crucially, we note that in this example, the algebraic multiplicity of the eigenvalue $c_1 = 0$ in the characteristic polynomial of A is two, and the geometric multiplicity of the eigenspace $\operatorname{null}(0I - A)$ corresponding to the eigenvalue 0 is also two. We will soon investigate this phenomenon further.

2.3 Diagonalization

Our ultimate objective throughout this chapter is to study the **canonical forms** of an $n \times n$ matrix A for some positive integer n. Put simply, a canonical form of a matrix is any representation of the matrix by a **similar** matrix that is (in a strict sense) in "simplest form." By definition, we say that a pair of $n \times n$ matrices A and B are similar provided that there exists an invertible matrix P such that $B = PAP^{-1}$. One of the most immediate and delightful canonical forms occurs when there exists a basis of eigenvectors for the matrix A. Explicitly, if the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ form a basis for real n-space, the best-case scenario is that $\mathbf{v}_1, \dots, \mathbf{v}_n$ are in fact eigenvectors of A corresponding to distinct eigenvalues $c_1, \dots, c_n$, respectively: indeed, in this case, we have that $A\mathbf{v}_i = c_i \mathbf{v}_i$ for each integer $1 \leq i \leq n$ by Proposition 2.2.5. Consider the real matrix P whose ith column is the eigenvector $\mathbf{v}_i$. By Proposition 2.3.2, we have that P is invertible so that

$$\begin{aligned} P^{-1}AP &= P^{-1}\begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{bmatrix} \\ &= P^{-1}\begin{bmatrix} c_1\mathbf{v}_1 & c_2\mathbf{v}_2 & \cdots & c_n\mathbf{v}_n \end{bmatrix} \\ &= \begin{bmatrix} c_1P^{-1}\mathbf{v}_1 & c_2P^{-1}\mathbf{v}_2 & \cdots & c_nP^{-1}\mathbf{v}_n \end{bmatrix} \\ &= \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix}. \end{aligned}$$

Consequently, A is similar to a diagonal matrix whose diagonal entries are the eigenvalues of A! We refer to the diagonal matrix $P^{-1}AP = \operatorname{diag}\{c_1, c_2, \dots, c_n\}$ as the **diagonalization** of A.

Definition 2.3.1. We say that a real $n \times n$ matrix A is **diagonalizable** if there exists an ordered basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of real n-space such that $A\mathbf{v}_i = c_i\mathbf{v}_i$ for some scalars $c_1, \dots, c_n$. Put another way, a real $n \times n$ matrix is diagonalizable if and only if there exists a basis of eigenvectors for A if and only if A can be represented by a diagonal matrix with respect to some ordered basis of real n-space.

Our first order of business is to provide a necessary and sufficient condition for the diagonalizability of a real $n \times n$ matrix. Like we mentioned in the exposition preceding the above definition, one starting point is that eigenvectors corresponding to distinct eigenvalues are linearly independent.

Proposition 2.3.2 (Eigenvectors Belonging to Distinct Eigenvalues Are Linearly Independent)**.** *Consider any $n \times n$ matrix A. If $\mathbf{v}_1, \dots, \mathbf{v}_k$ are any eigenvectors of A corresponding respectively to the distinct eigenvalues $c_1, \dots, c_k$ of A, then $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent.*

Proof. We proceed by induction on the number k of eigenvectors present. By definition, if c_1 is an eigenvalue of A corresponding to the eigenvector $\mathbf{v}_1$ of A, then $\mathbf{v}_1$ is a nonzero vector, hence $\mathbf{v}_1$ is linearly independent. Consider any eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ of A corresponding respectively to the pairwise distinct eigenvalues $c_1, \dots, c_k$ of A. We must show that if $a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$, then $a_1 = \cdots = a_k = 0$. Observe that if we apply A to the above relation of linear dependence, then

$$\mathbf{0} = A\mathbf{0} = A(a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k) = a_1A\mathbf{v}_1 + \cdots + a_kA\mathbf{v}_k = a_1c_1\mathbf{v}_1 + \cdots + a_kc_k\mathbf{v}_k \tag{2.3.1}$$

by assumption that $\mathbf{v}_i$ is an eigenvector of A corresponding to the eigenvalue c_i of A. On the other hand, if we multiply our original relation of linear dependence by c_1, then we find that

$$\mathbf{0} = c_1\mathbf{0} = c_1(a_1\mathbf{v}_1 + \cdots + a_k\mathbf{v}_k) = a_1c_1\mathbf{v}_1 + \cdots + a_kc_1\mathbf{v}_k. \tag{2.3.2}$$

By subtracting Equation (2.3.2) from Equation (2.3.1), we obtain a third equation

$$\mathbf{0} = a_2(c_1 - c_2)\mathbf{v}_2 + \cdots + a_k(c_1 - c_k)\mathbf{v}_k.$$

By induction, these $k - 1$ vectors are linearly independent, hence we conclude that $a_i(c_1 - c_i) = 0$ for each integer $2 \leq i \leq k$. Considering that c_1 and c_i are distinct eigenvalues for each integer $2 \leq i \leq k$, we must have that $c_1 - c_i$ is nonzero. Cancelling the factor of $c_1 - c_i$ from each identity $a_i(c_1 - c_i) = 0$ yields that $a_2 = \cdots = a_k = 0$, so our original relation of linear independence now states that $a_1\mathbf{v}_1 = \mathbf{0}$. But this implies that $a_1 = 0$ because $\mathbf{v}_1$ is nonzero by hypothesis. □

Corollary 2.3.3 (Every Real Matrix with Distinct Eigenvalues Is Diagonalizable)**.** *Consider any real $n \times n$ matrix A. Each of the following conditions guarantees that A is diagonalizable.*

1.) *The matrix A admits n distinct eigenvalues.*

2.) *The matrix A admits n linearly independent eigenvectors.*

3.) *The characteristic polynomial $\chi_A(x)$ splits into distinct linear factors.*

Proof. By the fourth part of the Fundamental Theorem of Subspaces of Real n-Space, every collection of n linearly independent vectors of real n-space form a basis for real n-space. By Proposition 2.3.2, eigenvectors corresponding to distinct eigenvalues are linearly independent, hence any collection of n eigenvectors corresponding to n distinct eigenvalues form a basis for real n-space By Definition 2.3.1, we conclude that A is diagonalizable. Even more, its matrix representation with respect to any ordered basis of eigenvectors of A corresponding to distinct eigenvalues of A is a diagonal matrix. Consequently, if A admits n distinct eigenvalues, then A must be diagonalizable because in this case, each of the n distinct eigenvalues of A corresponds to an eigenvector of A, i.e., there are n linearly independent eigenvectors of A. Last, the eigenvalues of A are the roots of the characteristic polynomial $\chi_A(x)$, and the roots of $\chi_A(x)$ are determined by its linear factors. □

Example 2.3.4. Consider the following 2×2 matrix A of Example 2.2.1.

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

We proved in that example that the eigenvalues of A are -1 and 3, hence by Corollary 2.3.3, we conclude that A is diagonalizable because it admits $n = 2$ distinct eigenvalues.

Example 2.3.5. Consider the following 3×3 matrix of Example 2.2.2.

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

We demonstrated previously that the eigenvalues of A are a, b, and c corresponding to the respective eigenvectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$. Consequently, A is diagonalizable. Of course, we did not need Corollary 2.3.3 to deduce this fact; we could have simply looked at the diagonal matrix A.

Example 2.3.6. Consider the following 3×3 matrix A of Example 2.2.3.

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

We demonstrated in the aforementioned example that A admits $n = 3$ distinct eigenvalues -1, 1, and 2, hence by Corollary 2.3.3, it follows that A is diagonalizable.

Example 2.3.7. Consider the following 3×3 matrix A of Example 2.2.4.

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

Even though A admits only two distinct eigenvalues 0 and 6, it turns out that A is diagonalizable! By Example 2.2.9, the eigenspace of A corresponding to the eigenvalue 0 has dimension two since $\operatorname{nullity}(0I - A) = 2$. Consequently, we must have that $\operatorname{nullity}(6I - A) = 1$ because the vectors in $\operatorname{null}(0I - A)$ and $\operatorname{null}(6I - A)$ are linearly independent and the dimension of real 3-space is three. We conclude that the two eigenvectors that span $\operatorname{null}(0I - A)$ and the one eigenvector that spans $\operatorname{null}(6I - A)$ comprise a basis for real 3-space, hence A is diagonalizable by Definition 2.3.1.

Example 2.3.7 illustrates that the conditions of Corollary 2.3.3 are sufficient but not necessary for the diagonalizability of A. Consequently, we seek more restrictive properties of A under which A is diagonalizable and for which A is not diagonalizable if the properties are not satisfied. Later in the course, we will discuss the mechanisms behind the following, but we omit the proof here.

Theorem 2.3.8 (Equivalent Conditions for Diagonalizability)**.** *Given any real $n \times n$ matrix A with distinct eigenvalues $c_1, \dots, c_k$, we have that A is diagonalizable if and only if*

1.) $\chi_A(x) = (x - c_1)^{e_1} \cdots (x - c_k)^{e_k}$ *and* $\operatorname{nullity}(c_i I - A) = e_i$ *for each integer* $1 \leq i \leq k$ *or*

2.) $\operatorname{nullity}(c_1 I - A) + \cdots + \operatorname{nullity}(c_k I - A) = \#(\textit{rows of } A)$.

Compiling the content of Corollary 2.3.3 and Theorem 2.3.8 yields the following.

Theorem 2.3.9 (Criteria for Diagonalizability)**.** *Consider any real $n \times n$ matrix A with distinct eigenvalues $c_1, \dots, c_k$. Given that any of the following criteria hold, the matrix A is diagonalizable.*

1.) *We have that $k = n$, i.e., the matrix A admits n distinct eigenvalues.*

2.) *We may factor the characteristic polynomial $\chi_A(x)$ into distinct linear factors.*

3.) *We may factor the characteristic polynomial $\chi_A(x) = (x - c_1)^{e_1} \cdots (x - c_k)^{e_k}$ into powers of distinct linear factors in such a manner that the algebraic multiplicity e_i of the eigenvalue c_i coincides with its geometric multiplicity $\operatorname{nullity}(c_i I - A)$ for each integer $1 \leq i \leq k$.*

4.) *We have that $\operatorname{nullity}(c_1 I - A) + \cdots + \operatorname{nullity}(c_k I - A) = \#(\textit{rows of } A) = \#(\textit{columns of } A)$.*

We have only encountered diagonalizable matrices so far in this section; however, it is unfortunately not true that every matrix is diagonalizable, as exhibited in the following example.

Example 2.3.10. Consider the following real 3×3 matrix.

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

Even though this matrix looks quite harmless and inconspicuous, it turns out that it is not diagonalizable. Explicitly, the eigenvectors of A do not form a basis for real 3-space, as we demonstrate next. Observe that the characteristic matrix $xI - A$ is the following upper-triangular matrix.

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

Consequently, the characteristic polynomial of A is $\chi(x) = x^3$ so that $c = 0$ is the only eigenvalue of A with algebraic multiplicity three. By the Criteria for Diagonalizability, A is diagonalizable if and only if the geometric multiplicity of $c = 0$ is three if and only if $\operatorname{nullity}(0I - A) = 3$. But reading off the rows of the matrix $0I - A$ yields the linear equation $-x_2 = 0$, hence we find that $x_2 = 0$ and x_1 and x_3 are free variables. We conclude that a typical eigenvector of A is of the form

$$\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \text{ so that } \operatorname{null}(0I - A) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Considering that $\operatorname{nullity}(0I - A) = 2 < 3$, we conclude that A is not diagonalizable.

