MA281: Introduction to Linear Algebra

Dylan C. Beck

Contents

Chapter 1

Matrices and Vector Spaces

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

1.1 Matrices and Matrix Addition

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

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

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

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

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

Our principal concern throughout this course are those $m \times n$ arrays whose 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 of any ring, but we will not.

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

Example 1.1.3. Consider the scenario of Example 1.1.1. We may assign to each player a real number called a "skill value" between 0 and 100, e.g., suppose that Alice has skill value a; Bob has skill value b; Carly has skill value c; and Daryl has skill value d. Under this convention, the matrices of Example 1.1.1 yield new matrices called "skill matrices"; they are given by

$$S = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ and } S^t = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

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

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

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

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

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

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

Example 1.1.6. Given any positive integers m and n, there is one and only one matrix consisting entirely of zeros: it is the $m \times n$ **zero matrix**, and it is denoted by $O_{m \times n}$. Often, if we are dealing with the case that $m = n$, then we will simply abbreviate the $n \times n$ zero matrix $O_{n \times n}$ as O_n.

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

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

is the **Kronecker delta**. Put another way, the $m \times n$ identity matrix is the unique $m \times n$ matrix whose (i, j)th entry is 1 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. Explicitly, we have that

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

Like with the zero matrix, we will simply abbreviate the $n \times n$ identity matrix $I_{n \times n}$ as I_n. Observe that the only nonzero components of I_n lie on its main diagonal, hence it is a **diagonal matrix**. Explicitly, a diagonal matrix is an $n \times n$ matrix consisting entirely of zeros off the main diagonal. Even more, by definition, I_n is the unique diagonal matrix whose nonzero entries are all one.

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

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

Observe that the first row of A becomes the first column of A^t (and likewise for the second row). Consequently, the transpose of any $1 \times n$ row vector is an $n \times 1$ column vector. We will also refer to A^t simply as the transpose of A; the process of computing A^t is called **transposition**.

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

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

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

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

Example 1.1.11. 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, if $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ is a real $m \times n$ matrix, then we will typically refer to any real number c as a **scalar**, and we define the **scalar multiple** of A by the scalar c as $cA = \begin{bmatrix} ca_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$. Essentially, we may view this as the sum of the matrix A with itself c times.

Example 1.1.12. Given any $m \times n$ matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$, let $-A = \begin{bmatrix} -a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$. We have that $A + (-A) = O_{m \times n} = -A + A$, and we say that $-A$ is the **additive inverse** of A.

Example 1.1.13. If $A = \begin{bmatrix} a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$ is any $m \times n$ matrix, then $A + A = \begin{bmatrix} 2a_{ij} \end{bmatrix}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$.

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

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

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

1.2 Rotation Matrices and Matrix Multiplication

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

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

Definition 1.2.1. Given any $m \times n$ matrix $A = \left[a_{ij}\right]_{\substack{1\le i\le m\\ 1\le j\le n}}$ and any $n \times r$ matrix $B = \left[a_{ij}\right]_{\substack{1\le i\le n,\\ 1\le j\le r}}$ 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 products of the (i, k)th entry of A and the (k, j)th entry of B for all integers $1 \le k \le n$.

Crucially, the order of A and B in the matrix product matters; the (right) matrix product BA is defined analogously. Be sure to note also that the number of rows of AB is the same as the number of rows of A, and the number of columns of AB is the same as the number of columns of B.

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

Example 1.2.2. Consider the following matrices.

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

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

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

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

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

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

We define the **dot product** $\cdot$ of the vectors v and w as the 1×1 vector

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

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 may be viewed as the $n \times 1$ vector

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

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

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

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

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

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

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

Even more, suppose that today, Alice has the propensity a to attend the concert and Bob has the propensity b to attend, and tomorrow, Alice has the propensity $0.9a + 0.5b$ to attend the concert and Bob has the propensity $0.1a + 0.5b$ to attend. If $p = \begin{bmatrix} a & b \end{bmatrix}^t$ is the "propensity vector," then tomorrow, the propensity that Alice and Bob will attend the concert is given by the matrix product

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

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

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

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

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

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

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

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

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

Example 1.2.6. If A is any $n \times n$ matrix, then the matrix product of A with itself is denoted simply by A^2; it is itself an $n \times n$ matrix, hence we may form the matrix product of A^2 with A. By Proposition 1.2.5, the matrices $(A^2)A$ and $A(A^2)$ are equal; they are denoted simply by A^3. Continuing in this manner, the k-fold product of A is $A^k = AA^{k-1} = A^{k-1}A$ for all integers $k \geq 2$.