We will therefore benefit from the development of tools to study matrices that are not diagonalizable. One natural question is whether a matrix that is not diagonalizable admits some other canonical form. Even though the matrix of Example 2.3.10 is not diagonalizable, it happens to be upper-triangular. Explicitly, an **upper-triangular** matrix is a square matrix whose entries below the main diagonal are zero. Conversely, a matrix is **lower-triangular** if it is the transpose of an upper-triangular matrix. Considering that the determinant of a matrix is equal to the determinant of its transpose and the characteristic polynomial of a matrix is therefore equal to the characteristic polynomial of its transpose, we may fix our attention on upper-triangular matrices.

One of the foremost features of such matrices is that the determinant of an upper-triangular matrix is the product of its diagonal entries; this affords a simple way to compute the characteristic polynomial of an upper-triangular matrix since its characteristic matrix is upper-triangular.

Proposition 2.3.11. *The determinant of a triangular matrix is the product of its diagonal entries.*

Proof. Considering that a lower-triangular matrix is the transpose of an upper-triangular matrix and the determinant of a matrix is equal to the determinant of its transpose by Proposition 1.10.12, we may prove the claim for upper-triangular matrices. We proceed by induction on the size n of an $n \times n$ upper-triangular matrix A. Every 2×2 diagonal matrix is of the following form.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}$$

Consequently, we have that $\det(A) = a_{11}a_{22}$, as desired. We will assume by induction that the claim holds for $(n-1) \times (n-1)$ upper-triangular matrices. Consider the following $n \times n$ matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

Expanding the determinant along the first column, we obtain the following identity.

$$\det(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix}$$

Considering that the determinant on the right-hand side is taken from an $(n-1) \times (n-1)$ matrix, it follows by our inductive hypothesis that $\det(A) = a_{11}a_{22} \cdots a_{nn}$ is the product of the diagonal. $\square$

Corollary 2.3.12. *Given any triangular $n \times n$ matrix A whose diagonal entries are $a_1, \dots, a_n$, the characteristic polynomial of A is given by $\chi_A(x) = (x - a_1) \cdots (x - a_n)$.*

Proof. Considering that xI is a diagonal matrix, it follows that $xI - A$ is a triangular matrix because the difference does not affect any components of A other than those lying on the diagonal of A. Observe that the diagonal components of $xI - A$ are simply the linear polynomials $x - a_1, \dots, x - a_n$, hence by Proposition 2.3.11, we conclude that $\chi_A(x) = \det(xI - A) = (x - a_1) \cdots (x - a_n)$. $\square$

We will soon return to the explore the ubiquity of upper-triangular matrices. Explicitly, we will demonstrate that every real $n \times n$ matrix admits a matrix representation that is upper-triangular! Put another way, if A is a real $n \times n$ matrix, then there exists an invertible matrix P such that $P^{-1}AP$ is upper-triangular. Considering the utility of such a matrix representation, we will spend a considerable amount of time working toward the construction of these matrices. We adopt the approach in this course of using another canonical form to construct this upper-triangular matrix.

2.4 Smith Normal Form

We turn our attention next to an indispensable tool in the theory of canonical forms for matrices. Explicitly, we will construct a canonical form for the characteristic matrix $xI - A$ of a real $n \times n$ matrix that will allow us to determine the minimal polynomial and characteristic polynomial of A.

Theorem 2.4.1 (Smith Normal Form)**.** *Given any real $n \times n$ matrix A and any indeterminate x, there exist invertible real $n \times n$ matrices P and Q and polynomials $p_1(x), p_2(x), \dots, p_\ell(x)$ such that*

$$P(xI - A)Q = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & p_1(x) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & p_2(x) & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & p_\ell(x) \end{bmatrix}$$

and the polynomials $p_i(x)$ are unique (up to sign) and satisfy that $p_1(x) \mid p_2(x) \mid \cdots \mid p_\ell(x)$. Even more, the non-constant polynomials are called **invariant factors***; the minimal polynomial of A is the largest invariant factor $p_\ell(x)$; and the characteristic polynomial of A is $p_1(x)p_2(x)\cdots p_\ell(x)$.*

Computing the Smith Normal Form for the characteristic matrix $xI - A$ of a real $n \times n$ matrix A amounts to carrying out some elementary row operations and elementary column operations on $xI - A$ to reduce the given matrix to the desired form. Explicitly, we will find that the invertible $n \times n$ matrix P is obtained from the $n \times n$ identity matrix by performing the specified elementary row operations on $xI - A$; likewise, the invertible $n \times n$ matrix Q is obtained from the $n \times n$ identity matrix by performing the specified elementary column operations on $xI - A$. We note that there are three elementary row (or column) operations that are valid in this scenario.

Definition 2.4.2 (Elementary Row and Column Operations)**.** Each of the following polynomial arithmetic operations are permissible to perform on the characteristic matrix $xI - A$ of a real $n \times n$ matrix A in order to reduce $xI - A$ to its unique Smith Normal Form.

1.) We may multiply any row (or column) of the matrix by a nonzero real number a.

2.) We may add any polynomial multiple of a row (or column) to another row (or column).

3.) We may interchange any pair of rows (or columns) of the matrix.

We continue using the shorthand $R_i \mapsto aR_i$ to denote the operation of multiplying the ith row of the matrix by a; we will use the shorthand $R_j + p(x)R_i \mapsto R_j$ to denote the operation of adding a polynomial multiple $p(x)$ of the ith row of the matrix to the jth row of the matrix (for any distinct indices i and j); and we will use the shorthand $R_i \leftrightarrow R_j$ to denote the operation of interchanging the ith and jth rows of the matrix. Each of these elementary row operations can also be performed with the ith and jth columns C_i and C_j of the matrix for any pair of distinct indices i and j.

Example 2.4.3. Let us compute the Smith Normal Form for $xI - A$ of the following 2×2 matrix.

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

We will keep track of the elementary row operations by performing each such operation on the 2×2 identity matrix; likewise, we will keep track of the column operations by manipulating the columns of the 2×2 identity matrix according to the column operations on $xI - A$.

$$xI - A = \begin{bmatrix} x-1 & 0 \\ -1 & x+1 \end{bmatrix}$$

1.) $C_2 + (x+1)C_1 \mapsto C_2$ $\qquad xI - A \sim \begin{bmatrix} x-1 & (x-1)(x+1) \\ -1 & 0 \end{bmatrix} \qquad Q \sim \begin{bmatrix} 1 & x+1 \\ 0 & 1 \end{bmatrix}$

2.) $R_1 \leftrightarrow R_2$ $\qquad xI - A \sim \begin{bmatrix} -1 & 0 \\ x-1 & (x-1)(x+1) \end{bmatrix} \qquad P \sim \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

3.) $R_2 + (x-1)R_1 \mapsto R_2$ $\qquad xI - A \sim \begin{bmatrix} -1 & 0 \\ 0 & (x-1)(x+1) \end{bmatrix} \qquad P \sim \begin{bmatrix} 0 & 1 \\ 1 & x-1 \end{bmatrix}$

4.) $-R_1 \mapsto R_1$ $\qquad xI - A \sim \begin{bmatrix} 1 & 0 \\ 0 & (x-1)(x+1) \end{bmatrix} \qquad P \sim \begin{bmatrix} 0 & -1 \\ 1 & x-1 \end{bmatrix}$

Consequently, the Smith Normal Form for $xI - A$ and the invertible matrices P and Q are as follows.

$$\operatorname{SNF}(xI - A) = P(xI - A)Q = \begin{bmatrix} 0 & -1 \\ 1 & x-1 \end{bmatrix} \begin{bmatrix} x-1 & 0 \\ -1 & x+1 \end{bmatrix} \begin{bmatrix} 1 & x+1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & (x-1)(x+1) \end{bmatrix}$$

Even more, the only invariant factor of A is $(x-1)(x+1)$, hence we have that $\mu_A(x) = (x-1)(x+1)$ and $\chi_A(x) = (x-1)(x+1)$. Last, the **elementary divisors** of A are $x-1$ and $x+1$. Later, we will discuss the notion of the elementary divisors of a square matrix at greater length; however, for now, we remark that the elementary divisors can be determined as the largest powers of the distinct linear factors of the invariant factors, hence in this case, they are $x-1$ and $x+1$.

Example 2.4.4. Let us compute the Smith Normal Form for $xI - A$ of the following 2×2 matrix.

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

We will keep track of the elementary row operations by performing each such operation on the 2×2 identity matrix; likewise, we will keep track of the column operations by manipulating the columns of the 2×2 identity matrix according to the column operations on $xI - A$.

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

1.) $C_1 \leftrightarrow C_2$ $\qquad xI - A \sim \begin{bmatrix} -1 & x \\ x & 0 \end{bmatrix}$ $\qquad Q \sim \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

2.) $R_2 + xR_1 \mapsto R_2$ $\qquad xI - A \sim \begin{bmatrix} -1 & x \\ 0 & x^2 \end{bmatrix}$ $\qquad P \sim \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix}$

3.) $C_2 + xC_1 \mapsto C_2$ $\qquad xI - A \sim \begin{bmatrix} -1 & 0 \\ 0 & x^2 \end{bmatrix}$ $\qquad Q \sim \begin{bmatrix} 0 & 1 \\ 1 & x \end{bmatrix}$

4.) $-R_1 \mapsto R_1$ $\qquad xI - A \sim \begin{bmatrix} 1 & 0 \\ 0 & x^2 \end{bmatrix}$ $\qquad P \sim \begin{bmatrix} -1 & 0 \\ x & 1 \end{bmatrix}$

Consequently, the Smith Normal Form for $xI - A$ and the invertible matrices P and Q are as follows.

$$\text{SNF}(xI - A) = P(xI - A)Q = \begin{bmatrix} -1 & 0 \\ x & 1 \end{bmatrix} \begin{bmatrix} x & -1 \\ 0 & x \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & x \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & x^2 \end{bmatrix}$$

Even more, the only invariant factor of A is x^2, hence the minimal polynomial and the characteristic polynomial of A are $\mu_A(x) = x^2$ and $\chi_A(x) = x^2$. Last, the only elementary divisor of A is x^2.

Going forward into the case of 3×3 matrices, out of want for simplicity, we will not concern ourselves with keeping track of the matrices P and Q; however, we note that (somewhat miraculously) in order to determine the invertible matrix P that converts A to its **Rational Canonical Form** or **Jordan Canonical Form**, it suffices to keep track *only* of the elementary row operations.

Example 2.4.5. Let us compute the Smith Normal Form for $xI - A$ of the following 3×3 matrix.

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

We will keep track of the elementary row operations and often abbreviate column operations.

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

1.) $C_1 \leftrightarrow C_2$ $\qquad xI - A \sim \begin{bmatrix} -1 & x-1 & -1 \\ x-2 & -2 & -2 \\ -3 & -3 & x-3 \end{bmatrix}$

2.) $R_2 + (x-2)R_1 \mapsto R_2$ $$xI - A \sim \begin{bmatrix} -1 & x-1 & -1 \\ 0 & (x-1)(x-2)-2 & -(x-2)-2 \\ -3 & -3 & x-3 \end{bmatrix}$$

3.) $R_3 - 3R_1 \mapsto R_3$ $$xI - A \sim \begin{bmatrix} -1 & x-1 & -1 \\ 0 & (x-1)(x-2)-2 & -(x-2)-2 \\ 0 & -3(x-1)-3 & x \end{bmatrix}$$

Perform column operations and simplify the result. $$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & x(x-3) & -x \\ 0 & -3x & x \end{bmatrix}$$

4.) $C_2 + (x-3)C_3 \mapsto C_2$ $$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -x \\ 0 & -3x + x(x-3) & x \end{bmatrix}$$

5.) $R_3 + R_2 \mapsto R_3$ $$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -x \\ 0 & x(x-6) & 0 \end{bmatrix}$$

6.) $C_2 \leftrightarrow C_3$, $-C_2 \mapsto C_2$ $$xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x(x-6) \end{bmatrix}$$

We note that this last matrix is by definition the Smith Normal Form for $xI - A$. Consequently, the invariant factors of A are x and $x(x-6)$; the elementary divisors of A are x, x, and $x-6$; the minimal polynomial of A is $\mu_A(x) = x(x-6)$; and the characteristic polynomial of A is $\chi_A(x) = x^2(x-6)$.

Example 2.4.6. Let us compute the Smith Normal Form for $xI - A$ of the following 3×3 matrix.

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

We will keep track of the elementary row operations and often abbreviate column operations; however, it is possible here to get away almost entirely with using column operations.

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

1.) $R_3 + \frac{1}{2}(x-1)R_1 \mapsto R_3$ $\qquad xI - A \sim \begin{bmatrix} x-1 & 0 & -2 \\ 0 & x-1 & 0 \\ \frac{1}{2}(x-1)^2 & 0 & 0 \end{bmatrix}$

2.) $C_1 \leftrightarrow C_3$ $\qquad xI - A \sim \begin{bmatrix} -2 & 0 & x-1 \\ 0 & x-1 & 0 \\ 0 & 0 & \frac{1}{2}(x-1)^2 \end{bmatrix}$

3.) $-\frac{1}{2}C_1 \mapsto C_1$ $\qquad xI - A \sim \begin{bmatrix} 1 & 0 & x-1 \\ 0 & x-1 & 0 \\ 0 & 0 & \frac{1}{2}(x-1)^2 \end{bmatrix}$

4.) $C_3 - (x-1)C_3 \mapsto C_3$ $\qquad xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & x-1 & 0 \\ 0 & 0 & \frac{1}{2}(x-1)^2 \end{bmatrix}$

5.) $2C_3 \mapsto C_3$ $\qquad xI - A \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & x-1 & 0 \\ 0 & 0 & (x-1)^2 \end{bmatrix}$

We note that this last matrix is the Smith Normal Form for $xI - A$. Consequently, the invariant factors of A are $x-1$ and $(x-1)^2$; the elementary divisors of A are $x-1$ and $(x-1)^2$; the minimal polynomial of A is $\mu_A(x) = (x-1)^2$; and the characteristic polynomial of A is $\chi_A(x) = (x-1)^3$.

Example 2.4.7. Observe that the characteristic matrix of the $n \times n$ zero matrix O is simply the $n \times n$ matrix xI. Consequently, the Smith Normal Form for the characteristic matrix of the $n \times n$ zero matrix is the diagonal matrix consisting of n copies of x along the main diagonal. Particularly, the invariant factors and the elementary divisors of O are $x, x, \dots, x$ (n times); the minimal polynomial of O is $\mu_O(x) = x$; and the characteristic polynomial of O is $\chi_O(x) = x^n$.

Example 2.4.8. Observe that the characteristic matrix of the $n \times n$ identity matrix I is the matrix $(x-1)I$. Consequently, the Smith Normal Form for the characteristic matrix of the $n \times n$ identity matrix is the diagonal matrix consisting of n copies of $x-1$ along the main diagonal. Particularly, the invariant factors and the elementary divisors of I are $x-1, x-1, \dots, x-1$ (n times); the minimal polynomial of I is $\mu_I(x) = x-1$; and the characteristic polynomial of I is $\chi_I(x) = (x-1)^n$.

We will come to find that the Rational Canonical Form for A is built out of the invariant factors of A; similarly, the Jordan Canonical Form for A is built out of the **elementary divisors** of A. By definition, the elementary divisors of A are the powers of the irreducible polynomial factors of the invariant factors of A. We have tacitly used this fact already, but let us do some more examples.

Example 2.4.9. Given that the invariant factors of a matrix A are $x-1$ and $(x-1)(x-2)$, the elementary divisors of A must be $x-1$, $x-1$, and $x-2$; this must be a 3×3 matrix with minimal polynomial $\mu_A(x) = (x-1)(x-2)$ and characteristic polynomial $\chi_A(x) = (x-1)^2(x-2)$.

Example 2.4.10. Given that the invariant factors of a matrix A are x, x^2, and $x^3(x+1)^2$, the elementary divisors of A must be x, x^2, x^3, and $(x+1)^2$; this must be an 8×8 matrix with minimal polynomial $\mu_A(x) = x^3(x+1)^2$ and characteristic polynomial $\chi_A(x) = x^6(x+1)^2$.

Example 2.4.11. Observe that there cannot be a matrix with invariant factors $x-1$ and $x+1$ because neither of these linear polynomials divides the other. Explicitly, they have distinct roots.

We provide an algorithm for determining the elementary divisors from the invariant factors.

Algorithm 2.4.12 (Converting Invariant Factors to Elementary Divisors)**.** Let A be a real $n \times n$ matrix whose invariant factors are known. Use the following to find the elementary divisors of A.

1.) Given the invariant factors $p_i(x)$ with $p_1(x) \mid p_2(x) \mid \cdots \mid p_\ell(x)$, express each invariant factor $p_i(x)$ as a product of distinct prime powers of irreducible polynomials.

2.) Construct an upper-triangular array whose ith column consists of the distinct prime powers of irreducible polynomials $q_{i1}(x)^{e_{i1}}, \dots, q_{ik}(x)^{e_{ik}}$ such that $p_i(x) = q_{i1}(x)^{e_{i1}} \cdots q_{ik}(x)^{e_{ik}}$.

3.) We obtain the elementary divisors of A as the components of the upper-triangular array.

Example 2.4.13. By the previous algorithm, if A admits an invariant factor $x(x-1)^2(x^2+1)^3$, then the elementary divisors of A corresponding to this invariant factor are x, $(x-1)^2$, and $(x^2+1)^3$.

Conversely, it is possible to ask for the invariant factors from the elementary divisors. We provide an algorithm for this task; however, we note that it is slightly more delicate than the last.

Algorithm 2.4.14 (Converting Elementary Divisors to Invariant Factors)**.** Let A be a real $n \times n$ matrix whose elementary divisors are known. Use the following to find the invariant factors of A.

1.) Find the irreducible polynomial $p(x)$ that appears the most times among the elementary divisors of A. Choose one arbitrarily if more than one polynomial fits this criterion.

2.) Create an array whose first row consists of all powers of $p(x)$ that appear as elementary divisors of A, listing these powers in non-decreasing order from left to right.

3.) Repeat the second step in the second row with the irreducible polynomial $q(x)$ that appears the second most times among the elementary divisors of A.

4.) Continue this process until all irreducible polynomials appearing as elementary divisors of A have been written in a row. One should end with an upper-triangular array.

5.) By multiplying the elements of each consecutive column, we obtain the invariant factors of A.

Example 2.4.15. Given that the elementary divisors of a matrix A are x, x, x^2, x^3, $x-1$, x^2+1, and x^2+1, the previous algorithm leads us to the following upper-triangular array.

$$\begin{array}{cccc} x & x & x^2 & x^3 \\ & & x^2+1 & x^2+1 \\ & & & x-1 \end{array}$$

Consequently, the invariant factors of A are the products of the columns of this array, i.e., they are x, x, $x^2(x^2+1)$, and $x^3(x-1)(x^2+1)$. We conclude that A is a 12×12 matrix with minimal polynomial $\mu_A(x) = x^3(x-1)(x^2+1)$ and characteristic polynomial $\chi_A(x) = x^7(x-1)(x^2+1)^2$.

Example 2.4.16. Given that the elementary divisors of a matrix A are x^2, x^2, x^2+x+1, and x^2+x+1, the previous algorithm leads us to the following upper-triangular array.

$$\begin{array}{cc} x^2 & x^2 \\ x^2+x+1 & x^2+x+1 \end{array}$$

Consequently, the invariant factors of A are the products of the columns of this array, i.e., they are $x^2(x^2+x+1)$ and $x^2(x^2+x+1)$. We conclude that A is an 8×8 matrix with minimal polynomial $\mu_A(x) = x^2(x^2+x+1)$ and characteristic polynomial $\chi_A(x) = x^4(x^2+x+1)^2$.

Example 2.4.17. Observe that there cannot be a 3×3 matrix with elementary divisors x^2 and x^2 because this would force the characteristic polynomial to be x^4, and this is impossible.

Example 2.4.18. Likewise, there cannot be any 3×3 matrices with elementary divisors x and x because this would force the characteristic polynomial to be x^2, and this is impossible.

2.5 Rational Canonical Form

Last section, we defined the Smith Normal Form of the characteristic matrix of a real $n \times n$ matrix. Essentially, the Smith Normal Form provides a generalization of the reduced row echelon form of a matrix with entries that do not belong to a **field**: polynomials do not admit multiplicative inverses, so a matrix whose entries consist of polynomials might not admit a typical reduced row echelon form consisting of zeros and ones; however, the Smith Normal Form guarantees that every such matrix can be placed in a unique diagonal form consisting of ones and polynomials along the diagonal in such a manner that each of the non-constant polynomials divides the next. Even more, the Smith Normal Form provides the invariant factors and elementary divisors of a real $n \times n$ matrix. We will see throughout this section and the next that this information leads to canonical forms that are (in a strict sense) "simplest" and from which the properties of a matrix can be easily deduced.

Given any monic polynomial $p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ of degree n, we define the **companion matrix** of the polynomial $p(x)$ as the following $n \times n$ matrix.

$$C_{p(x)} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & -\alpha_1 \\ 0 & 1 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\alpha_{n-1} \end{bmatrix}$$

Example 2.5.1. Observe that the companion matrix of any linear polynomial $x+c$ is $\begin{bmatrix} -c \end{bmatrix}$. Explicitly, the companion matrix of x is $\begin{bmatrix} 0 \end{bmatrix}$, and the companion matrix of $x-1$ is $\begin{bmatrix} 1 \end{bmatrix}$.

Example 2.5.2. Observe that the companion matrix of any quadratic polynomial x^2+ax+b is

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

Explicitly, the companion matrix of x^2+1 is given as follows.

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

Likewise, the companion matrix of $x^2 + x + 1$ is the following.

$$\begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$$

Crucially, the characteristic polynomial and minimal polynomial of the companion matrix of a monic polynomial $p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ are both simply $p(x)$.

Proposition 2.5.3. *Consider any monic polynomial $p(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0$ of positive degree n and the following companion matrix $C_{p(x)}$ of the polynomial $p(x)$.*

$$C_{p(x)} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\alpha_0 \\ 1 & 0 & \cdots & 0 & -\alpha_1 \\ 0 & 1 & \cdots & 0 & -\alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\alpha_{n-1} \end{bmatrix}$$

Both the characteristic polynomial and the minimal polynomial of $C_{p(x)}$ are equal to $p(x)$.