Proposition 1.2.7. *If A is any $m \times n$ matrix and B and C are any $n \times r$ matrices, then the product $A(B + C)$ is well-defined; $A(B + C) = AB + AC$; and $A(cB) = c(AB)$ for all scalars c.*

Proof. By Definition 1.1.10, the matrix sum $B + C$ is an $n \times r$ matrix, hence the product $A(B + C)$ is well-defined because the number of columns of A is equal to the number of rows of $B + C$. By Definition 1.2.1, the (i, j)th entry of $A(B + C)$ is the sum of the products of the (i, k)th entry of A and the $(k, j)$the entry of BC for all integers $1 \leq k \leq n$; the latter is by Definition 1.1.10 the sum of the (k, j)th entry of B and the (k, j)th entry of C. Because addition is multiplication is distributive, 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 the proof is similar to the proof of Proposition 1.2.5. $\square$

Ultimately, Proposition 1.2.7 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 $A(cB + C) = c(AB) + AC$. Even more, like matrix addition, matrix multiplication is compatible with transposition.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.3 Elementary Row and Column Operations

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

$$\begin{aligned} a_{1,1} x_1 + \cdots + a_{1,n} x_n &= b_1 \\ a_{2,1} x_1 + \cdots + a_{2,n} x_n &= b_2 \\ &\vdots \\ a_{m,1} x_1 + \cdots + a_{m,n} x_n &= b_m \end{aligned}$$

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

We say that the (real) scalars $\alpha_1, \dots, \alpha_n$ constitute a **solution** to a (real) $m \times n$ system of linear equations if it holds that $a_{i,1} \alpha_1 + \cdots + a_{i,n} \alpha_n = b_i$ for each integer $1 \leq i \leq m$.

Example 1.3.2. One can find many solutions to the matrix equation of Example 1.3.1. Explicitly, $\alpha_1 = 43$ and $\alpha_2 = \alpha_3 = 0$ or $\alpha_1 = 41$, $\alpha_2 = 1$, and $\alpha_3 = 0$ give rise to two distinct solutions.

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

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

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

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

$$\begin{aligned} a_{1,1} x_1 + \cdots + a_{1,n} x_n &= b_1 \\ a_{2,1} x_1 + \cdots + a_{2,n} x_n &= b_2 \\ &\vdots \\ a_{m,1} x_1 + \cdots + a_{m,n} x_n &= b_m \end{aligned}$$

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

$a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n = b_i$ of the $m \times n$ system of linear equations, i.e., the $m \times n$ matrix

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}.$$

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

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

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

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

Example 1.3.4. Rephrasing the results of 1.3.2, the matrices $\alpha_1 = \begin{bmatrix} 43 & 0 & 0 \end{bmatrix}$ and $\alpha_2 = \begin{bmatrix} 41 & 1 & 0 \end{bmatrix}$ give rise to two distinct solutions of the matrix equation of Example 1.3.3. On the other hand, put into the language of matrix equations, the information that $22 = x_1 + 2x_2$ and $15 = x_1 + x_2$ can be most efficiently synthesized by viewing the coefficients of these linear equations as rows of a matrix. Explicitly, we construct a matrix A whose first row is $\begin{bmatrix} 1 & 2 \end{bmatrix}$, corresponding to the respective coefficients of x_1 and x_2 in the equation $22 = x_1 + 2x_2$; the second row of the matrix A is $\begin{bmatrix} 1 & 1 \end{bmatrix}$, corresponding to the respective coefficients of x_1 and x_2 in the equation $15 = x_1 + x_2$. Once again, the column vector x^t consists of the variables x_1 and x_2 in distinct rows, and the column vector b^t consists of the integers 22 and 15 in distinct rows. Ultimately, yields the matrix equation

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

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

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

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

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

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

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

Example 1.3.5. Given the matrices A and b of Example 1.3.4, we obtain the augmented matrix

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

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

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

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

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

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

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

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

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

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

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

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

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

1.4 The Method of Gaussian Elimination in Linear Systems

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(1.) it lies in row echelon form;

(2.) its pivots are all 1; and

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1.5 Invertible Matrices

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

Proposition 1.5.1. *Let A be an $n \times n$ matrix. Every left inverse of A is also a right inverse of A and vice-versa. 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 $LA = I_n = AR$. By Proposition 1.2.5, we have that $L = LI_n = L(AR) = (LA)R = I_n R = R$. Consequently, L is a two-sided inverse of A. Even more, if L' is any two-sided inverse of A, then it is a right inverse of A so that $L' = L$. $\square$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(3.) If E is obtained from I_n by interchanging rows i and j of I_n, then E is its own inverse. □

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Equivalently, the displayed matrix above is $E_3 E_2 E_1 A$ for the 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 that $E_4 E_3 E_2 E_1 A = I_n$ so that $A^{-1} = E_4 E_3 E_2 E_1$. Explicitly, we obtain the following.

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

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

References

[Lan86] S. Lang. *Introduction to Linear Algebra*. second. Undergraduate Texts in Mathematics. Springer, 1986.