Proof. We will prove that the characteristic polynomial of $C_{p(x)}$ is equal to $p(x)$. We proceed by induction on the degree n of $p(x)$. Certainly, if $n = 1$, then the companion matrix of $p(x) = x + \alpha_0$ is annihilated by $p(x)$ because it holds that $C_{p(x)} = \begin{bmatrix} -\alpha_0 \end{bmatrix}$ so that $p(C_{p(x)}) = C_{p(x)} + \alpha_0 I = O$. We conclude in this case that $p(x)$ is the minimal polynomial of $C_{p(x)}$ by Proposition 2.1.9, hence it is the characteristic polynomial by Proposition 2.1.13. We will assume by induction that the claim holds for all monic polynomials of degree $n-1$. Consider the characteristic matrix $xI - C_{p(x)}$.

$$xI - C_{p(x)} = \begin{bmatrix} x & 0 & \cdots & 0 & \alpha_0 \\ -1 & x & \cdots & 0 & \alpha_1 \\ 0 & -1 & \ddots & 0 & \alpha_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & x + \alpha_{n-1} \end{bmatrix}$$

By definition, the characteristic polynomial of $C_{p(x)}$ is $\det(xI - C_{p(x)})$. Expanding the determinant along the first row yields $\det(xI - C_{p(x)}) = x \det(xI - C_{q(x)}) + (-1)^{n+1}\alpha_0 \det(A)$ for the matrices

$$xI - C_{q(x)} = \begin{bmatrix} x & 0 & \cdots & 0 & \alpha_1 \\ -1 & x & \cdots & 0 & \alpha_2 \\ 0 & -1 & \ddots & 0 & \alpha_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & x + \alpha_{n-1} \end{bmatrix} \text{ and } A = \begin{bmatrix} -1 & x & 0 & \cdots & 0 \\ 0 & -1 & x & \cdots & 0 \\ 0 & 0 & -1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

obtained as $(n-1) \times (n-1)$ submatrices of $xI - C_{p(x)}$ by deleting the first row and first column and the first row and nth column of $xI - C_{p(x)}$, respectively. Observe that $C_{q(x)}$ is the companion matrix of the monic polynomial $q(x) = x^{n-1} + \alpha_{n-1}x^{n-2} + \cdots + \alpha_3 x^2 + \alpha_2 x + \alpha_1$, hence by induction, the characteristic polynomial and the minimal polynomial of $C_{q(x)}$ are both $q(x)$. Particularly, it follows that $x\det(xI - C_{q(x)}) = xq(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_3 x^3 + \alpha_2 x^2 + \alpha_1 x = p(x) - \alpha_0$. On the other hand, we note that A is an upper-triangular matrix with $n-1$ copies of -1 along the diagonal, hence we conclude by Proposition 2.3.11 that $\det(A) = (-1)^{n-1}$ and $(-1)^{n+1}\alpha_0 \det(A) = \alpha_0$. Combined, these two calculations reveal that $\det(xI - C_{p(x)}) = p(x) - \alpha_0 + \alpha_0 = p(x)$, as desired.

Even though it is a bit contrived, we will prove that $p(x)$ is the minimal polynomial of $C_{p(x)}$ by demonstrating that no monic polynomial of strictly lesser degree annihilates $C_{p(x)}$. Observe that for the $n \times 1$ standard basis vector $\mathbf{e}_1$ consisting of one in the first row and zeros elsewhere, we have that $C_{p(x)}\mathbf{e}_1 = \mathbf{e}_2$ so that $C_{p(x)}^2\mathbf{e}_1 = C_{p(x)}\mathbf{e}_2 = \mathbf{e}_3$ and $C_{p(x)}^k\mathbf{e}_1 = \mathbf{e}_{k+1}$ for all integers $1 \leq k \leq n-1$. Consequently, for any monic polynomial $q(x) = x^{n-1} + \beta_{n-2}x^{n-2} + \cdots + \beta_2 x^2 + \beta_1 x + \beta_0$, we have that $q(C_{p(x)})\mathbf{e}_1 = \mathbf{e}_n + \beta_{n-2}\mathbf{e}_{n-1} + \cdots + \beta_2\mathbf{e}_3 + \beta_1\mathbf{e}_2 + \beta_0\mathbf{e}_1$. We conclude that $q(C_{p(x)})$ is nonzero, hence there cannot be a monic polynomial of degree less than n that annihilates $C_{p(x)}$. □

Given any (real) matrices $A_1, \dots, A_k$ such that A_i is an $n_i \times n_i$ matrix for each integer $1 \leq i \leq k$, the **direct sum** of $A_1, \dots, A_k$ is the (real) $(n_1 + \cdots + n_k) \times (n_1 + \cdots + n_k)$ matrix

$$A_1 \oplus \cdots \oplus A_k = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix}$$

constructed by arranging the matrices $A_1, \dots, A_k$ along the main diagonal and completing the matrix with zeros elsewhere. We refer to a square matrix of this form as a **block diagonal** matrix.

Example 2.5.4. Every diagonal matrix can be realized as a block diagonal matrix whose components along the main diagonal are simply 1×1 matrices. Explicitly, we have the following.

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} a_{11} \end{bmatrix} \oplus \begin{bmatrix} a_{22} \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} a_{nn} \end{bmatrix}$$

Example 2.5.5. By definition, the direct sum of a 1×1 and a 2×2 matrix matrix is a 3×3 block diagonal matrix. Explicitly, the direct sum is a matrix of the following form.

$$\begin{bmatrix} a_{11} \end{bmatrix} \oplus \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & b_{11} & b_{12} \\ 0 & b_{21} & b_{22} \end{bmatrix}$$

Block diagonal matrices behave in a civilized manner with respect to taking determinants and computing their characteristic matrices. Consequently, the determinant, characteristic polynomial, and minimal polynomial of a block diagonal matrix can be easily deduced as follows.

Proposition 2.5.6. *Given any square matrices $A_1, \dots, A_k$, we have that*

$$\det(A_1 \oplus \cdots \oplus A_k) = \det(A_1) \cdots \det(A_k).$$

Proof. By definition of the direct sum of matrices, we have the following.

$$A_1 \oplus \cdots \oplus A_k = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix}$$

By Theorem 1.10.13, there exist scalars $\alpha_1, \dots, \alpha_k$ such that $\det(A_i) = \alpha_i \det(\text{RREF}(A_i))$ for each integer $1 \leq i \leq k$. Considering that the matrix $A_1 \oplus \cdots \oplus A_k$ is block diagonal, performing elementary row operations on any submatrix A_i does not affect any of the other submatrices, hence we may reduce each of the matrices $A_1, \dots, A_k$ to its reduced row echelon form at the cost of some scalar.

$$\det(A_1 \oplus \cdots \oplus A_k) = \alpha_1 \cdots \alpha_k \begin{vmatrix} \text{RREF}(A_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \text{RREF}(A_k) \end{vmatrix}$$

Either the reduced row echelon form of each of the matrices $A_1, \dots, A_k$ is the appropriately-sized identity matrix, or the reduced row echelon form of some matrix possesses a zero row. Certainly, in the first case, the determinant of the matrix in the above displayed equation is one, and we conclude that $\det(A_1 \oplus \cdots \oplus A_k) = \alpha_1 \cdots \alpha_k$. Even more, the determinant of each matrix A_i satisfies that $\det(A_i) = \alpha_i$, hence it holds that $\det(A_1 \oplus \cdots \oplus A_k) = \det(A_1) \cdots \det(A_k)$. Conversely, if the reduced row echelon form of some matrix possesses a zero row, then the determinant of the matrix in the above displayed equation is zero so that $\det(A_1 \oplus \cdots \oplus A_k) = 0 = \det(A_1) \cdots \det(A_k)$. $\square$

Corollary 2.5.7. *Given any square matrices $A_1, \dots, A_k$ with respective characteristic polynomials $\chi_1(x), \dots, \chi_k(x)$, the characteristic polynomial of $A_1 \oplus \cdots \oplus A_k$ is $\chi_1(x) \cdots \chi_k(x)$.*

Proof. Considering that $xI - (A_1 \oplus \cdots \oplus A_k) = (xI - A_1) \oplus \cdots \oplus (xI - A_k)$, the claim follows immediately from the definition of the characteristic polynomial and Proposition 2.5.6. $\square$

Proposition 2.5.8. *Given any square matrices $A_1, \dots, A_k$ with respective minimal polynomials $\mu_1(x), \dots, \mu_k(x)$, the minimal polynomial of $A_1 \oplus \cdots \oplus A_k$ is $\operatorname{lcm}(\mu_1(x), \dots, \mu_k(x))$.*

Proof. We claim that for any polynomial $p(x)$, we have that $p(A_1 \oplus \cdots \oplus A_k) = p(A_1) \oplus \cdots \oplus p(A_k)$. Considering that the identity $\alpha(A_1 \oplus \cdots \oplus A_k) = (\alpha A_1) \oplus \cdots \oplus (\alpha A_k)$ clearly holds, it suffices to prove that $(A_1 \oplus \cdots \oplus A_k)^n = (A_1^n) \oplus \cdots \oplus (A_k^n)$ for any positive integer n: indeed, we have that

$$(A_1 \oplus \cdots \oplus A_k)^2 = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix} \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k \end{bmatrix} = \begin{bmatrix} A_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_k^2 \end{bmatrix} = (A_1^2) \oplus \cdots \oplus (A_k^2)$$

because the only nonzero entries of this matrix product come from the rows and columns corresponding to the matrix A_i for each integer $1 \leq i \leq k$. Certainly, it is possible to repeat this process for any positive integer n, hence the desired identity $p(A_1 \oplus \cdots \oplus A_k) = p(A_1) \oplus \cdots \oplus p(A_k)$ holds.

Consider the least common multiple $p(x) = \operatorname{lcm}(\mu_1(x), \dots, \mu_k(x))$ of the minimal polynomials of $A_1, \dots, A_k$. By definition, for each integer $1 \leq i \leq k$, there exists a polynomial $q_i(x)$ such that $p(x) = \mu_i(x) q_i(x)$, from which it follows that $p(x)$ annihilates the matrices $A_1, \dots, A_k$. Consequently, we find that $p(x)$ annihilates $A_1 \oplus \cdots \oplus A_k$, hence by Proposition 2.1.9, we conclude that $p(x)$ must be divisible by the minimal polynomial $\mu(x)$ of $A_1 \oplus \cdots \oplus A_k$. Conversely, if $\mu(x)$ annihilates the direct sum $A_1 \oplus \cdots \oplus A_k$, then it must annihilate each of the matrices $A_1, \dots, A_k$ because it holds by the previous paragraph that $\mu(A_1 \oplus \cdots \oplus A_k) = \mu(A_1) \oplus \cdots \oplus \mu(A_k)$, and the latter is equal to the zero matrix if and only if $\mu(A_i)$ is equal to the zero matrix for each integer $1 \leq i \leq k$. By Proposition 2.1.9, $\mu(x)$ is divisible by $\mu_1(x), \dots, \mu_k(x)$, hence it is divisible by $p(x) = \operatorname{lcm}(\mu_1(x), \dots, \mu_k(x))$. $\square$

We are at last ready to construct the Rational Canonical Form of a real $n \times n$ matrix.

Definition 2.5.9 (Rational Canonical Form)**.** Consider any (real) $n \times n$ matrix A with invariant factors $p_1(x), p_2(x), \dots, p_\ell(x)$ whose companion matrices are $C_{p_1(x)}, C_{p_2(x)}, \dots, C_{p_\ell(x)}$, respectively. We define the **Rational Canonical Form** of A as the (real) $n \times n$ matrix

$$\operatorname{RCF}(A) = C_{p_1(x)} \oplus C_{p_2(x)} \oplus \cdots \oplus C_{p_\ell(x)} = \begin{bmatrix} C_{p_1(x)} & 0 & 0 & 0 \\ 0 & C_{p_2(x)} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & C_{p_\ell(x)} \end{bmatrix}$$

Example 2.5.10. Let us compute the Rational Canonical Form for the matrix of Example 2.4.3.

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

We proved in that example that the only invariant factor of A is $(x-1)(x+1) = x^2-1$. Consequently, the Rational Canonical Form for A is the companion matrix of this quadratic polynomial.

$$\operatorname{RCF}(A) = C_{x^2-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Example 2.5.11. Let us compute the Rational Canonical Form for the matrix of Example 2.4.4.

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

We proved in that example that the only invariant factor of A is x^2. Like the previous example, the Rational Canonical Form for A must be the companion matrix of x^2.

$$\operatorname{RCF}(A) = C_{x^2} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Example 2.5.12. Let us compute the Rational Canonical Form for the matrix of Example 2.4.5.

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

We proved in that example that the invariant factors of A are x and $x(x-6) = x^2-6x$. Consequently, the Rational Canonical Form for A is the direct sum of the companion matrices of x and $x^2 - 6x$.

$$\operatorname{RCF}(A) = C_x \oplus C_{x^2-6x} = \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 0 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 6 \end{bmatrix}$$

Example 2.5.13. Let us compute the Rational Canonical Form for the matrix of Example 2.4.6.

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

Considering that the invariant factors of A are $x-1$ and $(x-1)^2 = x^2-2x+1$ by the example, the Rational Canonical Form for A is the direct sum of the companion matrices of $x-1$ and x^2-2x+1.

$$\mathrm{RCF}(A) = C_{x-1} \oplus C_{x^2-2x+1} = \begin{bmatrix}1\end{bmatrix} \oplus \begin{bmatrix}0 & -1\\ 1 & 2\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & -1\\ 0 & 1 & 2\end{bmatrix}$$

Example 2.5.14. Consider any matrix A whose invariant factors are $x-1$ and $(x-1)(x-2)$. Observe that any such matrix must be a 3×3 matrix. By definition, the Rational Canonical Form for such a matrix is the direct sum of the companion matrices of $x-1$ and $(x-1)(x-2) = x^2-3x+2$.

$$\mathrm{RCF}(A) = C_{x-1} \oplus C_{x^2-3x+2} = \begin{bmatrix}1\end{bmatrix} \oplus \begin{bmatrix}0 & -2\\ 1 & 3\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & -2\\ 0 & 1 & 3\end{bmatrix}$$

Example 2.5.15. Consider any matrix A whose invariant factors are x, x^2, and $x^3(x+1)^2$. Observe that any such matrix must be an 8×8 matrix. By definition, the Rational Canonical Form for such a matrix is the direct sum of the companion matrices of x, x^2, and $x^3(x+1)^2 = x^5+2x^4+x^3$.

$$\mathrm{RCF}(A) = C_x \oplus C_{x^2} \oplus C_{x^5+2x^4+x^3} = \begin{bmatrix}0\end{bmatrix} \oplus \begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix} \oplus \begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & -1\\ 0 & 0 & 0 & 1 & -2\end{bmatrix}$$

Example 2.5.16. Consider any matrix A whose invariant factors are x, x, $x^2(x^2+1) = x^4+x^2$, and $x^3(x-1)(x^2+1) = x^3(x^3-x^2+x-1) = x^6-x^5+x^4-x^3$. Observe that any such matrix must be a 12×12 matrix. By definition, the Rational Canonical Form for such a matrix is the direct sum of the companion matrices of x, x, x^4+x^2, and $x^6-x^5+x^4-x^3$.

$$\begin{aligned}\mathrm{RCF}(A) &= C_x \oplus C_x \oplus C_{x^4+x^2} \oplus C_{x^6-x^5+x^4-x^3}\\ &= \begin{bmatrix}0\end{bmatrix} \oplus \begin{bmatrix}0\end{bmatrix} \oplus \begin{bmatrix}0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & -1\\ 0 & 0 & 1 & 0\end{bmatrix} \oplus \begin{bmatrix}0\end{bmatrix} \oplus \begin{bmatrix}0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & -1\\ 0 & 0 & 1 & 0\end{bmatrix} \oplus \begin{bmatrix}0 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 1\\ 0 & 0 & 0 & 1 & 0 & -1\\ 0 & 0 & 0 & 0 & 1 & 1\end{bmatrix}\end{aligned}$$

Example 2.5.17. Consider any matrix A with two invariant factors of $x^2(x^2+x+1) = x^4+x^3+x^2$. Observe that any such matrix must be an 8×8 matrix, and the Rational Canonical Form for such a matrix must be the direct sum of the companion matrix of $x^4+x^3+x^2$ with itself.

$$\mathrm{RCF}(A) = C_{x^4+x^3+x^2} \oplus C_{x^4+x^3+x^2} = \begin{bmatrix}0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & -1\\ 0 & 0 & 1 & -1\end{bmatrix} \oplus \begin{bmatrix}0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & -1\\ 0 & 0 & 1 & -1\end{bmatrix}$$

2.6 Jordan Canonical Form

Like the Rational Canonical Form, the **Jordan Canonical Form** of an $n \times n$ matrix is a block diagonal matrix built as a direct sum of square matrices that are obtained from the Smith Normal Form of the characteristic matrix. Explicitly, suppose that A is a (real) $n \times n$ matrix with elementary divisors $(x - c_{i1})^{e_{i1}}, \dots, (x - c_{ik})^{e_{ik}}$. We refer to the following $e_{ij} \times e_{ij}$ upper-triangular matrix $J_{(x-c_{ij})^{e_{ij}}}$ as the **Jordan matrix** (or **Jordan block**) corresponding to the elementary divisor $(x - c_{ij})^{e_{ij}}$.

$$J_{(x-c_{ij})^{e_{ij}}} = \begin{bmatrix} c_{ij} & 1 & 0 & \cdots & 0 \\ 0 & c_{ij} & 1 & \cdots & 0 \\ 0 & 0 & c_{ij} & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \cdots & c_{ij} \end{bmatrix}$$

Put another way, the Jordan matrix corresponding to the elementary divisor $(x - c_{ij})^{e_{ij}}$ is the $e_{ij} \times e_{ij}$ upper-triangular matrix consisting of c_{ij} on the diagonal and ones along the **superdiagonal**.

Example 2.6.1. By definition, the Jordan matrix corresponding to any linear polynomial $x + c$ is the 1×1 matrix $J_{x+c} = \begin{bmatrix} -c \end{bmatrix}$. One might recognize this as the companion matrix of $x + c$.

Example 2.6.2. By definition, the Jordan matrix corresponding to the polynomial $(x - 1)^2$ is the 2×2 upper-triangular matrix with ones along the diagonal and ones along the superdiagonal.

$$J_{(x-1)^2} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Example 2.6.3. By definition, the Jordan matrix corresponding to the polynomial $(x + 3)^3$ is the 3×3 upper-triangular matrix with -3s along the diagonal and ones along the superdiagonal.

$$J_{(x+3)^3} = \begin{bmatrix} -3 & 1 & 0 \\ 0 & -3 & 1 \\ 0 & 0 & -3 \end{bmatrix}$$

Definition 2.6.4 (Jordan Canonical Form)**.** Consider any (real) $n \times n$ matrix A with elementary divisors $(x - c_{i1})^{e_{i1}}, (x - c_{i2})^{e_{i2}}, \dots, (x - c_{ik})^{e_{ik}}$ and their corresponding Jordan matrices $J_{(x-c_{i1})^{e_{i1}}}$, $J_{(x-c_{i2})^{e_{i2}}}, \dots, J_{(x-c_{ik})^{e_{ik}}}$. We define the **Jordan Canonical Form** of A as the $n \times n$ matrix

$$\mathrm{JCF}(A) = J_{(x-c_{i1})^{e_{i1}}} \oplus J_{(x-c_{i2})^{e_{i2}}} \oplus \cdots \oplus J_{(x-c_{ik})^{e_{ik}}} = \begin{bmatrix} J_{(x-c_{i1})^{e_{i1}}} & 0 & 0 & 0 \\ 0 & J_{(x-c_{i2})^{e_{i2}}} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & J_{(x-c_{ik})^{e_{ik}}} \end{bmatrix}$$

Example 2.6.5. Let us compute the Jordan Canonical Form for the matrix of Example 2.4.3.

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

We proved in that example that the elementary divisors of A are $x - 1$ and $x + 1$. Consequently, the Jordan Canonical Form for A is the direct sum of the 1×1 Jordan matrices J_{x-1} and J_{x+1}.

$$\mathrm{JCF}(A) = J_{x-1} \oplus J_{x+1} = \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Example 2.6.6. Let us compute the Jordan Canonical Form for the matrix of Example 2.4.4.

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

We proved in that example that the only elementary divisor of A is x^2. Like the previous example, the Jordan Canonical Form for A must be the 2×2 Jordan matrix J_{x^2}.

$$\mathrm{JCF}(A) = J_{x^2} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Example 2.6.7. Let us compute the Jordan Canonical Form for the matrix of Example 2.4.5.

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

We proved in that example that the elementary divisors of A are x, x, and $x-6$. Consequently, the Jordan Canonical Form for A is the direct sum of the 1×1 Jordan matrices J_x, J_x, and J_{x-6}.

$$\mathrm{JCF}(A) = J_x \oplus J_x \oplus J_{x-6} = \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

Example 2.6.8. Let us compute the Jordan Canonical Form for the matrix of Example 2.4.6.

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

By the example, the elementary divisors of A are $x-1$ and $(x-1)^2$, hence the Jordan Canonical Form for A is the direct sum of the 1×1 Jordan matrix J_{x-1} and the 2×2 Jordan matrix $J_{(x-1)^2}$.

$$\mathrm{JCF}(A) = J_{x-1} \oplus J_{(x-1)^2} = \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Example 2.6.9. Consider any matrix A whose elementary divisors are $x-1$, $x-1$, and $x-2$. Observe that any such matrix must be a 3×3 matrix. By definition, the Jordan Canonical Form for such a matrix is the direct sum of the Jordan matrices J_{x-1}, J_{x-1}, and J_{x-2}.

$$\mathrm{JCF}(A) = J_{x-1} \oplus J_{x-1} \oplus J_{x-2} = \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Example 2.6.10. Consider any matrix A whose elementary divisors are x, x^2, x^3, and $(x+1)^2$. Observe that any such matrix must be an 8×8 matrix. By definition, the Jordan Canonical Form for such a matrix is the direct sum of the Jordan matrices corresponding to x, x^2, x^3, and $(x+1)^2$.

$$\mathrm{JCF}(A) = J_x \oplus J_{x^2} \oplus J_{x^3} \oplus J_{(x+1)^2} = \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} -1 & 1 \\ 0 & -1 \end{bmatrix}$$

Example 2.6.11. Consider any real matrix A whose invariant factors are x, x, $x^2(x^2+1)$, and $x^3(x-1)(x^2+1)$. Observe that both roots of the polynomial x^2+1 are complex numbers: indeed, the roots of x^2+1 are i and $-i$. Consequently, if we view A a real matrix, then A does not admit a Jordan Canonical Form. Explicitly, the Jordan Canonical Form is built from the Jordan matrices corresponding to powers of linear polynomials: if x^2+1 is an elementary divisor of A, then viewed as a real polynomial, this polynomial does not **split** as a product of linear polynomials; however, if we view A as a matrix whose entries are complex numbers, then we may view x^2+1 as a polynomial with complex coefficients, hence it is permissible to factor x^2+1 as $(x+i)(x-i)$. Under this lens, the elementary divisors of A are $x, x, x^2, x^3, x-1, x-i, x+i, x-i$, and $x+i$. Consequently, the Jordan Canonical Form for A is the following 12×12 complex upper-triangular matrix.

$$\begin{aligned} \operatorname{JCF}(A) &= J_x \oplus J_x \oplus J_{x^2} \oplus J_{x^3} \oplus J_{x-1} \oplus J_{x-i} \oplus J_{x+i} \oplus J_{x-i} \oplus J_{x+i} \\ &= \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 1 \end{bmatrix} \oplus \begin{bmatrix} i \end{bmatrix} \oplus \begin{bmatrix} -i \end{bmatrix} \oplus \begin{bmatrix} i \end{bmatrix} \oplus \begin{bmatrix} -i \end{bmatrix} \end{aligned}$$

Example 2.6.12. Consider any matrix A with elementary divisors of x^2, x^2, x^2+x+1, and x^2+x+1. Observe that the Jordan Canonical Form for such a matrix exists if and only if we view A as a matrix with complex entries: indeed, the polynomial x^2+x+1 has two complex roots $-\frac{1}{2}+\frac{\sqrt{3}}{2}i$ and $-\frac{1}{2}-\frac{\sqrt{3}}{2}i$. Consequently, the Jordan Canonical Form for A is the following.

$$\operatorname{JCF}(A) = J_{x^2} \oplus J_{x^2} \oplus J_{x+\frac{1}{2}-\frac{\sqrt{3}}{2}i} \oplus J_{x+\frac{1}{2}+\frac{\sqrt{3}}{2}i} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \begin{bmatrix} -\frac{1}{2}+\frac{\sqrt{3}}{2}i \end{bmatrix} \oplus \begin{bmatrix} -\frac{1}{2}-\frac{\sqrt{3}}{2}i \end{bmatrix}$$

Remark 2.6.13. Examples 2.6.11 and 2.6.12 raise an important point regarding the Jordan Canonical Form of a square matrix A: it exists if and only if the elementary divisors of A are all power of linear polynomials. Consequently, if we want the Jordan Canonical Form to exist for any square matrix, we must assume that the entries of our matrix lie in an **algebraically closed** field, i.e., we must ensure that the characteristic polynomial of our matrix can be written as a product of (not necessarily distinct) linear polynomials. Often, the caveat with the Jordan Canonical Form is that it is an upper-triangular matrix with entries in the complex numbers. Conversely, the Rational Canonical Form of a matrix always exists; however, it is rarely an upper-triangular matrix. Even still, in most cases, the Jordan Canonical Form is preferable to the Rational Canonical Form because of its upper-triangular form. One can prove that the determinant of a matrix is the product of its eigenvalues, hence the product of the eigenvalues of a real matrix must be a real number. We could have predicted this based on the fact that complex roots come in conjugate pairs whose product is a real number. Even more, the **trace** of a matrix is the sum of the diagonal components of the matrix; this can be achieved as the sum of the eigenvalues. Once again, if the matrix is real, then the sum of its eigenvalues is a real number because each conjugate pair of complex eigenvalues sum to a real number. Consequently, the requirement to pass to the complex numbers is not detrimental.

2.7 Chapter Overview

This section is currently under construction.

Chapter 3

Linear Transformations of Vector Spaces

By now, we are sufficiently familiar with the theory of $n \times n$ matrices whose entries lie in a field. Culminating in the construction of canonical forms, our studies have led us to develop sophisticated machinery to understand both the algebraic and geometric properties of real $n \times n$ matrices. Our principal aim throughout this chapter is to recognize that real n-space and real $m \times n$ matrices belong to a more general notion of vector spaces and linear transformations of vector spaces that are ubiquitous throughout mathematics and physics. Explicitly, we will demonstrate that every linear transformation of real n-space is uniquely determined by a real $m \times n$ matrix and vice-versa. Even more, we will explore other Euclidean vector spaces, e.g., real polynomials and real functions.

3.1 Linear Transformations of Euclidean Spaces

We begin our discussion by introducing a class of "ideal" functions from real n-space to real m-space for positive integers m and n. We recall that a **function** $f : X \to Y$ from a nonempty set X to a nonempty set Y is a relation such that for each element x in X, there is one and only one element $y = f(x)$ in Y. Each function from a nonempty set X to a nonempty set Y induces the following.

1.) We recall that the **domain** D_f of the function $f : X \to Y$ consists of all x-values for which $y = f(x)$ is well-defined. Consequently, we have that $D_f = \{x \in X \mid y = f(x) \text{ is well-defined}\}$. Often, we will assume that the domain of a function $f : X \to Y$ is simply X, but in some cases (such as those arising in calculus), it is required to determine the domain of f explicitly.

2.) We refer to the set Y as the **codomain** of the function $f : X \to Y$. We say that $y = f(x)$ is the **image** of the element x in X. Crucially, it is not necessarily true that every element of Y must be the image of some element in X. Explicitly, the collection $f(A)$ of all elements y in Y for which $y = f(x)$ for some element x in A is called the **image** of A under f. Consequently, we have that $f(A) = \{y \in Y \mid y = f(x) \text{ for some element } x \text{ in } A\}$. We are familiar with the image of the entire domain X under f: it is called the **range** R_f of f, i.e., $R_f = f(X)$.

3.) Given any nonempty subset B of Y, the collection $f^{-1}(B)$ of all elements x in X for which $y = f(x)$ is an element of B is called the **inverse image** of B under f. Consequently, we have that $f^{-1}(B) = \{y \in B \mid y = f(x) \text{ for some element } x \text{ in } X\}$; this could be empty!

We turn our attention next to the structure-preserving functions between real n-space.

Definition 3.1.1. Given any positive integers m and n, we say that a function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a **linear transformation** if and only if it preserves addition and scalar multiplication, i.e.,

1.) $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$ for all vectors $\mathbf{v}$ and $\mathbf{w}$ of real n-space and

2.) $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$ for all vectors $\mathbf{v}$ in real n-space and all scalars α.

Conveniently, it is possible to summarize the above pair of **linearity** conditions as follows.

Proposition 3.1.2. *Given any positive integers m and n, the function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if $T(\alpha\mathbf{v} + \mathbf{w}) = \alpha T(\mathbf{v}) + T(\mathbf{w})$ for all vectors $\mathbf{v}$ and $\mathbf{w}$ and scalars α.*

Proof. Certainly, if the function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then by Definition 3.1.1, it holds that $T(\alpha\mathbf{v} + \mathbf{w}) = T(\alpha\mathbf{v}) + T(\mathbf{w}) = \alpha T(\mathbf{v}) + T(\mathbf{w})$ for all vectors $\mathbf{v}$ and $\mathbf{w}$ in real n-space and all scalars α. Conversely, if we assume that $T(\alpha\mathbf{v} + \mathbf{w}) = \alpha T(\mathbf{v}) + T(\mathbf{w})$ for all vectors $\mathbf{v}$ and $\mathbf{w}$ of real n-space and all scalars α, then in particular, we may evaluate $T(\mathbf{0})$ to find that

$$T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0}).$$

Cancelling $T(\mathbf{0})$ from both sides, we find that $T(\mathbf{0}) = \mathbf{0}$. Consequently, it follows that

1.) $T(\mathbf{v} + \mathbf{w}) = T(1\mathbf{v} + \mathbf{w}) = 1T(\mathbf{v}) + T(\mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$ and

2.) $T(\alpha\mathbf{v}) = T(\alpha\mathbf{v} + \mathbf{0}) = \alpha T(\mathbf{v}) + T(\mathbf{0}) = \alpha T(\mathbf{v}) + \mathbf{0} = \alpha T(\mathbf{v})$

for all vectors $\mathbf{v}$ and $\mathbf{w}$ in real n-space and scalars α; as such, the asserted claim holds. □

Even more, we collect in the next proposition two useful properties of linear transformations.

Proposition 3.1.3 (Basic Properties of Linear Transformations of Euclidean Spaces)**.** *Consider any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ defined for any positive integers m and n.*

1.) *We have that $T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_n T(\mathbf{v}_n)$ for all vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in real n-space and all scalars $\alpha_1, \ldots, \alpha_n$. 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 $T(\mathbf{0}) = \mathbf{0}$ for the respective zero vectors $\mathbf{0}$ of real n-space and real m-space.*

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 $n = 1$, so we may assume inductively that $T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_n T(\mathbf{v}_n)$ for all vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ and all scalars $\alpha_1, \ldots, \alpha_n$. By definition of a linear transformation, we have that

$$T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n + \alpha_{n+1}\mathbf{v}_{n+1}) = T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n) + T(\alpha_{n+1}\mathbf{v}_{n+1}).$$

By hypothesis, the first summand is equal to $\alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_n T(\mathbf{v}_n)$, from which it follows that $T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n + \alpha_{n+1}\mathbf{v}_{n+1}) = \alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_n T(\mathbf{v}_n) + \alpha_{n+1} T(\mathbf{v}_{n+1})$, as desired.

On the matter of the second property, we use the linearity of the function T to first recognize that $T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$. On the other hand, the zero vector satisfies that $\mathbf{0} + \mathbf{0} = \mathbf{0}$, hence we have that $T(\mathbf{0}) + T(\mathbf{0}) = T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0})$. Cancelling $T(\mathbf{0})$ from both sides yields $T(\mathbf{0}) = \mathbf{0}$. □

Example 3.1.4. We claim that the function $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x, y) = (-y, x)$ is a linear transformation called the **orthogonalization** of (x, y) in the xy-plane.

1.) Given any vectors $\mathbf{v} = [x_1, y_1]$ and $\mathbf{w} = [x_2, y_2]$, we have that $\mathbf{v} + \mathbf{w} = [x_1 + x_2, y_1 + y_2]$ and
$$T(\mathbf{v} + \mathbf{w}) = T(x_1 + x_2, y_1 + y_2) = (-y_1 - y_2, x_1 + x_2) = (-y_1, x_1) + (-y_2, x_2) = T(\mathbf{v}) + T(\mathbf{w}).$$

2.) Given any vector $\mathbf{v} = [x, y]$ and any scalar α, we have that $\alpha\mathbf{v} = [\alpha x, \alpha y]$ and
$$T(\alpha\mathbf{v}) = T(\alpha x, \alpha y) = (-\alpha y, \alpha x) = \alpha(-y, x) = \alpha T(\mathbf{v}).$$

By definition, we conclude that $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x, y) = (-y, x)$ is a linear transformation.

Example 3.1.5. We claim that the function $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (x, y)$ is a linear transformation called the **projection** of (x, y, z) into the xy-plane.

1.) Given any vectors $\mathbf{v} = [x_1, y_1, z_1]$ and $\mathbf{w} = [x_2, y_2, z_2]$, we have that
$$T(\mathbf{v}+\mathbf{w}) = T(x_1+x_2, y_1+y_2, z_1+z_2) = (x_1+x_2, y_1+y_2) = (x_1, y_1)+(x_2, y_2) = T(\mathbf{v})+T(\mathbf{w}).$$

2.) Given any vector $\mathbf{v} = [x, y, z]$ and any scalar α, we have that $\alpha\mathbf{v} = [\alpha x, \alpha y, \alpha z]$ and
$$T(\alpha\mathbf{v}) = T(\alpha x, \alpha y, \alpha z) = (\alpha x, \alpha y) = \alpha(x, y) = \alpha T(\mathbf{v}).$$

By definition, we conclude that $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x, y, z) = (x, y)$ is a linear transformation.

Example 3.1.6. We claim that the function $T : \mathbb{R} \to \mathbb{R}^3$ defined by $T(x) = (x, x^2, x^3)$ is not a linear transformation: indeed, it is possible to deduce this based purely on the fact that x^2 and x^3 are not linear functions; however, a concrete example to illustrate that T is not linear is that
$$T(2) = (2, 2^2, 2^3) = (2, 4, 8) \neq (2, 2, 2) = 2(1, 1, 1) = 2T(1).$$

Example 3.1.7. We will explicitly determine in this example all linear transformations $T : \mathbb{R} \to \mathbb{R}$ from the real line to itself; then, we will give a geometric interpretation of the image of such functions in the Cartesian plane. By Definition 3.1.1, for every real number x, we must have that
$$T(x) = xT(1) = T(1)x.$$
Consequently, the image $T(x)$ of x under T for any real number x is uniquely determined by the image $T(1)$ of 1 under T. Even more, if we denote $T(1) = m$, then $T(x) = mx$ for all real numbers x, hence the linear transformation $T : \mathbb{R} \to \mathbb{R}$ are precisely the lines through the origin.

Often, it is more tedious than it is difficult to determine according to Definition 3.1.1 whether a function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation; however, our next theorem provides a classification of all linear transformations of real n-space that eliminates our need to refer to the definition.

Theorem 3.1.8 (Linear Transformations of Euclidean Spaces and Real Matrices)**.** *Given any positive integers m and n, a function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if there exists a real $m \times n$ matrix A such that $T(\mathbf{v}) = A\mathbf{v}$ for every vector $\mathbf{v}$ in real n-space.*

Proof. We prove first that if A is any real $m \times n$ matrix, then the function $T_A(\mathbf{v}) = A\mathbf{v}$ is a linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$. Considering any vector $\mathbf{v}$ in real n-space as a real $n \times 1$ matrix, it follows that $A\mathbf{v}$ is a real $m \times 1$ matrix, hence $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is well-defined by its domain and codomain. Even more, for any vectors $\mathbf{v}$ and $\mathbf{w}$ in real n-space and any scalar α, we have that

$$T_A(\alpha\mathbf{v} + \mathbf{w}) = A(\alpha\mathbf{v} + \mathbf{w}) = A(\alpha\mathbf{v}) + A\mathbf{w} = \alpha A\mathbf{v} + A\mathbf{w} = \alpha T_A(\mathbf{v}) + T_A(\mathbf{w})$$

because Matrix Multiplication Is Distributive. By Proposition 3.1.2, T_A is a linear transformation.

Conversely, we demonstrate that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then there exists a real $m \times n$ matrix A such that $T(\mathbf{v}) = A\mathbf{v}$ for every vector $\mathbf{v}$ in real n-space. Considering that T is a linear transformation, every coordinate of $T(\mathbf{v})$ must be a linear combination of the coordinates of $\mathbf{v}$. Consequently, if we assume that the coordinates of $\mathbf{v}$ are given by $\mathbf{v} = [v_1, v_2, \dots, v_n]$, then there must exist real numbers $a_{i1}, a_{i2}, \dots, a_{in}$ for each integer $1 \leq i \leq m$ such that

$$T(\mathbf{v}) = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + \cdots + a_{1n}v_n \\ a_{21}v_1 + a_{22}v_2 + \cdots + a_{2n}v_n \\ \vdots \\ a_{m1}v_1 + a_{m2}v_2 + \cdots + a_{mn}v_n \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} v_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} v_2 + \cdots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} v_n.$$

By Remark 1.3.20, we find that $T(\mathbf{v}) = A\mathbf{v}$ for the real $m \times n$ matrix A with (i, j)th entry a_{ij}. $\square$

We refer to the matrix of Theorem 3.1.8 as the **matrix representation** of the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ (with respect to the standard basis vectors of real n-space). Eventually, we will come to find that the matrix representation of a linear transformation is an indispensable tool in the theory of linear transformations of vector spaces; however, for now, it is enough to witness the utility of the matrix representation in the context of linear transformations of real n-space.

Example 3.1.9. We will determine in this example whether the function $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(x, y, z) = (y, x - y + z, x)$ is a linear transformation. Based on intuition, at a glance, we are inclined to believe that T is a linear transformation because each component of its image is a linear function of x, y, and z. Concretely, we may determine the matrix representation as follows.

$$T(x, y, z) = \begin{bmatrix} y \\ x - y + z \\ x \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} y + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} z = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Consequently we conclude that T is a linear transformation with the above matrix representation.

Example 3.1.10. Let us next illustrate how to determine a formula for the image a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by computing the matrix representation of T according to the images of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ of real n-space under T. Explicitly, let us assume that

$$T(1, 0) = (1, 3, 5) \text{ and } T(0, 1) = (2, 4, 6)$$

so that $T : \mathbb{R}^2 \to \mathbb{R}^3$. Every vector $\mathbf{v}$ in real 2-space can be written as a linear combination of $[1, 0]$ and $[0, 1]$ according to the fact that $\mathbf{v} = [x, y] = [x, 0] + [0, y] = x[1, 0] + y[0, 1]$, hence we find that

$$T(\mathbf{v}) = T(x[1, 0] + y[0, 1]) = xT(1, 0) + yT(0, 1) = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} x + \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} y = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Consequently, the matrix representation of T is provided by the above real 3×2 matrix, and the formula for the image of a vector $\mathbf{v} = [x, y]$ under T is given by $T(x, y) = (x + 2y, 3x + 4y, 5x + 6y)$.

Example 3.1.11. On the other hand, it is possible to determine a formula for the image a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by computing the matrix representation of T according to the images of any linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ of real n-space under T. Concretely, suppose that

$$T(1,1,0) = (1,0,1) \text{ and } T(1,0,1) = (0,2,2) \text{ and } T(0,1,1) = (-1,2,1)$$

so that $T : \mathbb{R}^3 \to \mathbb{R}^3$. Every vector $\mathbf{v} = [x, y, z]$ in real 3-space can be written as a linear combination of $[1, 1, 0]$, $[1, 0, 1]$, and $[0, 1, 1]$ because these vectors are linearly independent: indeed, to determine the coefficients a, b, and c of $[x, y, z]$ with respect to the basis vectors $[1, 1, 0]$, $[1, 0, 1]$, and $[0, 1, 1]$, it suffices to solve the following 3×3 matrix equation using Gaussian Elimination or matrix inversion.

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Using Gaussian Elimination, we reduce an augmented matrix to reduced row echelon form.

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 1 & 0 & 1 & y \\ 0 & 1 & 1 & z \end{array}\right] \overset{(1.)}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & -1 & 1 & y - x \\ 0 & 1 & 1 & z \end{array}\right] \overset{(2.)}{\sim} \left[\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & -1 & 1 & y - x \\ 0 & 0 & 2 & z + y - x \end{array}\right] \overset{(3.)}{\sim} \left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{1}{2}(x + y - z) \\ 0 & 1 & 0 & \frac{1}{2}(z - y + x) \\ 0 & 0 & 1 & \frac{1}{2}(z + y - x) \end{array}\right]$$

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

(2.) We employed the elementary row operation $R_3 + R_2 \mapsto R_3$.

(3.) We employed the elementary row operations $\frac{1}{2}R_3 \mapsto R_3$, $-R_2 + R_3 \mapsto R_2$, and $R_1 - R_2 \mapsto R_1$.

Considering the rows of the above augmented matrix as the formula for T, we conclude that

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

Equivalently, our above computation yields the matrix representation of T since we have that

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

Considering the intimate relationship between a linear transformation T of real n-space and its matrix representation by a real $m \times n$ matrix A, it is not surprising that the analogy between a linear transformation and its matrix representation gives rise to the following important notions.

Definition 3.1.12. Given any positive integer n, we say that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is **invertible** if and only if the standard matrix representation A of T is invertible. Explicitly, we define the **inverse transformation** of an invertible linear transformation T as the linear transformation $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ represented by the matrix inverse A^{-1} of the real $n \times n$ matrix A.

Example 3.1.13. Each of the linear transformations of Examples 3.1.4, 3.1.9, and 3.1.11 is invertible because their matrix representations are invertible; the linear transformations of Examples 3.1.5 and 3.1.10 are not invertible because these transformations are not represented by square matrices.

Definition 3.1.14. Given any positive integer m and n and any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$,

1.) the **range** of T consists of all vectors in $\mathbb{R}^m$ that are the image of some vector in $\mathbb{R}^n$, i.e.,

$$\operatorname{range}(T) = \{\mathbf{w} \in \mathbb{R}^m \mid \mathbf{w} = T(\mathbf{v}) \text{ for some vector } \mathbf{v} \in \mathbb{R}^n\} \text{ and}$$

2.) the **kernel** of T consists of all vectors in $\mathbb{R}^n$ whose image in $\mathbb{R}^m$ is the zero vector, i.e.,

$$\ker(T) = \{\mathbf{v} \in \mathbb{R}^n \mid T(\mathbf{v}) = \mathbf{0}\}.$$

Crucially, if $T : \mathbb{R}^n \to \mathbb{R}^m$ admits a real $m \times n$ matrix representation A, the range of T is the column space of A and the kernel of T is the null space of A, i.e., $\operatorname{range}(T) = \operatorname{col}(A)$ and $\ker(T) = \operatorname{null}(A)$. Consequently, the range and kernel of a linear transformation are subspaces of real n-space, and we may therefore define the **rank** and **nullity** of a linear transformation according to the formulae

$$\operatorname{rank}(T) = \operatorname{rank}(A) \text{ and } \operatorname{nullity}(T) = \operatorname{nullity}(A).$$

By the Rank Equation for A, we obtain the following identity for the linear transformation T.

$$\operatorname{rank}(T) + \operatorname{nullity}(T) = \#(\text{columns of the matrix representation } A)$$

Example 3.1.15. Each of the linear transformations of Examples 3.1.4, 3.1.9, and 3.1.11 has rank equal to the number of columns of its matrix representation by Corollary 1.8.13 because their matrix representations are invertible, hence the nullity of each transformation is zero. Likewise, the linear transformations of Examples 3.1.5 and 3.1.10 have rank equal to the number of columns of their matrix representations because their matrix representations have as many pivots as columns.

3.2 Vector Spaces

Going forward, we will formally 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 3.2.1. Consider the set S that consists of the first five positive integers 1, 2, 3, 4, and 5. We note that the elements of S are precisely the integers 1, 2, 3, 4, and 5, hence 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 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 only keeps track of what objects belong to it. Likewise, the number of times an element of S appears in the set 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 false 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 3.2.2. 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

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

Back to our example of the real numbers, the objects in $\mathbb{R}$ are denoted by x, and the requirement for set membership in $\mathbb{R}$ 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 place a vertical bar $\mid$ to signify the phrase "such that"; and finally, we list the condition under which an object belongs to the set S in question.

Example 3.2.3. Consider the collection $\mathbb{R}^{m\times n}$ of real $m\times n$ matrices. We note that 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 3.2.4. Consider the collection $\mathbb{R}[x]$ of real polynomials in indeterminate x. We note that 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 1.3.4, 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 3.2.5. We say that a pair $(V, +)$ is a (real) **vector space** if the following hold.

1.) (Closure Under Addition) We have that $\mathbf{u} + \mathbf{v} \in V$ for any pair of elements $\mathbf{u}, \mathbf{v} \in V$.

2.) (Associativity of Addition) We have that $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$.

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

4.) (Additive Identity) There exists an element $\mathbf{0}_V \in V$ such that $\mathbf{v} + \mathbf{0}_V = \mathbf{v}$ for all $\mathbf{v} \in V$.

5.) (Additive Inverse) For any element $\mathbf{v} \in V$, we have that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}_V$ for some $-\mathbf{v} \in V$.

6.) (Closure Under Scalar Multiplication) We have that $\alpha\mathbf{v} \in V$ for all (real) scalars α and $\mathbf{v} \in V$.

7.) (Multiplicative Identity) We have that $1\mathbf{v} = \mathbf{v}$ for each element $\mathbf{v} \in V$.

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

9.) (Distributive Law I) We have that $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ for all (real) scalars α and $\mathbf{u}, \mathbf{v} \in V$.

10.) (Distributive Law II) We have that $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{v} + \beta\mathbf{v}$ for all (real) scalars α, β, and $\mathbf{v} \in V$.

We will henceforth refer to the elements $\mathbf{v}$ of any vector space V as (real) **vectors**.

Combined, the first five properties of Definition 3.2.5 ensure that any vector space V constitutes an **abelian group** with respect to the particular addition defined on its elements. Group theory is an essential branch of study in modern algebra, but we will not concern ourselves with their study here; however, we will come to find that different notions of addition are required for different vector spaces, e.g., the familiar addition of vectors in real n-space, addition of (real) $m \times n$ matrices, and addition of real polynomials of degree $\leq n$ for some non-negative integer n. Our next proposition confirms the fact (we have taken for granted) that $\mathbb{R}^n$ forms a real n-dimensional vector space.

Proposition 3.2.6. *Real n-space $\mathbb{R}^n$ forms a real vector space of dimension n.*

Proof. We define addition of points in real n-space **componentwise** by

$$[x_1, x_2, \dots, x_n] + [y_1, y_2, \dots, y_n] = [x_1 + y_1, x_2 + y_2, \dots, x_n + y_n].$$

Considering that addition of real numbers constitutes an associative and commutative binary operation on the real numbers, the first three conditions of Definition 3.2.5 are satisfied. Even more, the zero vector in $\mathbb{R}^n$ is the n-tuple $\mathbf{0} = [0, 0, \dots, 0]$, and for any real n-tuple $[x_1, x_2, \dots, x_n]$, we have that $-[x_1, x_2, \dots, x_n] = [-x_1, -x_2, \dots, -x_n]$. We conclude that the fourth and fifth conditions of the definition hold, hence we may turn our attention to scalar multiplication in $\mathbb{R}^n$. We define $\alpha[x_1, x_2, \dots, x_n] = [\alpha x_1, \alpha x_2, \dots, \alpha x_n]$ for any real number α and any real n-tuple $[x_1, x_2, \dots, x_n]$. Considering that multiplication of real numbers constitutes an associative, commutative, and distributive binary operation on the real numbers, it follows that $\mathbb{R}^n$ is a real vector space. Last, the dimension of $\mathbb{R}^n$ is n since the standard basis of $\mathbb{R}^n$ consists of the vectors $\mathbf{e}_i$ whose ith coordinate is 1 and whose other coordinates are 0, i.e., $\mathbf{e}_1 = [1, 0, \dots, 0]$, $\mathbf{e}_2 = [0, 1, \dots, 0]$, and so on. □

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

Example 3.2.7. 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 1.3.6, 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 1.3.13, 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 1.3.13, 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 1.3.13, 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 1.3.13, we have that $(c + d)A = cA + dA$ for all real numbers c and d and all real $m \times n$ matrices A.

Example 3.2.8. Consider the collection $F(\mathbb{R}, \mathbb{R})$ of real functions $f : \mathbb{R} \to \mathbb{R}$. We may define function addition so that if $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are any functions, then $f + g$ is the function satisfying $(f + g)(x) = f(x) + g(x)$ for all real numbers x, and we may define scalar multiplication so that $(\alpha f)(x) = \alpha f(x)$. Observe that the following hold, hence $F(\mathbb{R}, \mathbb{R})$ is a real vector space.

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

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

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

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

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

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

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

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

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

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

Given any vector $\mathbf{0}_V$ of a vector space V satisfying property (4.) of Definition 3.2.5, we say that $\mathbf{0}_V$ is a **zero vector**. We demonstrate that a vector space V has one and only one zero vector.

Proposition 3.2.9. *Given any vector space $(V, +)$, let $\mathbf{0}_V$ be a zero vector of V.*

1.) *Given any vector $\mathbf{u} \in V$ satisfying that $\mathbf{u} + \mathbf{v} = \mathbf{v}$ for every vector $\mathbf{v} \in V$, it must hold that $\mathbf{u} = \mathbf{0}_V$. Consequently, the zero vector of a vector space is unique.*

2.) *We have that $0\mathbf{v} = \mathbf{0}_V$ for all vectors $\mathbf{v} \in V$.*

Proof. (1.) Observe that if $\mathbf{u}$ is any vector of V with the property that $\mathbf{u} + \mathbf{v} = \mathbf{v}$ for every vector $\mathbf{v}$ of V, then it holds $\mathbf{u} + \mathbf{0}_V = \mathbf{u}$ by definition of a zero vector $\mathbf{0}_V$. Conversely, we have that $\mathbf{u} + \mathbf{0}_V = \mathbf{0}_V$ by assumption. We conclude therefore that $\mathbf{u} = \mathbf{u} + \mathbf{0}_V = \mathbf{0}_V$ so that $\mathbf{u} = \mathbf{0}_V$.

(2.) Given any vector $\mathbf{v} \in V$, we obtain a vector $0\mathbf{v} \in V$ satisfying that $0\mathbf{v} = (0+0)\mathbf{v} = 0\mathbf{v} + 0\mathbf{v}$. Consequently, by properties (2.) and (5.) of Definition 3.2.5, there exists a vector $-0\mathbf{v}$ of V such that $0\mathbf{v} = 0\mathbf{v} + \mathbf{0}_V = 0\mathbf{v} + [0\mathbf{v} + (-0\mathbf{v})] = (0\mathbf{v} + 0\mathbf{v}) + (-0\mathbf{v}) = 0\mathbf{v} + (-0\mathbf{v}) = \mathbf{0}_V$. □

Generally, throughout all of mathematics, one of the primary means of classifying an object is to study its subobjects. Given any vector space V, we say that a subset W of V is a **vector subspace** of V (or simply a subspace of V) if the ten properties of Definition 3.2.5 hold for W with respect to the addition and scalar multiplication of V. We provide next a short criterion for subspaces.

Proposition 3.2.10 (Three-Step Subspace Test)**.** *Given any subset W 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 $\mathbf{0}_V$ is an element of W.*

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

3.) *We have that $\alpha\mathbf{w}$ is an element of W for all scalars α and all vectors $\mathbf{w} \in W$.*

Proof. Certainly, if W is a vector subspace of V, then by Definition 3.2.5, it satisfies the second and third properties listed above. Even more, we may consider the zero vector $\mathbf{0}_W$ of W. Considering that W is a subset of V, we may view $\mathbf{0}_W$ as an element of V so that $\mathbf{0}_W + \mathbf{0}_W = \mathbf{0}_W = \mathbf{0}_W + \mathbf{0}_V$. Cancelling $\mathbf{0}_W$ from both sides of this identity yields that $\mathbf{0}_W = \mathbf{0}_V$, 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 3.2.5. 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 $-\mathbf{w}$ is an element of W for all vectors $\mathbf{w} \in W$; then, by the second property above, we have that $\mathbf{w} + (-\mathbf{w})$ is an element of W that satisfies $\mathbf{w} + (-\mathbf{w}) = \mathbf{0}_V$. □

Example 3.2.11. 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 3.2.12. 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 1.3.14; 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 3.2.13. Consider the real vector space $F(\mathbb{R}, \mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ and its subset $\mathcal{C}^1(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ such that f' is continuous. Clearly, the zero function $O : \mathbb{R} \to \mathbb{R}$ is continuous. Likewise, the sum of continuous functions is a continuous function, 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 3.2.14. Consider the real vector space $\mathcal{C}^1(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ such that f' is continuous. Consider the set $W = \{f \in \mathcal{C}^1(\mathbb{R}) \mid f(0) = 0\}$. Clearly, the zero function $O : \mathbb{R} \to \mathbb{R}$ lies in W because it satisfies that $O(0) = 0$; the sum of any functions f and g of W lies in W because we have that $(f+g)(0) = f(0) + g(0) = 0 + 0 = 0$; and the scalar multiple αf of a function $f \in W$ satisfies that $(\alpha f)(0) = \alpha f(0) = \alpha \cdot 0 = 0$, so it must lie in W for all real numbers α. We conclude that W is a real vector subspace of $\mathcal{C}^1(\mathbb{R})$ by the Three-Step Subspace Test.

Example 3.2.15. Consider the real vector space $\mathbb{R}^{n \times n}$ of real $n \times n$ matrices. Consider the subset $W = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is invertible}\}$. Observe that the $n \times n$ zero matrix $O_{n \times n}$ is not invertible, hence it does not lie in W. By the Three-Step Subspace Test, we conclude that W is not a vector subspace of $\mathbb{R}^{n \times n}$. Even more, the subset $W' = \{A \in \mathbb{R}^{n \times n} \mid A \text{ is not invertible}\}$ does not constitute a vector subspace of V: all though the $n \times n$ zero matrix $O_{m \times n}$ lies in W', this set does not satisfy the first property of Definition 3.2.5 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 3.2.16 (Two-Step Subspace Test)**.** *Given any nonempty subset W 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 $\mathbf{v} + \mathbf{w}$ is an element of W for any pair of vectors $\mathbf{v}, \mathbf{w} \in W$.*

2.) *We have that $\alpha \mathbf{w}$ is an element of W for all scalars α and all vectors $\mathbf{w} \in W$.*

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 $-\mathbf{w}$ is an element of W for all elements $\mathbf{w}$ of W. Likewise, if the first condition holds, then by assumption that W is nonempty, we have that $\mathbf{0}_V = \mathbf{w} + (-\mathbf{w})$ is an element of W; we are done by the Three-Step Subspace Test. □

Proposition 3.2.17 (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 $\alpha \mathbf{v} + \beta \mathbf{w} \in W$ for any vectors $\mathbf{v}, \mathbf{w} \in W$ and scalars α, β.*

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

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

Proposition 3.2.18. *Consider any vector space $(V, +)$ with a pair of vector subspaces U and W.*

1.) *Let $U + W$ denote the collection of all vectors $\mathbf{u} + \mathbf{w}$ such that $\mathbf{u}$ is a vector of U and $\mathbf{w}$ 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 $\mathbf{v}$ such that $\mathbf{v}$ 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. We leave this as an exercise for the reader to prove by the Three-Step Subspace Test. □

3.3 Chapter Overview

This section is currently under construction.

References

[Con22] K. Conrad. *The Minimal Polynomial and Some Applications*. 2022. URL: https://kconrad.math.uconn.edu/blurbs/linmultialg/minpolyandappns.pdf.

[DF04] D.S. Dummit and R.M. Foote. *Abstract Algebra*. 3rd ed. John Wiley & Sons, Inc., 2004.

[FB95] J.B. Fraleigh and R.A. Beauregard. *Linear Algebra*. 3rd ed. Addison-Wesley Publishing Company, 1995.

[HK71] K. Hoffman and R. Kunze. *Linear Algebra*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1971.

[Lan86] S. Lang. *Introduction to Linear Algebra*. 2nd ed. Undergraduate Texts in Mathematics. Springer-Verlag New York, Inc., 1986.

[Lan87] S. Lang. *Introduction to Linear Algebra*. 3rd ed. Undergraduate Texts in Mathematics. Springer-Verlag New York, Inc., 1987.

[McK22] J. McKinno. *The Principal Axis Theorem*. 2022. URL: https://www.math.uwaterloo.ca/~jmckinno/Math225/Week7/Lecture2m.pdf.

[Moo68] J.T. Moore. *Elements of Linear Algebra and Matrix Theory*. International Series in Pure and Applied Mathematics. McGraw-Hill Book Company, 1968.

[Smi17] K.E. Smith. *The Spectral Theorem*. 2017. URL: http://www.math.lsa.umich.edu/~kesmith/SpectralTheoremW2017.pdf.

[Str06] G. Strang. *Linear Algebra and Its Applications*. 4th ed. Cengage Learning, 2006.